ФГБОУ ВО «Пензенский государственный университет», г. Пенза;
ФГКОУ ВО «Воронежский институт МВД России», г. Воронеж;
ФАУ «ГНИИИ ПТЗИ ФСТЭК России», г. Воронеж;
ФГБОУ ВО «Липецкий государственный педагогический университет», г. Липецк;
ФГБОУ ВО «Рязанский радиотехнический университет», г. Рязань;
ФГБОУ ВО «Оренбургский государственный университет», г. Оренбург;
ФГБОУ ВО «МИРЭА – Российский технологический университет», г. Москва;
ФГУП «18 ЦНИИ» МО РФ, г. Москва;
ФГАНУ «Институт сверхвысокочастотной полупроводниковой электроники имени В. Г. Мокерова Российской академии наук» (ИСВЧПЭ РАН), г. Москва;
АО «Пензенский научно-исследовательский электротехнический институт», г. Пенза;
Пензенский филиал АО «Научно-технический центр "Атлас"», г. Пенза;
ООО «Научно-техническое предприятие "Криптософт"», г. Пенза;
АО «Научно-производственное предприятие "Рубин"», г. Пенза;
АО «Производственное объединение "Электроприбор"», г. Пенза;
АО «Радиозавод», г. Пенза;
АО «Системы управления», г. Москва;
Общероссийская общественная организация «Российское научно-техническое общество радиотехники, электроники и связи имени А. С. Попова», г. Тула;
«Научно-исследовательский и конструкторский институт радиоэлектронной техники» филиал ФГУП НПЦ «ПО "Старт"» имени М. В. Проценко;
«Петербургский государственный университет путей сообщения императора Александра I», г. Санкт-Петербург;
Филиал АО «ПНИЭИ» Научно-исследовательское предприятие «Аргус»;
ООО Научно-производственная фирма «Кристалл»;
Филиал Военной академии имени Петра Великого, г. Серпухов;
Обособленное подразделение ОАО «ИнфоТеКС», г. Пенза;
ООО «НПФ "КРУГ"», г. Пенза;
ООО «Научно-производственное предприятие "БиоКрипт"», г. Пенза;
ООО «АЛГОМАТ», г. Калининград

Безопасность информационных технологий

Сборник научных статей по материалам II Всероссийской научно-технической конференции
(г. Пенза, 3 июня 2020 г.)

Пенза Издательство ПГУ 2020

УДК 681.322
Б39

Безопасность информационных технологий : сб. науч. ст. по Б39 материалам II Всерос. науч.-техн. конф. (г. Пенза, 3 июня 2020 г.). – Пенза : Изд-во ПГУ, 2020. – 248 с.

ISBN 978-5-907364-51-6

Рассматриваются различные аспекты безопасности информационных технологий. Публикуемые материалы прошли рецензирование.

Издание предназначено для специалистов по безопасности информационных технологий, преподавателей, аспирантов, докторантов и студентов вузов.

УДК 681.322

URL: https://tsib.pnzgu.ru/BIT

Приказ
о подготовке и проведении II Всероссийской научно-технической конференции «Безопасность информационных технологий» № 376/о от 20.05.2020.

ISBN 978-5-907364-51-6

УДК 004.056

В. И. Волчихин

О ВОЗРАСТАНИИ РОЛИ БЕЗОПАСНОСТИ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В СОВРЕМЕННЫХ УСЛОВИЯХ

Аннотация. Пандемия коронавируса на фоне новых вызовов в области информационной безопасности ставит новые задачи по ее совершенствованию. Использование on-line сервисов выдвигает требования по высоконадежной локальной и дистанционной аутентификации личности по ее неповторимым биометрическим данным, а также защите персональных данных.

V. I. Volchikhin

THE INCREASING ROLE OF INFORMATION TECHNOLOGY SECURITY IN TODAY'S ENVIRONMENT

Abstract. The coronavirus pandemic, against the backdrop of new challenges in the field of information security, poses new challenges to improve it. The use of on-line services puts forward requirements for highly reliable local and remote identity authentication based on its unique biometric data, as well as the protection of personal data.

В январе 2020 г. Всемирный экономический форум (ВЭФ) выделил пять основных угроз: замедляющуюся экономику и социальную напряженность, изменение климата, сокращение биоразнообразия видов, проблемы с кибербезопасностью и новые трудности, с которыми сталкивается здравоохранение.

В числе основных угроз миру ВЭФ также назвал кибератаки, мошенничество и кражу персональных данных. Отдельно в докладе выделен искусственный интеллект, отмечено, что в настоящее время нельзя оценить полный потенциал искусственного интеллекта или его риски [1].

Во время вспышки новой коронавирусной инфекции COVID-19 весь мир столкнулся с уникальным вызовом в области информационной безопасности [2].

Статистика всплеска кибератак в области социальной инженерии и переход на дистанционную работу обусловливает необходимость в совершенствовании нормативно-методической базы защищенного удаленного доступа, в первую очередь для информационных систем, подлежащих защите в соответствии с российским законодательством и требованиями Федеральной службы по техническому и экспертному контролю России и Национального координационного центра по компьютерным инцидентам [3, 4].

Очевидно, что идея дистанционной работы и обучения начала массовую апробацию и она найдет свою нишу в современном технологическом укладе нашей страны.

Анализ сложившейся ситуации еще раз подчеркнет приоритетность создания информационных технологий нового поколения в современную эпоху нашей страны.

Открывая II Всероссийскую научно-техническую конференцию «Безопасность информационных технологий», хочу сказать, что она проходит дистанционно в достаточно сложных условиях режима всемирной пандемии COVID-19, эти факторы вносят коррективы как в программу конференции, так и в обсуждение научных докладов.

При этом отмечаю, что актуальность рассматриваемых на конференции докладов только возрастает, так как они посвящены вопросам защиты информации, локальной и дистанционной аутентификации личности по его неповторимым биометрическим данным, защите персональных данных.

Желаю всем участникам конференции крепкого здоровья, новых творческих успехов.

Библиографический список

1. Давосский форум назвал главные риски 2020 года // Fobes. – URL: https://www.forbes.ru/newsroom/obshchestvo/391219-davosskiy-forum-nazval-glavnye-riski-2020-goda

2. Марков, А. Информационная безопасность в условиях пандемии COVID-19 / А. Марков. – URL: https://russiancouncil.ru/analytics-and-comments/analytics/informatsionnaya-bezopasnost-v-usloviyakh-pandemii-covid-19/?sphrase_id=35369216

3. Об угрозах безопасности информации, связанных с пандемией коронавируса (COVID-19.). НКЦКИ, 2020. – ALRT-20200320.1 (20 марта 2020 г.) – 4 с. – URL: https://safe-surf.ru/upload/ALRT/ALRT-20200320.1.pdf

4. Рекомендации по обеспечению безопасности объектов критической информационной инфраструктуры при реализации дистанционного режима исполнения должностных обязанностей работниками субъектов критической информационной инфраструктуры : письмо ФСТЭК России № 240/84/389 от 20 марта 2020 г. – 2020. – 3 с. – URL: https://fstec.ru/compo-nent/ attachments/download/ 2711

Для цитирования:

Волчихин, В. И. О возрастании роли безопасности информационных технологий в современных условиях / В. И. Волчихин // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 3–5.

УДК 519.2

В. А. Фунтиков

НОВЫЕ ПОДХОДЫ К ВЫСОКОНАДЕЖНОЙ БИОМЕТРИКО-НЕЙРОСЕТЕВОЙ АУТЕНТИФИКАЦИИ МОБИЛЬНОГО ПОЛЬЗОВАТЕЛЯ В УСЛОВИЯХ МИРОВОЙ ПАНДЕМИИ КОРОНАВИРУСА COVID-19

Аннотация. Рассмотрен вариант использования двух искусственных нейронных сетей: «глубокого» «широкого» обучения на персональных биометрических данных пользователя при формировании биометрического контейнера – хранителя ключа, на котором шифруются конфиденциальные сведения. Все операции проводятся в специально разработанной доверенной среде вычисления.

В. А. Фунтиков, сопредседатель оргкомитета конференции, генеральный директор АО «Пензенский научно-исследовательский электротехнический институт», кандидат технических наук

V. A. Funtikov

NEW APPROACHES TO HIGHLY RELIABLE BIOMETRIC-NEURAL NETWORK AUTHENTICATION OF MOBILE USER IN THE CONDITIONS OF THE GLOBAL PANDEMIC OF COVID-19 CORONAVIRUS

Abstract. The option of using two artificial neural networks is considered: "deep" "broad" training on the user's personal biometric data in the formation of a biometric container – the key custodian on which sensitive information is encrypted. All operations are conducted in a specially designed trusted computing environment.

Работа нашей конференции проходит дистанционно, и этот факт еще раз подчеркивает актуальность защиты информации в открытом информационном пространстве, особенно таком, как Интернет.

В настоящее время учеными и специалистами АО «ПНИЭИ» разрабатываются уникальные технические решения высоконадежной инженерной реализации с созданием и глубокой экспертизой доверенных аппаратно-программных сред.

Специалисты института одними из первых в мировой практике синтезировали большие и сверхбольшие искусственные нейронные сети. Это направление исследований открыло перспективу создания высоконадежных биометрико-криптографических систем защиты информации массового применения для электронного документооборота, электронной коммерции, электронных платежей, а также высоконадежной дистанционной аутентификации пользователей этих электронных систем, что важно в сегодняшних непростых условиях пандемии коронавируса COVID-19 [1, 2].

Уязвимость широко развитой в настоящее время технологии электронной подписи – отсутствие надежной, нерасторжимой связи кода ключа подписи с ее владельцем, что органически присуще собственноручной подписи [2]. Поэтому возникает реальная опасность в том, что ключ подписи может быть похищен, передан владельцем третьему лицу. И наконец, ключ может быть скопирован и владелец может вообще не узнать о его хищении. При этом все риски, связанные с хищением ключа подписи, лежат целиком и полностью на владельце ключа, а инструменты, обеспечивающие сохранность ключа подписи, в настоящее время отсутствуют. Указанная уязвимость не позволяет обеспечить надежную аутентификацию и создает потенциальную возможность для мошеннических действий третьих лиц.

Одним из действенных путей устранения данной уязвимости может стать привязка кода ключа к биометрии самого пользователя [2, 3]. Однако использование существующей технологии биометрической аутентификации не в полной мере обеспечивает безопасность ключа. Рассмотрим более подробно суть этой технологии. Берется искусственная нейронная сеть глубокого обучения, которую путем подачи на нее большого количества биометрических образов одного типа (лицо, голос, радужная оболочка глаза и др.) обучают выделению тех или иных физиологических признаков (лицо, радужная оболочка глаза, почерк, голос, вены и др.). Процесс обучения требует больших вычислительных ресурсов и продолжается 1–2 месяца в зависимости от мощности ЭВМ [4–6]. В результате обучения создается нейронная сеть, специализирующаяся на выделении физиологических признаков какого-либо типа (лицо, голос и др.).

Далее специализированная нейронная сеть эксплуатируется в двух режимах – в режиме обучения узнаванию конкретного человека и собственно режиме узнавания конкретного человека.

В режиме обучения узнаванию на вход специализированной сети подается биометрический образ конкретного человека, из которого сеть выделяет физиологические признаки, вычисляет их значения и сохраняет в памяти в связке с идентификатором человека (ФИО, логин и т.п.). По сути, сохраненные значения физиологических признаков являются эталоном биометрического образа конкретного человека.

В режиме узнавания специализированная нейронная сеть также вычисляет значения физиологических признаков предъявленного биометрического образа и сравнивает их с эталоном (эталонами) конкретного человека или множества людей.

Достоинством технологии является то, что для работы специализированной нейронной сети не требуется больших вычислительных ресурсов – достаточно ресурсов обычного смартфона.

Однако с точки зрения информационной безопасности указанная технология обладает существенными недостатками – эталоны биометрических образов и защищаемая информация (как правило, это пароли, криптографические ключи и персональные данные) хранятся в открытом виде и, как следствие, имеют низкую стойкость к различного рода атакам [2, 3].

Кроме того, для разрешения доступа к хранимым секретам или управлению оборудованием используется однобитовое решающее правило «Свой/Чужой», легко модифицируемое хакерскими методами.

Помимо недостатков в части информационной безопасности традиционная технология биометрической аутентификации обладает ограничениями, важными с точки зрения эксплуатации. В частности, аутентификация возможна только на том устройстве, на котором проходило обучение. Перенос процедуры аутентификации на другое устройство или дистанционная аутентификация требует использования криптографически защищенных каналов связи [7].

Указанные недостатки и ограничения снимаются при использовании разработанной сотрудниками института отечественной технологии нейросетевого преобразования биометрических данных человека в криптографические ключи [8].

В основе предлагаемой технологии аутентификации лежит использование криптографических механизмов: шифрования и элек-

тронной подписи. Эти технологии используются давно и зарекомендовали себя как наиболее надежные. Однако и у этих технологий имеются недостатки, связанные с оборотом и хранением ключей [2].

В государственных системах задача обеспечения безопасности при обороте и хранении ключей возложена на режимно-секретные и иные специально созданные органы, которые оснащены необходимым оборудованием и специальными средствами. У обычного гражданина специальных средств обеспечения безопасности хранения ключей нет. Ситуация обостряется тем, что ключи должны быть не только защищены от утраты и хищений, но и доступны практически в любой момент.

Кроме того, ключи электронной подписи могут быть не только украдены или скопированы, но и переданы по сговору. Все это означает, что ключи электронной подписи, в отличие от рукописной подписи, не позволяют точно установить личность человека, что, как следствие, создает предпосылки для мошенничества.

С целью исключения указанных недостатков специалистами АО «ПНИЭИ» была разработана технология нейросетевого преобразования биометрических характеристик в криптографический ключ. Эта технология направлена на устранение сложностей, связанных с хищением ключей, а также на обеспечение невозможности передачи ключей электронной подписи другим лицам [9–13].

По мнению наших специалистов, основными источниками уязвимостей в прикладных системах, требующих подтверждения личности, являются следующие:

1. Использование в качестве аутентифицирующей информации паролей (PIN-кодов), OTP-кодов (одноразовых паролей, в том числе получаемых по SMS) и криптографических ключей, хранящихся в незашифрованном виде. Эта информация может быть подобрана, похищена, скопирована или передана третьим лицам по сговору. Именно это обстоятельство является непреодолимой уязвимостью в системах аутентификации, основанных на знаниях (на владении тайной информацией). Поэтому для аутентификации необходимо использовать физиологические особенности людей, которые не могут быть подобраны, скопированы или переданы по сговору.

2. Хранение критической информации и выполнение операций с критической информацией в недоверенной среде. Если аутентификация выполняется под контролем программ злоумышленника, то от успешных атак не спасут никакие способы аутенти-

фикации и никакие криптографические алгоритмы. Критически важная информация (пароли, ключи шифрования или подписи, номера счетов и т.д.) будет перехвачена, скопирована или изменена и использована во вред клиенту или системе.

Суть технологии состоит в последовательном использовании двух искусственных нейронных сетей. Первая сеть – сеть «глубокого» обучения – специализируется на выделении из «сырых»[1] биометрических данных тех данных, которые являются уникальными для конкретного человека – персональных[2] биометрических данных. Персональные биометрические данные подаются на вход второй – «широкой» – нейронной сети, которая специализируется на формировании криптографических ключей, сформированных установленным порядком и введенным в нее на период обучения.

В отличие от первой нейронной сети вторая сеть не требует предварительного обучения. Эксплуатация второй сети не требует больших вычислительных ресурсов, а ее обучение проходит в реальном времени. Совокупность двух обученных нейронных сетей образует нейросетевой биометрический контейнер, способный формировать из биометрических данных криптографические ключи. Полученный на выходе такого контейнера ключ далее используется для шифрования/расшифрования любой конфиденциальной информации. После использования ключ стирается. Таким образом, в оборудовании, оснащенном биометрическим контейнером, конфиденциальные данные в постоянной памяти хранятся в зашифрованном виде, а ключ шифрования не хранится вообще – он формируется только при предъявлении биометрических данных «своего» и после шифрования/расшифрования сразу же стирается.

Состав биометрического контейнера приведен на рис. 1.

Шаг 1. На вход автомата обучения подается криптографический ключ (в принципе, любая цифровая последовательность, которая может интерпретироваться как ключ, логин, пароль, банковский счет, персональные данные и др.).

Шаг 2. На вход специализированной нейронной сети 1 подаются «сырые» биометрические данные человека, которого надо узнавать.

---

[1] «Сырые» биометрические данные – это полный набор биометрических данных, содержащий как значимые (присущие только конкретному человеку), так и незначимые данные.

[2] Персональные биометрические данные содержат только значимые (уникальные для конкретного человека) данные.

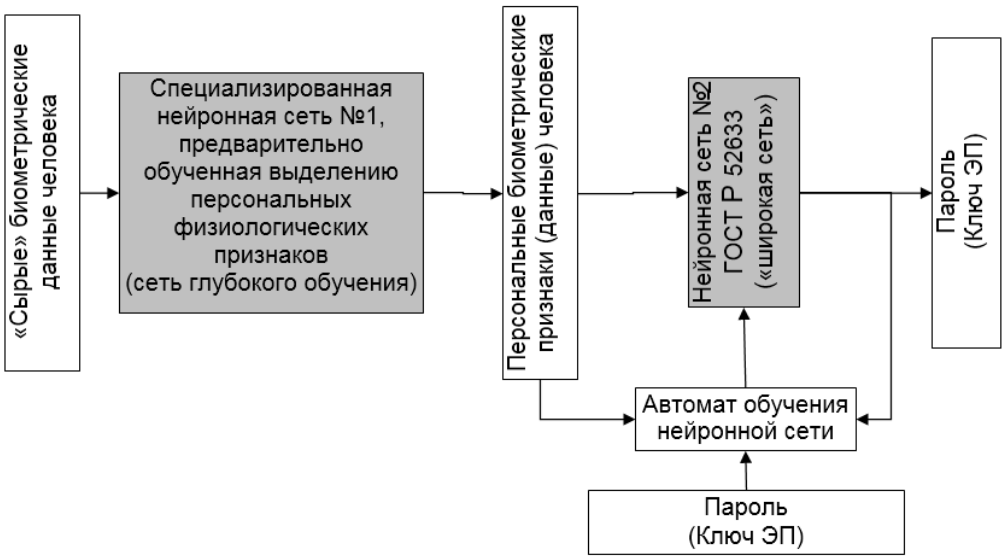


Рис. 1. Биометрический контейнер

Порядок обучения биометрического контейнера

Шаг 3. Специализированная нейронная сеть 1 выделяет из «сырых» биометрических данных персональные биометрические данные.

Шаг 4. Персональные биометрические данные подаются на входы нейронной сети 2 и автомата обучения.

Шаг 5. Автомат обучения рассчитывает коэффициенты нейронной сети.

Шаг 6. Нейронная сеть суммирует персональные биометрические данные с коэффициентами, рассчитанными автоматом, и подает их на выход, который в режиме обучения заведен на вход автомата обучения.

Шаг 7. Автомат обучения реализует метод контролируемого обучения с учителем. Он сравнивает выходные данные нейронной сети с требуемой последовательностью (криптографическим ключом). Если выходные данные равны ключу, то обучение завершается, иначе повторяются операции 5–7.

После обучения биометрические данные и криптографический ключ уничтожаются. В компьютере остается только биометрический контейнер – хранитель ключа, на котором шифруются конфиденциальные сведения.

При необходимости использования конфиденциальных сведений пользователь предъявляет компьютеру свои биометрические

данные, из которых контейнер восстанавливает ключ и расшифровывает конфиденциальные сведения. Важно, чтобы программа шифрования/расшифрования уничтожала ключ сразу после использования.

Приведенное описание является упрощенным, в частности, в нем не учитывается то, что обучение осуществляется не по одному, а по нескольким образам. Исчерпывающее описание структуры контейнера и описание работы нейронной сети 2 приведено в ГОСТ Р 52633 [15].

Следует также отметить, что обучение нейронной сети 1 узнаванию человека не производится и эталоны его биометрического образа не хранятся.

Порядок восстановления пароля/ключа ЭП с помощью биометрического контейнера

Шаг 1. На вход специализированной нейронной сети 1 подаются «сырые» биометрические данные человека, которого надо узнавать.

Шаг 2. Специализированная нейронная сеть 1 выделяет из «сырых» биометрических данных персональные биометрические данные.

Шаг 3. Персональные биометрические данные подаются на входы нейронной сети 2.

Шаг 4. Нейронная сеть 2 суммирует персональные биометрические данные с ранее рассчитанными коэффициентами, результат подается на выход.

Если на вход обученного биометрического контейнера подать биометрические данные «своего», то на его выходе будет сформирован требуемый криптографический ключ.

Если на вход биометрического контейнера подать биометрические данные «чужого», то на его выходе будет сформирована цифровая последовательность, отличная от требуемого ключа.

Как видим, в этой схеме вообще нет решающего правила «Свой/Чужой», нет узнавания человека, а на выходе контейнера есть только правильный или неправильный криптографический ключ, на котором шифруются конфиденциальные данные.

Попытки использовать полученную на выходе цифровую последовательность в качестве ключа будут успешными только в случае, если контейнеру были предъявлены биометрические данные «Свой».

Если биометрические данные принадлежат другому человеку, то на выходе нейронной сети тоже получится ключ, но отличающийся от ключа владельца.

Как уже говорилось, в качестве источника биометрических данных могут использоваться физиологические особенности голоса, лица, почерка, рисунка вен и т.д. Они могут использоваться в качестве самостоятельного или дополнительного фактора аутентификации. Однако не все биометрические характеристики человека подходят для надежной аутентификации. Так, отпечатки пальцев, рисунок ладони и другие статические биометрические характеристики человека, в том числе радужная оболочка глаз, могут быть подделаны. Вопрос только в наличии у злоумышленников соответствующего технологического оборудования. Гораздо более стойкими к копированию и подделке являются динамические биометрические характеристики (динамика рукописного почерка) и биометрические характеристики, дополненные динамическим вводом паролей, например лицо с жестами, которые используются в качестве паролей[1]. Крайне желательно, чтобы для выявления жестов использовалось не программное обеспечение, а нейронная сеть. При этом на выходе нейронной сети одно и то же лицо без жестов и с разными жестами давали сильно разные значения персональных биометрических данных.

Применительно к сотовым телефонам в качестве источника биометрических данных более всего подходят лицо с тайными жестами и голос. Это вызвано тем, что все телефоны имеют микрофоны и оснащено видеокамерами. Для повышения показателей узнаваемости можно использовать мультибиометрические данные – одновременно использовать лицо и почерк, лицо и палец, лицо и пароль, почерк и пароль и т.д. Следует отметить, что выбор типа биометрических данных влияет только выбор специализированной нейронной сети 1 – сети глубокого обучения.

В настоящее время в АО «ПНИЭИ» имеется нейронная сеть глубокого обучения, обученная выделению персональных биометрических данных из лица. Эта сеть реализует наиболее современную на сегодняшний день технологию компьютерного зрения – сверточную искусственную нейронную сеть из 39 слоев, параметры которой получены в результате глубокого обучения. Сеть глу-

[1] Под жестом понимаются повороты головы, моргание, открытие рта и т.д. Для защиты от предъявления, записанного видео система может потребовать от клиента выполнение любых случайным образом выбранных жестов.

бокого обучения преобразует двумерное изображение лица в набор биометрических параметров, значения которых зависят от физиологических особенностей человека. Сеть обеспечивает вероятность ошибочного распознавания на уровне 0,0001 %.

Обеспечение доверенной среды в клиентских устройствах является весьма сложной, а до недавнего момента практически невыполнимой задачей. Для ее решения в состав клиентских устройств должен быть включен доверенный вычислительный элемент. Этот элемент не должен иметь аппаратных «закладок», должен иметь только доверенное ПО, должна быть обеспечена невозможность выполнения в нем посторонних программ, критически важная информация должна обрабатываться только в нем и ни при каких обстоятельствах не покидать его пределы.

В настоящее время отечественными предприятиями микроэлектронной промышленности совместно с АО «ПНИЭИ» с учетом требований действующей нормативной базы разработаны образцы доверенных специализированных изделий в виде SIM- и microSD-карт, в которых доверие к вычислителю определяется за счет следующих факторов:

1) разработка вычислителя и его ПО (загрузчика) проводится исключительно российскими организациями, при соблюдении специальных требований ФСБ России и проведении процедур верификации;

2) аппаратная среда и ПО вычислителя содержат специальные элементы и механизмы, исключающие возможность извлечения из вычислителя критической информации – в нашем случае криптографических ключей.

Суть защиты от извлечения ключей заключается в том, что в микросхеме, реализующей функции шифрования, ключи никогда не покидают пределы специальной микросхемы, так как:

– ключи хранятся в области памяти, недоступной по внешним контактам микросхемы;

– в программном обеспечении вычислителя отсутствует функция передачи ключей во внешние устройства;

– при загрузке нового ПО область памяти, в которой хранятся ключи, гарантированно стирается.

Главное достоинство совокупности использования описанных выше технологий заключается в формировании ключей из биометрических данных человека и возможности отказа от хранения ключей на носителях. Для формирования ключей важно иметь под

рукой смартфон с доверенным вычислителем (в виде SIM-, microSD-карты).

Указанное достоинство позволяет применять совокупность перечисленных технологий в ряде других областей:

– проведение платежей с помощью телефона и протокола NFC;

– системы документооборота, в которых электронная подпись становится действительно подписью, а не печатью;

– действительно электронные удостоверения личности, не привязанные к конкретному носителю и оформленные в виде файлов. Без указанных технологий создание новых паспортов, оформленных в виде приложений в смартфонах, анонсированное Д. А. Медведевым, невозможно;

– электронные удостоверения прав предъявителя, не раскрывающие его личность (персональные данные). Такие удостоверения доказывают лишь принадлежность предъявителя к какому-либо ведомству, например к налоговой инспекции;

– системы электронного голосования;

– медицинские системы, в том числе связанные с обезличиванием историй болезни, назначением курса лечения и подписью рецептов.

Библиографический список

1. Об угрозах безопасности информации, связанных с пандемией коронавируса (COVID-19.). НКЦКИ, 2020. – ALRT-20200320.1 (20 марта 2020 г.) – 4 с. – URL: https://safe-surf.ru/upload/ALRT/ALRT-20200320.1.pdf

2. Рекомендации по обеспечению безопасности объектов критической информационной инфраструктуры при реализации дистанционного режима исполнения должностных обязанностей работниками субъектов критической информационной инфраструктуры. ФСТЭК России, 2020 : письмо ФСТЭК России № 240/84/389 от 20 марта 2020 г. – 3 с. – URL: https://fstec.ru/component/attachments/download/2711

3. Технология использования больших нейронных сетей для преобразования нечетких биометрических данных в код ключа доступа / Б. С. Ахметов, А. И. Иванов, В. А. Фунтиков, А. В. Безяев, Е. А. Малыгина. – Алматы : LEM, 2014. – 144 с. – URL: http://portal.kazntu.kz/files/publicate/2014-06-27-11940.pdf

4. Пат. 46397 Российская Федерация. Способ защиты персональных данных биометрической идентификации и аутентификации / Иванов А. И., Фунтиков В. А., Ефимов О. В. ; № 2007124112/09 ; заявл. 26.06.2007 ; опубл. 10.02.2009, Бюл. № 4.

5. Галушкин, А. И. Нейронные сети: история развития / А. И. Галушкин, Я. З. Цыпкин. – Москва : Радиотехника, 2001. – 840 с.

6. Минский, М. Персептроны / М. Минский, С. Пейперт. – Москва : Мир, 1971. – 261 с.

7. Иванов, А. И. Автоматическое обучение больших искусственных нейронных сетей / А. И. Иванов. – Пенза : Изд-во ПГУ, 2013. – 30 с. – URL: http://пниэи.рф/activity/science/

8. Иванов, А. И. Биометрическая аутентификация личности: обращение матриц нейросетевых функционалов в пространстве метрики Хемминга / А. И. Иванов, Е. А. Малыгина // Вопросы защиты информации. – 2015. – № 1. – С. 23–29.

9. Иванов, А. И. Среда моделирования «БиоНейроАвтограф». Программный продукт создан лабораторией биометрических и нейросетевых технологий, размещен с 2009 г. на сайте АО «ПНИЭИ» / А. И. Иванов, О. С. Захаров. – URL: http://пниэи.рф/activity/science/noc/bioneuroautugraph.zip

10. Майоров, А. В. Оценка стойкости защищенных нейросетевых преобразователей биометрия-код с использованием больших баз синтетических биометрических образов / А. В. Майоров, С. А. Сомкин, А. П. Юнин, А. Ж. Акмаев // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2018. – № 4 (48). – С. 65–74. – DOI 10.21685/2072-3059-2018-4-6.

11. Иванов, А. И. Прогнозирование значений энтропии длинных кодовых последовательностей, порождаемых естественными и искусственными языками / А. И. Иванов, Ю. К. Язов, Д. Н. Надеев, Е. А. Малыгина // Инфокоммуникационные технологии. – 2014. – Т. 12, № 2. – С. 12–14.

12. Майоров, А. В. Быстрые алгоритмы тестирования нейросетевых механизмов биометрико-криптографической защиты информации / А. Ю. Малыгин, В. И. Волчихин, А. И. Иванов, В. А. Фунтиков. – Пенза : Изд-во ПГУ, 2006. – 161 с.

13. Нейросетевая защита персональных биометрических данных / В. И. Волчихин, А. И. Иванов, И. Г. Назаров, В. А. Фунтиков, Ю. К. Язов. – Москва : Радиотехника, 2012. – 160 с.

14. Нейросетевое преобразование биометрического образа человека в код его личного криптографического ключа. – Москва : Радиотехника, 2008. – (Сер. «Нейрокомпьютеры и их применение» Кн. № 29). – 87 с.

15. ГОСТ Р 52633.0–2006. Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации. – Москва : Стандартинформ, 2007. – 24 с.

Для цитирования:

Фунтиков, В. А. Новые подходы к высоконадежной биометрико-нейросетевой аутентификации мобильного пользователя в условиях мировой пандемии коронавируса COVID-19 / В. А. Фунтиков // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 6–16.

УДК 004.056; 004.89; 57.087

В. С. Князьков, А. И. Иванов, А. В. Безяев

НЕОБХОДИМОСТЬ РАСШИРЕНИЯ ФУНКЦИОНАЛЬНЫХ ВОЗМОЖНОСТЕЙ ГОМОМОРФНОГО ШИФРОВАНИЯ ДЛЯ ЗАЩИТЫ НЕЙРОСЕТЕВЫХ РЕШАЮЩИХ ПРАВИЛ БИОМЕТРИЧЕСКИХ ПРИЛОЖЕНИЙ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Аннотация. Рассматривается проблема криптографической защиты решающих правил искусственного интеллекта биометрических приложений. Описаны угрозы, возникающие при использовании международной спецификации BioAPI при применении шифрования биометрических шаблонов обычной криптографией.

Сформулирована проблема защиты нейросетевых решающих правил гомоморфным шифрованием, обусловленная тем, что существующие его реализации ориентированы на использование только двух типов операций (сложение и умножение).

Сделано предположение о том, что добавление к двум операциям сложения и умножения третьей операции сравнения с порогом значительно расширит функциональные возможности гомоморфного шифрования. В этом случае появляется возможность использования защиты гомоморфным шифрованием нейронных сетей, автоматически обученных по ГОСТ Р 52633.5.

V. S. Knyazkov, A. I. Ivanov, A. V. Bezyaev

THE NEED TO EXPAND THE FUNCTIONALITY OF HOMOMORPHIC ENCRYPTION TO PROTECT NEURAL NETWORK CRUCIAL RULES OF BIOMETRIC APPLICATIONS OF ARTIFICIAL INTELLIGENCE

Abstract. The problem of cryptographic protection of the decisive rules of artificial intelligence of biometric applications is considered. The threats posed by the international BioAPI specification when using the encryption of biometric templates by conventional cryptography are described. The problem of protecting neural network decisive rules with homomorphic encryption is formulated, due to the fact that its existing implementations are focused on the use of only two types of operations (addition and multiplication). It has been suggested that adding to the two operations of addition and multiplication of the third comparison operation with the threshold will greatly enhance the functionality of homomorphic encryption. In this case, it is possible to use homomorphic encryption protection of neural networks automatically trained by GOST P 52633.5.

ВВЕДЕНИЕ (история создания и проблемы BioAPI)

Первый стандарт по биометрии создавался как национальный стандарт США в 1993 году, он регламентировал процедуру сжатия рисунка отпечатка пальца перед отправкой их по факсу. Еще одним важнейшим стандартом США является спецификация BioAPI (создана в марте 2000 года, а уже в сентябре этого же года спецификация была реализована на платформе MS Windows [1, 2]). То, что спецификация BioAPI рождалась как национальный стандарт США – это уже история, про которую все забыли.

Новая станица биометрической стандартизации началась в 2002 году с создания международного комитета ISO/IEC JTC1 sc37 (Биометрия). Этот технический комитет с момента своего создания по настоящее время создал, ввел в действие и разрабатывает в настоящее время примерно 153 стандартов. Столь значительная продуктивность, видимо обусловлена использованием госбюджета США, в какой-то форме финансирующего написание международных биометрических стандартов для всех. Свои национальные стандарты США по биометрии уже не пишут, в этом нет нужды. В частности, первая версия стандарт BioAPI-2000 с учетом решения проблем защиты так называемых биометрических шаблонов сегодня превратился в 6 международных стандартов [3–7]. Блок схема реализации BioAPI в криптографически защищенном исполнении приведена на рис. 1, для нее актуальны две существенные угрозы.

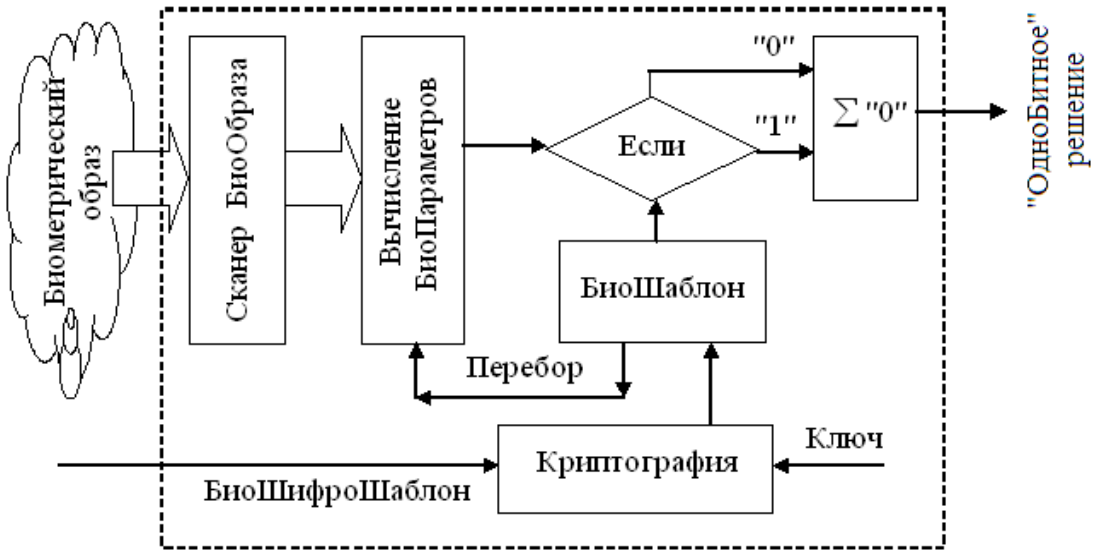


Рис. 1. Защищенная обработка биометрических данных в соответствии с идеологией международной спецификации BioAPI

Во-первых, зашифрованный кем-то биометрический шаблон (БиоШаблон) перед его использованием нужно расшифровать. То есть в доверенной вычислительной среде должен храниться криптографический ключ для расшифрования. Где-то в программном обеспечении биометрического приложения должен быть спрятан «ключ под ковриком».

Во-вторых, спецификация BioAPI ориентирована на использование однобитных решающих правил. Даже если фирма производитель программного обеспечения не публикует его, хакеры рано или поздно обнаружат местоположение в программном коде последнего бита. Изменение всего одного бита приводит к катастрофе, биометрический замок начинает пускать в квартиру всех, кроме ее хозяина.

Перспективы перехода на использование гомоморфного шифрования

К сожалению, полностью доверять, вроде бы респектабельным международным стандартам, нельзя. Одной из причин этого является желание «большого брата» подмять под себя перспективный рынок международных биометрических паспортов в 2000 г., через создание национального стандарта BioAPI. Во время публичного обсуждения BioAPI-2000 биометрическую и криптографическую общественность США удалось успокоить, пообещав появление в ближайшие два-три года гомоморфной криптографии. Вроде бы гомоморфная криптография должна была решить все проблемы применения БиоШаблонов спецификации BioAPI, соответствующая блок схема приведена на рис. 2.

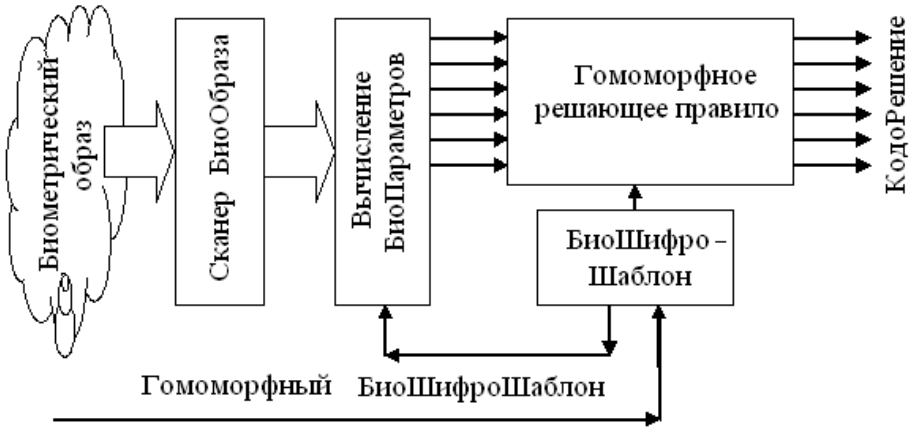


Рис. 2. Обработка биометрических данных, защищенная гомоморфным шифрованием и применением гомоморфного решающего правила

В рис. 2 нет ключа расшифровывания, так как гомоморфное решающее правило способно работать с зашифрованными данными [8–10], исчезла угроза «ключ под ковриком». Кроме того, исчезает проблема решения в один бит, так как гомоморфное решающе правило выдает код, смысловое содержание которого можно узнать только после расшифровывания этого кода.

Необходимо так же отметить, что первые схемы частично гомоморфного шифрования действительно были созданы в конце прошлого века [8, 10] (до 2000 года, когда велось публичного обсуждения BioAPI), однако на них было технически невозможно построить хоть какое-то решающее правило. История развития идеи гомоморфного шифрования отражена в табл. 1.

Таблица 1

Год создания	Авторы криптосхемы гомоморфного шифрования	Основные характеристики достигнутого уровня гомоморфизма
1978 г.	Рональд Риверст, Леонард Адлеман, Майкл Дертузосом (*авторы криптосхемы RSA, ввели понятие гомоморфное шифрование*)	Высказана идея, схемы шифрования пока нет
1982 г.	Шафри Гольдвассер, Сильвио Микали *Частичный гомоморфизм*	Одно умножение
1998 г.	Тацуки Омато, Сигенори Утиямо *Частичный гомоморфизм*	Одно умножение
1999 г.	Паскаль Пэйе *Частичный гомоморфизм*	Одно умножение
2005 г.	Дэн Бонех, Ю Чжин Го, Коби Ниссом *Частичный гомоморфизм*	Одно умножение Неограниченное число сложений
2009 г.	Кейг Дженри *Полный гомоморфизм, очень сложные вычисления, накапливание ошибок*	Неограниченное число сложений Неограниченное число умножений
2012 г.	Цвик Бракерски, Крейг Генри, Видон Вайтунакон *Полный гомоморфизм, приемлемые по сложности вычисления, накапливание ошибок*	Неограниченное число сложений. Неограниченное число умножений

Из табл. 1 видно, что работоспособные схемы гомоморфного шифрования появились только в 2012 году (через 12 лет после спецификации BioAPI). Казалось бы, уже с 2012 года должны быть запущены работы по применению гомоморфного шифрования для

защиты биометрических шаблонов BioAPI. Этого почему-то не произошло, нет ни одного разработанного стандарта по гомоморфному шифрованию, нет сообщений о положительном опыте применения гомоморфного шифрования для защиты биометрических шаблонов BioAPI.

Причина пробуксовки применения гомоморфного шифрования для защиты биометрии проста: выяснилось, что гомоморфные решающие правила не могут быть как угодно большими. Начиная с некоторого размера гомоморфные шифротексты перестают однозначно расшифровываться. Чем длиннее шифротекст, тем больше вероятность, что гомоморфное решение не сможет верно расшифроваться на гомоморфном ключе. Такая ситуация не возникает в классическом шифровании, классическое шифрование (симметричное или асимметричное) позволяет шифровать и расшифровывать тексты любой длинны.

Отечественный подход к обеспечению безопасности обработки хранения и транспорта персональных биометрических данных

Проблемы BioAPI для специалистов по информационной безопасности были понятны всегда, в связи с этим в России как некоторая альтернатива разрабатывается пакет стандартов, ориентированный на использование нейросетевых решающих правил [11]. На рис. 3 приведена блок-схема использования нейросети, обученной по ГОСТ Р 52633.5–2011 [12] для безопасной обработки биометрических данных.

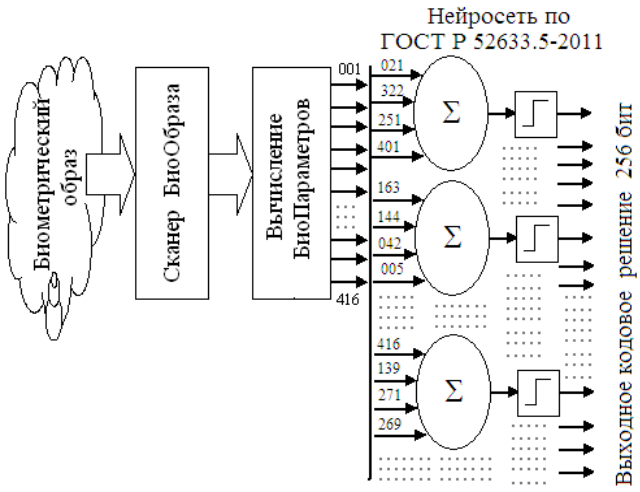


Рис. 3. Нейросетевое преобразование относительно бедных («сырых») биометрических параметров в код длинного криптографического ключа

Из рис. 3 видно, что проблема последнего бита решается за счет использования выходного кода «Свой» длинной 256 бит. Злоумышленник ничего не знающий об этом коде должен его подбирать, а это технически сложная задача. Кроме того, в ближайшее время в России должен появиться документ [13], регламентирующий применение обычных криптографических механизмов для шифрования таблиц связей и таблиц весовых коэффициентов, обученных по ГОСТ Р 52633.5–2011 [12] нейронов.

Таким образом, нейронные сети, стандартизованные в России пакетом из 7 стандартов с номерами ГОСТ Р 52633.xx-20xx следует рассматривать как эффективные решающие правила для искусственного интеллекта биометрических приложений, таблицы которых могут быть зашифрованы обычной криптографией.

Перспективы совершенствования гомоморфного шифрования для защиты данных нейросетевых решающих правил

Рассмотрим ситуацию, при которой для защиты нейросетевых решающих правил будет использоваться полностью гомоморфное шифрование. Очевидно, что, используя его мы легко можем скрыть линейную часть обработки каждым нейроном. Для описания сумматора любого из нейронов достаточно операций сложения и умножения (достаточно уже достигнутого уровня гомоморфизма, смотри нижние строки табл. 1). Однако существующие схемы гомоморфного шифрования не ориентированы на третью операцию сравнения с порогом. По этой причине защитить гомоморфным шифрованием результат нейросетевой обработки сегодня можно только частично. Для полной защиты нейросе6тевой обработки необходимо синтезировать полностью или частично гомоморфные схемы шифрования для трех операций (сложения, умножения, сравнения с порогом).

Можно ли построить схемы шифрования гомоморфные относительно трех базовых операций и какое число таких операций достаточно для практических нужд пока неизвестно. Необходимо начинать работы в этом направлении. Еще одним аргументом в пользу перспектив поиска сочетания гомоморфного шифрования и искусственных нейронных сетей является высокая корректирующая способность последних [14]. При работе в шумах обычно шифротексты накрывают кодами с обнаружением и исправлением ошибок. Формально шифрокод гомоморфного решающего правила

мы можем накрыть кодом с обнаружением и исправлением ошибок, встроив его в криптосхему. Опыт имеющийся у авторов этой статьи показывает, что корректирующая способность искусственных нейронных сетей много выше, корректирующей способности классических кодов, построенных для обнаружения и исправления ошибок [14, 15]. Получается, что тенденция роста вероятности ошибок расшифровывания данных существующих схем гомоморфного шифрования может быть устранена полностью или частично при объединении процедур гомоморфного шифрования с нейронными сетями. Обычные самокорректирующиеся коды не способных исправлять порядка 20 % ошибочных разрядов, тогда как нейросетевые корректоры после обучения легко корректируют 30 % и более ошибочных разрядов кода.

Библиографический список

1. BioAPI // Википедия. – URL: https://ru.wikipedia.org/wiki/BioAPI
2. Болл, Руд. Руководство по биометрии : пер. с англ. / Руд Болл, Джонатан Х. Коннел, Шарат Панканти, Налини К. Ратха, Эндрю У. Сеньор. – Москва : Техносфера, 2007. – 368 с.
3. ГОСТ Р ИСО/МЭК 19784-1–2007. Автоматическая идентификация. Идентификация биометрическая. Биометрический программный интерфейс. Ч. 1. Спецификация биометрического программного интерфейса. – Москва : Стандартинформ, 2009.
4. ГОСТ Р ИСО/МЭК 19785-4–2012. Информационные технологии. Биометрия. Единая структура форматов обмена биометрическими данными. Ч. 4. Спецификация формата блока защиты информации. – Москва : Стандартинформ, 2013.
5. ISO/IEC 24745:2011. Information technology – Security techniques – Biometric information protection. – URL: https://iso.org/obp/ui/#iso:std:52946:en
6. ISO/IEC 24761:2009. Information technology – Security techniques – Authentication context for biometrics. – URL: https://iso.org/standard/41531.html
7. ISO/IEC 19792:2009. Information technology. Security techniques. Security evaluation of biometrics. – URL: https://iso.org/obp/ui/#iso:std:51521:en
8. Гомоморфное шифрование // Википедия. – URL: https://ru.wikipedia.org/wiki/Гомоморфное_шифрование
9. Варновский, Н. П. Гомоморфное шифрование / Н. П. Варновский, А. В. Шокуров // Труды института системного программирования / под ред. В. П. Иванникова. – Москва : ИСП РАН, 2006. – Т. 12. – С. 27–36.
10. Полностью гомоморфное шифрование // Википедия. – URL: https://ru.wikipedia.org/wiki/Полностью_гомоморфное_шифрование
11. ГОСТ Р 52633.0–2006. Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации». – Москва : Стандартинформ, 2007. – 25 с.

12. ГОСТ Р 52633.5–2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия – код доступа. – Москва : Стандартинформ, 2012. – 20 с.

13. Техническая спецификация ТК 26 «Криптографическая защита информации. Защита нейросетевых биометрических контейнеров с использованием криптографических алгоритмов». Окончание публичного обсуждения и голосование по технической спецификации ожидается в 2020 г. – URL: https://tc26.ru/discussions/

14. Безяев, А. В. Оптимизация структуры самокорректирующегося био-кода, хранящего синдромы ошибок в виде фрагментов хеш-функций / А. В. Безяев, А. И. Иванов, Ю. В. Фунтикова // Вестник Уральского федерального округа. Безопасность в информационной сфере. – 2014. – № 3 (13). – С. 4–14.

15. Безяев, А. В. Биометрико-нейросетевая аутентификация: обнаружение и исправление ошибок в длинных кодах без накладных расходов на избыточность : препринт / А. В. Безяев. – Пенза : Изд-во ПГУ, 2020. – 40 с.

Для цитирования:

Князьков, В. С. Необходимость расширения функциональных возможностей гомоморфного шифрования для защиты нейросетевых решающих правил биометрических приложений искусственного интеллекта / В. С. Князьков, А. И. Иванов, А. В. Безяев // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 17–24.

УДК 519.2

А. Е. Сулавко, А. И. Иванов

НАСТРОЙКА И БАЛАНСИРОВКА ДВУХМЕРНЫХ ГИПЕРБОЛИЧЕСКИХ КВАНТОВАТЕЛЕЙ БАЙЕСА В БИНАРНОМ ИСПОЛНЕНИИ, ОБЕСПЕЧИВАЮЩИХ РАВНОВЕРОЯТНЫЕ СОСТОЯНИЯ РАЗРЯДОВ ВЫХОДНОГО КОДА ДЛЯ ОБРАЗОВ «ЧУЖОЙ»

Аннотация. Рассматривается требование к квантователю Байеса, обеспечивающему равновероятные выходные состояния «0» и «1» для бинарного исполнения при тестировании случайными биометрическими образами «Чужой» с произвольной корреляционной сцепленностью. При этом применение подобных квантователей компрометирует знак коэффициента корреляции, подобный уровень компрометации биометрии вполне допустим, если сети искусственных нейронов Байеса используются совместно с парольной аутентификацией, сетями персептронов и/или с сетями квадратичных нейронов.

A. E. Sulavko, A. I. Ivanov

SETTING UP AND BALANCING BAYES' TWO-DIMENSIONAL HYPERBOLIC QUANTA IN A BINARY PERFORMANCE THAT PROVIDE AN EQUAL STATE OF OUTPUT CODE DISCHARGES FOR "ALIEN" IMAGES

Abstract. The requirement for Bayes quantum, providing equal-probable output states "0" and "1" for binary performance when testing random biometric images of "Alien" with arbitrary correlational grip is considered. At the same time, the use of such quantums compromises the correlation factor, a similar level of biometric compromise is quite permissible if the networks of artificial neurons of Bayes are used in conjunction with password authentication, perseptron networks and/or with networks of quadratic neurons.

При создании средств искусственного интеллекта принципиально важно иметь автоматическую процедуру настройки параметров решающего правила и обеспечить эффективное хэширование (перемешивание) данных образов «Чужой». То есть злоумышленник, пытаясь подставлять данные случайных образов «Чужой» на входы защищенного решающего правила искусственного интел-

лекта не должен получить статистически достоверной информации о уязвимостях защиты, например, анализируя вероятности появления состояний «0» и состояний «1» в большом числе разрядов длинного выходного кода. Это все справедливо к любым средствам защиты информации: как защиты на физическом уровне, так и защиты криптографическими механизмами.

Для определенности рассмотрим процедуру настройки двухмерных гиперболических квантователей Байеса или нейронов Байеса [1–3]. Автоматическая настройка этих конструкций выполняется центрированием, нормированием и сортировкой биометрических данных образа «Свой», как это показано на рис. 1.

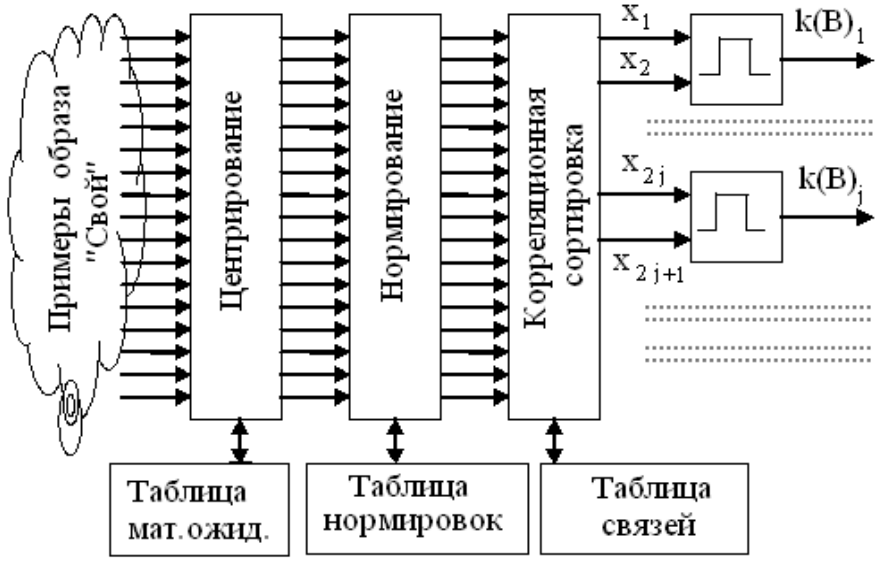


Рис. 1. Настройка сети двухмерных нейронов Байеса сортировкой, выделяющей сильные положительные корреляционные связи

Обычно из биометрических образов удается извлекать сотни биометрических параметров. Так программное обеспечение свободного доступа «БиоНейроАвтограф» [4, 5] преобразует динамику воспроизведения рукописных образов в 416 биометрических параметров. Это означает, что на этих данных можно построить десятки тысяч двухмерных квадратичных нейронов и гиперболических нейронов Байеса. Так как нейроны Байеса наиболее эффективны для сильно коррелированных данных, настройка сети выполняется сортировкой, выделяющей наиболее сильно связанные между собой пары биометрических параметров (например, мы можем выделить 256 пар биометрических параметров с модулем коэффициента корреляции между ними более 0.85). При этом, желательно избежать повторения обрабатываемых сетью параметров.

Результаты подобной сортировки будут являться связями рассматриваемой нейронной сети и записываются в таблицу связей. Заметим, что эта таблица должна связывать между собой обнаруженные пары номеров входных параметров и помечать один из номеров знаком минус, если, вычисленный модуль коэффициента корреляции отрицателен.

Кроме того, должно быть выполнено нормирование данных образа «Свой» и получена таблица нормировок. Еще одной важной для гиперболического квантования операцией является операция центрирования, обрабатываемых биометрических данных. Далее она выполняется при обработке любых биометрических данных как образа «Свой», так и образов «Все Чужие». Для ее выполнения необходимо сформировать таблицу математических ожиданий биометрических параметров образа «Свой». В итоге, мы имеем, три итоговых таблицы обученной распознавать примеры образа «Свой» сети искусственных нейронов, как это показано на рис.1.

Полностью автоматическое формирование трех таблиц рис. 1 на малых выборках от 16 до 24 примеров образа «Свой» эквивалентно формированию основ еще одного национального стандарта России, регламентирующего требования к сетям нейронов Байеса с гиперболическими квантователями. В связи с указом № 490 В. В. Путина от 10.10.2019 «О развитии искусственного интеллекта в Российской Федерации» принципиально важным является создание стандартов по искусственному интеллекту. В том числе нужны стандарты, регламентирующие требования к доверенным приложениям искусственного интеллекта в защищенном исполнении. На текущий момент ряд таких национальных стандартов создан, их названия даны в первых трех строках табл. 1.

Следует отметить, что на сегодняшний день уже стандартизованы процедуры полностью автоматического обучения сетей из искусственных нейронов, осуществляющих обогащение бедных входных данных путем их накопления в линейном пространстве (строка 1 табл. 1). Разработан так же проект стандарта, регламентирующего процедуры автоматического обучения искусственных нейронов, осуществляющих обогащение бедных входных данных накоплением в квадратичном пространстве (строка 2 в табл. 1). Только после полной автоматизации процедур обучения сетей искусственных нейронов, возможен переход к следующему этапу защиты таблиц обученной нейронной сети криптографическими механизмами. Положительный пример, реализации криптографи-

ческой защиты таблиц обученной сети искусственных нейронов с накоплением в линейном пространстве отражен в, соответствующей, технической спецификации (строка 3 табл. 1).

Таблица 1

№	Номер и название национальных стандартов России
1	ГОСТ Р 52633.5-2011 «Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа». Головной разработчик АО «Пензенский научно-исследовательский электротехнический институт»
2	Проект стандарта ГОСТ Р 52633.xx-20xx «Защита информации. Техника защиты информации. Автоматическое обучение сетей квадратичных нейронов с многоуровневым квантованием биометрических данных». Головной разработчик ФГБОУ ВО «Пензенский государственный университет», находится в ТК 362 на этапе подготовки к публичному обсуждению
3	Техническая спецификация «Криптографическая защита информации. Защита нейросетевых биометрических контейнеров с использованием криптографических алгоритмов». Проведено публичное обсуждение первой и второй редакции членами ТК 26 «Криптографическая защита информации». Разработчик АО «Пензенский научно-исследовательский электротехнический институт»
4	Проект стандарта ГОСТ Р xx-20xx «Искусственный интеллект. Автоматическое обучение сетей, состоящих из искусственных нейронов Байеса с гиперболическими квантователями на малых выборках примеров, распознаваемого образа. Находится на этапе разработки, головной исполнитель ФГБОУ ВО «Омский государственный технический университет»

Необходимо отметить, что рассматриваемые в данной статье искусственные нейроны Байеса [1–3] хорошо дополняют квадратичные нейроны и нейроны с линейным накоплением. Нейроны Байеса хорошо работают с сильно коррелированными данными, с которыми не способны работать другие типы искусственных нейронов. В связи с этим, необходимо разработать национальный стандарт, регламентирующий особенности автоматического обучения сетей из искусственных нейронов Байеса (строка 4 табл. 1).

Как было показано выше, формировать три таблицы, отражающие результаты обучения нейронов Байеса не сложно. Основной отличительной особенностью нейронов Байеса является использование ими многомерных гиперболических функций для квантования пространства. Поясним ситуацию на простейшем примере гиперболического квантования двухмерного пространства (рис. 2).

Из рис. 2 видно, что две гиперболы, использующие как асимптоты оси координат x_1, x_2 всегда будут давать на выходе квантователя состояния вероятность появления состояний «0» меньше, чем вероятность появления состояний «1». Для того, что бы сбалансировать вероятности появления всех состояний P(«0») = P(«1») = 0.5 при тестировании образами «Все Чужие» необходимо отказаться от ортогональных асимптот и перейти к асимптотам, отклоненным от ортогональных не угол – φ. Очевидно, что, задав положения асимптот верхней и нижней гипербол мы фактически сжимаем их возможное многообразие [7]. Остается единственный параметр этих гипербол, который определяет минимальное расстояние между ними.

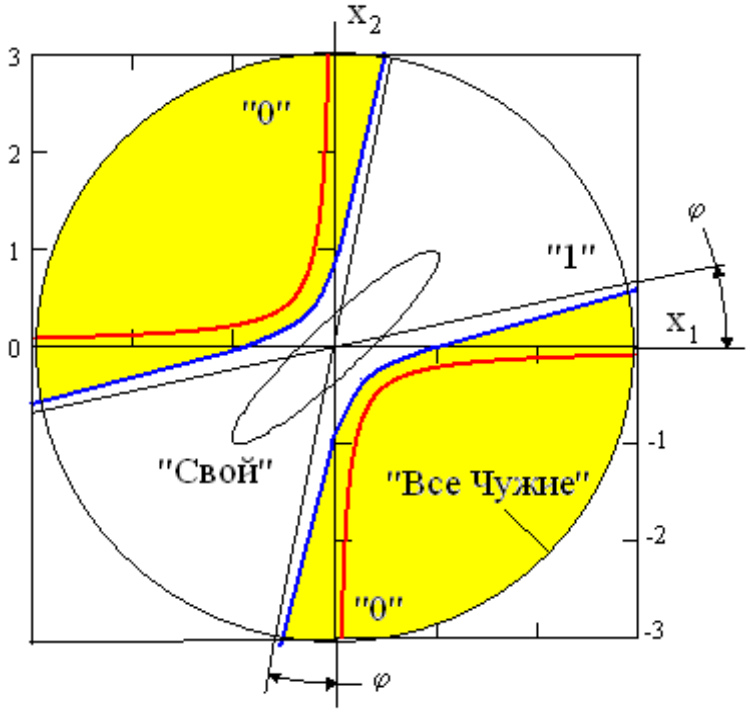


Рис. 2. Примеры гипербол, которые могут быть использованы при квантовании двухмерных данных

Это минимальное расстояние должно быть таким, чтобы данные образа «Свой» с минимальным значением взаимной коррелированности (например, 0.85) попадали в пространство между верхней и нижней гиперболами.

Доверие к приложениям искусственного интеллекта во много определяется тем на каких данные оно было обучено, достаточен ли объем обучающей выборки, и каким алгоритмом выполнялось обучение. Комплекс этих вопросов уже решен в России для сетей искусственных нейронов с накоплением в линейных пространствах

и с накоплением данных в квадратичных пространствах для биометрических данных. В связи с появлением в России технического комитета по стандартизации № 164 «Искусственный интеллект» авторы данной работы считают, что именно в рамках этого нового технического комитета следует разрабатывать новый национальный стандарт, регламентирующий автоматическое обучение сетей из искусственных нейронов Байеса.

Библиографический список

1. Akhmetov, B. B. Multivariate Statistical Analysis of Handwritten Images via Higher Order Correlation Coefficients / B. B. Akhmetov, A. I. Ivanov, P. S. Lozhnikov // Control and Communications (SIBCON) (Moscow, Russia, 12–14 May 2016). – Moscow, 2016. – P. 1–3. – DOI 10.1109/SIBCON.2016.7491759.

2. Иванов, А. И. Оценка надежности верификации автографа на основе искусственных нейронных сетей, сетей многомерных функционалов Байеса и сетей квадратичных форм / А. И. Иванов, П. С. Ложников, А. Е. Сулавко // Компьютерная оптика. – 2017 . – Т. 41, № 5. – С. 765–774.

3. Ложников, П. С. Биометрическая защита гибридного документооборота / П. С. Ложников. – Новосибирск : Из-во СО РАН, 2017. – 130 с.

4. Иванов, А. И. Среда моделирования «БиоНейроАвтограф». Программный продукт создан лабораторией биометрических и нейросетевых технологий, размещен с 2009 г. на сайте АО «ПНИЭИ» / А. И. Иванов, О. С. Захаров. – URL: http://пниэи.рф/activity/science/noc/bioneuroautograph.zip

5. Иванов, А. И. Автоматическое обучение больших искусственных нейронных сетей в биометрических приложениях : учеб. пособие / А. И. Иванов. – Пенза, 2013. – 30 с. – URL: http://пниэи.рф/activity/science/noc/tm_IvanovAI.pdf

6. Гипербола (математическая) // Википедия. – URL: https://ru.wikipedia.org/wiki/Гипербола_(математическая)

Для цитирования:

Сулавко, А. Е. Настройка и балансировка двухмерных гиперболических квантователей Байеса в бинарном исполнении, обеспечивающих равновероятные состояния разрядов выходного кода для образов «Чужой» / А. Е. Сулавко, А. И. Иванов // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 25–30.

УДК 519.2

А. В. Безяев, А. В. Елфимов, А. И. Иванов

МЕХАНИЗМЫ ПРОТИВОДЕЙСТВИЯ УНИЧТОЖЕНИЮ И ДЕГРАДАЦИИ ИСКУССТВЕННЫХ НЕЙРОНОВ, ОБЕСПЕЧИВАЮЩИЕ ВЫСОКИЙ УРОВЕНЬ НАДЕЖНОСТИ РАБОТЫ НЕЙРОСЕТЕВОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Аннотация. Доверие к приложениям искусственного интеллекта во многом определяется наличием механизмов обнаружения и исправления ошибок, возникающих по разным причинам. Ранее основное внимание уделялось ошибкам, возникающим из-за нестабильности биометрических данных. В статье показано, что те же самые механизмы работают и при ошибках, возникающих из-за медленной деградации нейронов вплоть до их полного физического уничтожения, например в жестких условиях воздействия высокого уровня космической радиации.

A. V. Bezyaev, A. V. Elfimov, A. I. Ivanov

MECHANISMS TO COUNTERACT THE DESTRUCTION AND DEGRADATION OF ARTIFICIAL NEURONS, ENSURING A HIGH LEVEL OF RELIABILITY OF NEURAL NETWORK ARTIFICIAL INTELLIGENCE

Abstract. Confidence in artificial intelligence applications is largely determined by the availability of mechanisms to detect and correct errors that occur for a variety of reasons. Previously, the focus was on errors arising from the instability of biometric data. The article shows that the same mechanisms work in errors caused by the slow degradation of neurons up to their complete physical destruction, for example, in harsh conditions of exposure to high levels of cosmic radiation.

В природе повреждение головного мозга живого существа не всегда приводит к гибели его носителя. Люди и животные восстанавливаются и продолжают жить после травм головы и инсультов. Тем же свойством должны обладать искусственные нейронные сети как носители искусственного интеллекта.

В настоящее время большое внимание уделяется «глубоким» искусственными нейронным сетям (ИНС) со структурой, приведенной на рис. 1.

Очевидным является то, что вырождение или уничтожение нейронов нижних слоев «глубокой» ИНС не должно приводить к «катастрофе» инверсии команды управления. Однако, если будет уничтожен последний нейрон или подменен последний бит нейросе-

тевого решающего правила, то последствия могут стать катастрофическими. Попытки скрыть последний бит решающего правила программными методами мало эффективны, атакующий хакер рано или поздно найдет в программе последний бит и подменит его.

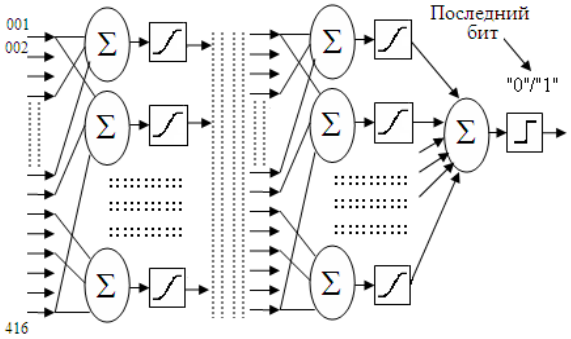


Рис. 1. Типовая структура многослойной нейронной сети «глубокого» обучения с обобщением входных данных до последнего выходного бита

Для того, чтобы лишить злоумышленника возможности атаки через подмену последнего бита в России разработана технология преобразования входных данных нейронной сети в ее выходной код криптографического ключа. Структурная схема такого нейросетевого преобразования приведена на рис. 2.

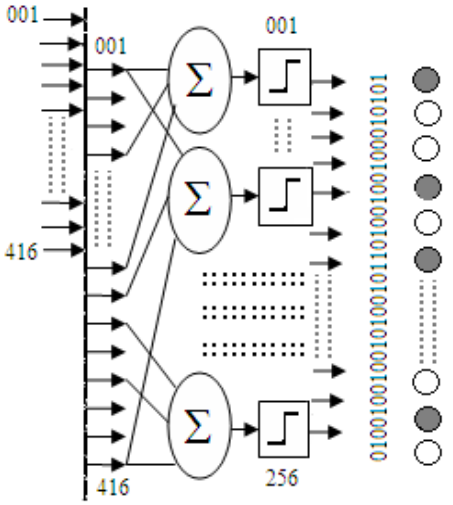


Рис. 2. Однослойная «широкая» сеть из 256 персептронов

Такое техническое решение не оставляет хакерам возможности реализации атаки на «последний» бит, им приходится искать комбинацию ключа из 256 бит, что является задачей высокой вычислительной сложности.

Одним из важнейших свойств «широких» искусственных нейронных сетей является то, что они автоматически учатся на малых обучающих выборках объемом 20 примеров образа «Свой» [1]. Если речь идет о приложениях требующих высокого уровня дове-

рия к решениям, принимаемым искусственным интеллектом, то обучающая выборка может быть увеличена, например, до 40 или 60 примеров биометрического образа «Свой».

Очевидно, что при обучении на выборке в 20 примеров ИНС должна принимать решение «Свой» с доверительной вероятностью 0.95, при обучении на 40 примерах доверительная вероятность увеличивается до 0.975. В ряде приложений искусственного интеллекта доверительная вероятность верного решения должна составлять 0.99999 и более. То есть обучение ИНС должно выполняться на 100 000 примеров образа «Свой», что не всегда технически реализуемо.

Добиться любой заданной доверительной вероятности, удается применением специальных механизмов обнаружения и исправления ошибок [2, 3]. Все эти механизмы строятся на формировании системы синдромов ошибок подсчетом хэш-функций. Схема вычислений показана на рис. 3. Для первого 16-ти битного фрагмента вычисляют первую хэш-функцию, а в качестве синдрома ошибки выбирают часть разрядов этой хэш-функции. Далее вычисляют хэш-функции для 32, 48, …, 256 бит, в конечном итоге получая 16 синдромов ошибок.

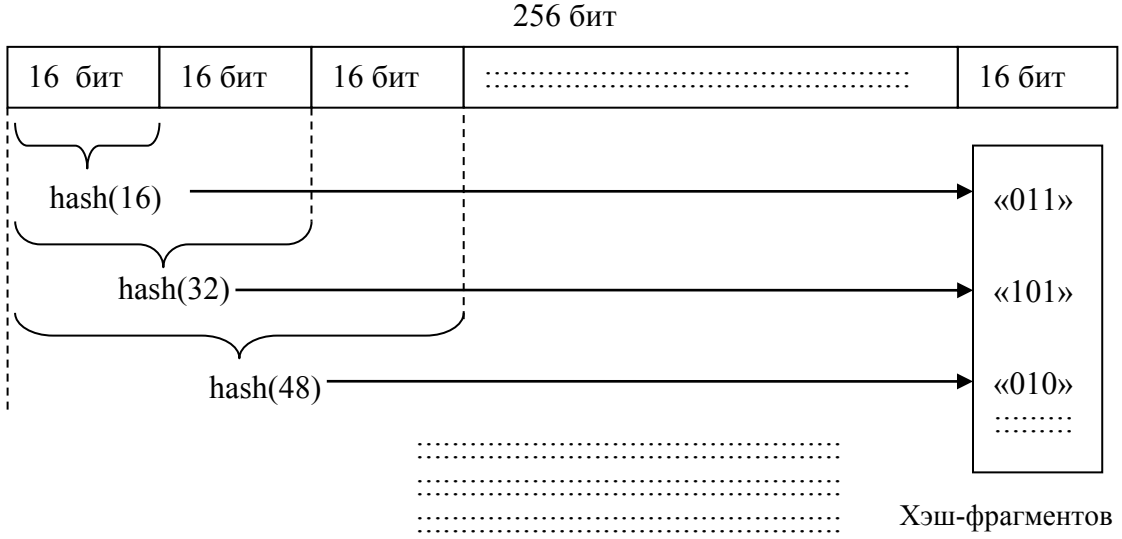


Рис. 3. Безопасная схема рекурсивного формирования эталонных хэш-фрагментов

При процедуре обнаружения ошибок вычисления повторяют в том же порядке, сравнивая новые результаты хэширования с эталонными. Ошибка считается обнаруженной, если вновь вычисленные хэш-фрагменты не совпадают с эталонными. Вероятность вер-

ного обнаружения ошибок определяется длиной хранимого хэш-фрагмента. При хранении одного эталонного разряда ошибка в первых 16 битах обнаруживается с вероятностью $2^{-16} \approx 0.000015$. При хранении трех эталонных разрядов ошибка в первых 16 битах обнаруживается с вероятностью 2^{-48}. Получается компактный и весьма эффективный механизм обнаружения ошибок.

После того, как кодовый блок с ошибкой (с ошибками) обнаружен, необходимо выполнять их поиск и корректировку. В рамках гипотезы одной ошибки при поиске и корректировке инвертировать один разряд в кодовой последовательности пред ее хэшированием, стремясь добиться совпадения вычисляемого хэш-фрагмента с эталонным. Если это не получилось, то приходится переходить к гипотезе наличия двух ошибок в проверяемом блоке и вычислять хэш-функции уже для всех возможных взаимных положений двух ошибок.

Очевидно, что рост числа обнаруживаемых и корректируемых ошибок в коде экспоненциально увеличивает время поиска и их корректировки путем простого перебора. Значительно сократить время поиска и корректировки ошибок удается в случае модуляции статических данных нейросетевой обработки [4]. Модуляция выполняется добавлением случайных мутаций по терминологии ГОСТ Р 52633.2 [5] каждому из 416 биометрических параметров, если они получены в среде моделирования «БиоНейроАвтограф» [6].

Очевидно, что при мутациях нулевой амплитуды ничего не должно измениться показатель стабильности любого разряда всегда единичный:

$$w_i = 2 \cdot \left| 0.5 - P("0_i") \right| = 2 \cdot \left| 0.5 - P("1_i") \right|. \qquad (1)$$

Однако с ростом амплитуды шума мутаций показатель стабильности части разрядов снижается, что отображено на рис. 4.

Для ускорения вычислений [4] инвертирование разрядов при вычислении хэш-функций ведут только по отношению к наиболее нестабильным битам. Обычно у кода примера образа «Свой» нестабильных разрядов в десятки раз меньше стабильных, что и дает значительный выигрыш в сокращении времени вычислений.

Следует подчеркнуть, что появление ошибок в выходном коде может быть связано как с естественной нестабильностью биометрического образа человека, так и с медленной деградацией искусственных нейронов. И та и другая причина дает одно и те же последствия. То есть, рассматриваемый алгоритм корректировки ошибок инвариантен к источнику появления ошибок и может быть применен в любом случае.

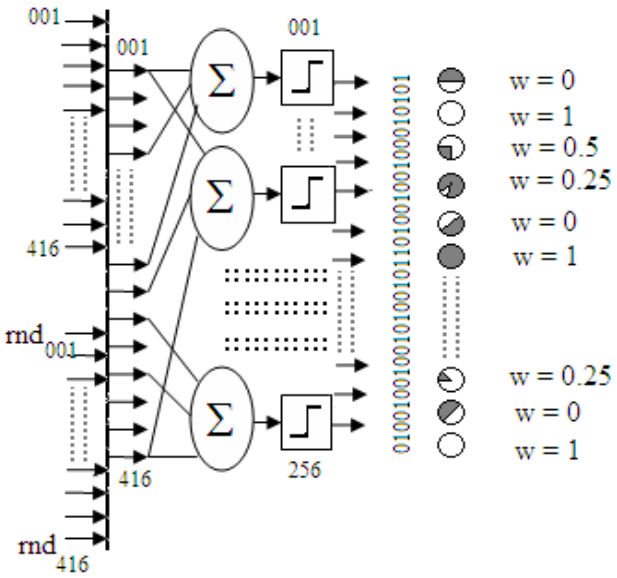


Рис. 4. Модуляция «статических» данных одного БиоПримера случайными мутациями для наблюдения «дрожания» выходных состояний разрядов нейросети

В том случае, если доверенная вычислительная среда реализована аппаратно, и находится тяжелых радиационных условиях, возможна ситуация, когда один из нейронов физически уничтожен. В этом случае, мы будем наблюдать стабильное состояние выходного разряда «уничтоженного» нейрона, независимо от подаваемых на входы нейронной сети случайных состояний. Обнаружение такого «убитого» разряда не сложно. Самым простым способом устранения этой ошибки является маскирование «убитого» разряда верным кодовым состоянием. Последнее эквивалентно компрометации одного бита криптографического ключа.

Очевидно, так же, что чем больше искусственных нейронов будет уничтожено или сильно деградирует, тем больше бит выходного криптографического ключа придется компрометировать маскированием. В этом контексте мы легко можем оценить время постепенного угасания надежности решений искусственного интеллекта и, соответственно, угасания доверия к нему.

Заключение

Криптографическая защита решающих правил искусственного интеллекта – это только один из аспектов доверия к искусственному интеллекту. Принципиально важным является то, что далеко не любое нейросетевой решающее правило искусственного интеллекта поддается его защите криптографическими механизмами от атак хакеров, от медленной деградации нейронов и от полного

уничтожения нейронов. В данной статье мы попытались показать, что в контексте перечисленных выше угроз «широкие» нейронные сети выгоднее «глубоких» нейронных сетей. По крайней мере «широкие» нейронные сети стандартизованы в России и обладают высоким уровнем нейропластичности. Предположительно, что уровень нейропластичности «широких» нейронных сетей может быть сопоставим с уровнем нейропластичности естественных нейронных сетей нашего головного мозга.

Библиографический список

1. ГОСТ Р 52633.5–2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия – код доступа. – Москва : Стандартинформ, 2012. – 20 с.

2. Пат. 2406143 Российская Федерация, G06K 9/03, G07D 7/00. Способ безопасной биометрической аутентификации / Безяев А. В., Иванов А. И. – № 2008126704 ; заявка от 30.06.2008, опубл. 10.12.2010, Бюл. № 34.

3. Безяев, А. В. Оптимизация структуры самокорректирующегося биокода, хранящего синдромы ошибок в виде фрагментов хеш-функций / А. В. Безяев, А. И. Иванов, Ю. В. Фунтикова // Вестник Уральского федерального округа. Безопасность в информационной сфере. – 2014. – № 3 (13). – С. 4–14.

4. Волчихин, В. И. Оценка эффекта ускорения вычислений, обусловленного поддержкой квантовой суперпозиции при корректировке выходных состояний нейросетевого преобразователя биометрии в код / В. И. Волчихин, А. И. Иванов, А. В. Безяев, А. В. Елфимов, А. П. Юнин // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2017. – № 1. – С. 43–55.

5. ГОСТ Р 52633.2–2010. Защита информации. Техника защиты информации. Требования к формированию синтетических биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации. – Москва : Стандартинформ, 2011. – 17 с.

6. Иванов, А. И. Среда моделирования «БиоНейроАвтограф». Программный продукт создан лабораторией биометрических и нейросетевых технологий, размещен с 2009 г. на сайте АО «ПНИЭИ» для свободного доступа / А. И. Иванов, О. С. Захаров. – URL: http://пниэи.рф/activity/science/noc/bioneuroautograph.zip

Для цитирования:

Безяев, А. В. Механизмы противодействия уничтожению и деградации искусственных нейронов, обеспечивающие высокий уровень надежности работы нейросетевого искусственного интеллекта / А. В. Безяев, А. В. Елфимов, А. И. Иванов // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 31–36.

УДК 519.2

С. В. Качалин, К. Н. Савинов, Н. А. Иванова, Т. А. Золотарева

МИНИМАЛЬНЫЙ ФУНКЦИОНАЛ КАЛЬКУЛЯТОРА, ВЫПОЛНЯЮЩЕГО НЕЙРОСЕТЕВУЮ РЕГУЛЯРИЗАЦИЮ ВЫЧИСЛЕНИЯ КОЭФФИЦИЕНТОВ КОРРЕЛЯЦИИ НА МАЛЫХ ВЫБОРКАХ БИОМЕТРИЧЕСКИХ ДАННЫХ

Аннотация. Дан перечень минимальных функциональных возможностей для разрабатываемого в настоящее время калькулятора корреляционных коэффициентов малых выборок. Обосновывается необходимость элементов интерфейса калькулятора и уровня их взаимосвязи. Даны ограничения на динамический диапазон наиболее важных параметров интерфейса.

S. V. Kachalin, K. N. Savinov, N. A. Ivanova, T. A. Solotareva

MINIMUM FUNCTION OF A CALCULATOR THAT PERFORMS NEURAL NETWORK REGULARIZATION OF CORRELATION RATIOS CALCULATIONS ON SMALL SAMPLES OF BIOMETRIC DATA

Abstract. A list of minimal functionality for the currently developed calculator of correlation ratios of small samples is given. The need for elements of the calculator interface and the level of their relationship is substantiated. Restrictions are given on the dynamic range of the most important interface parameters.

Введение

На текущий момент нейросетевые преобразователи биометрий в код криптографического ключа обогащают «сырые» данные в линейном пространстве и обучаются автоматически по алгоритму ГОСТ Р 52633.5–2011 [1] на обучающих выборках объемом от 16 до 26 примеров образа «Свой». Предположительно в ближайшем будущем в России будет принят стандарт, регламентирующие еще один алгоритм обучения квадратичных нейронов с многоуровневыми квантователями [2].

В связи с наметившимся переходом к использованию квадратичных нейронов возникает интерес к задаче регуляризации обучения нейронов Махалонобиса [3] и нейронов Байеса [4, 5]. Проблема состоит в том, что коэффициенты корреляции между двумя пара-

метрами на малых выборках по формуле Пирсона плохо вычисляются. Необходимо выполнить нейросетевую регуляризацию статистических вычислений, например, по способу, изложенному в работе [6]. Если выполнить вычисление коэффициентов корреляции разными способами [7–9], то данные в разных шкалах легко обобщаются, если построить и обучить, соответствующий нейросетевой регуляризатор вычислений.

Интерфейс нейросетевого регуляризатора вычисления коэффициентов корреляции

Очевидно, что экранные формы и интерфейс управления калькулятором вычисления коэффициентов корреляции будет слабо зависеть от того, сколько нейронных сетей будет в нем использовано. Очевидно так же, что интерфейс регуляризатора вычислений должен обеспечивать ввод данных из внешнего файла «rr.txt» и редактирование этих данных. Фрагмент экранной формы, одной из возможных реализаций интерфейса ввода данных приведен на рис. 1.

	4	5	6	7	8	9	10	11	12	13
x	-1.686	0.044	-0.121	0.556	2.192	0.809	0.985	0.862	0.916	0.673
y	0.771	0.306	0.011	-0.764	-0.409	-0.674	-0.132	1.018	0.181	0.513

Рис. 1. Фрагмент интерфейса, определяющий загрузку данных их файла rr.txt и их корректировку

Следует отметить, что длина загружаемого файла может быть любой, однако вычисления с использованием нейросетевых корректоров для длинных файлов в первой версии невозможно. Продукт создается под размер обучающих выборок наиболее востребованных в биометрии от 16 до 26 примеров. При таких размерах выборки калькулятор должен выдавать два результата, как это показано на рис. 2.

Результат вычислений

Корреляция Пирсона	r =	-0.439	±Δ	0.139
НейроКорректировка	rr =	0.142	±Δ	0.042

Рис. 2. Фрагмент интерфейса, отвечающий за вывод результата вычислений при корректных условиях по объему малой выборки входных данных

В случае введения файлов с не допустимой длинной (менее 16 или более 26), отказ от корректировки должен отображаться в явной форме, например, как это показано на рис. 3.

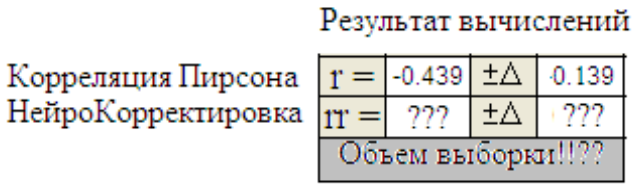


Рис. 3. Фрагмент интерфейса, отвечающий за вывод результата при загрузке данных некорректного объема (менее 16 или более 26)

Важнейшим элементом обеспечения доверия к новому классу программ (нейросетевым корректорам ошибок вычисления коэффициентов корреляции) является наличие встроенного режима тестирования. Фрагмент интерфейса тестирования приведен на рис. 4.

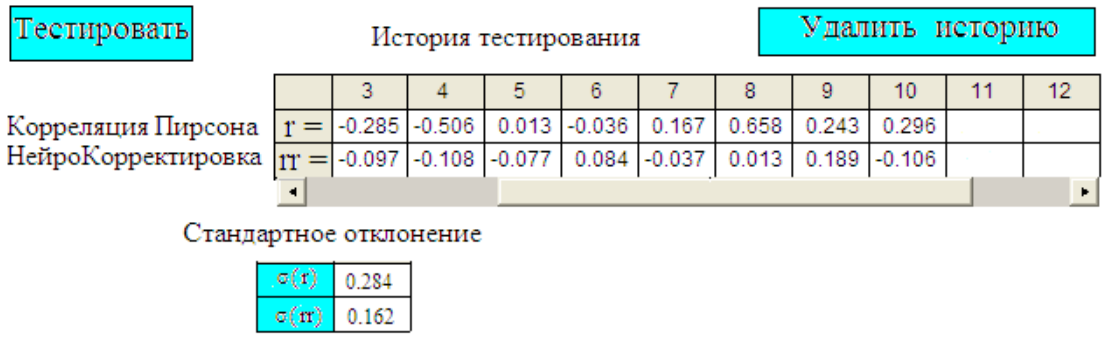


Рис. 4. Фрагмент интерфейса, отвечающий за вывод результата истории тестирования

При инициировании поля «ТЕСТИРОВАТЬ» программа реализации калькулятора должна создавать случайную последовательность «x» из 16 чисел, получая их функцией rnd(1), а так же получать такую же последовательность «y». Для каждой из последовательностей вычисляют коэффициенты корреляции, далее результаты размещают в файл «история_тестирования.txt». Параллельно должен формироваться файл «rnd2.txt», столбцы которого имеют по 32 отсчета (первые 16 – отсчеты «x», вторые 16 – отсчеты «y»). Наличие этого файла позволяет пользователю контролировать то, на сколько корректно калькулятор выполняет вычисление коэффициентов корреляции по формуле Пирсона.

Оценку коэффициента повышения точности вычислений калькулятора с нейросетевой регуляризацией следует оценивать как отношение стандартных отклонений:

$$k = \frac{\sigma(r)}{\sigma(rr)}. \quad (1)$$

При этом выигрыш по эквивалентному увеличению объема выборки исходных данных следует вычислять как квадрат отношения (1). Проверить корректность этих вычислений, можно воспользовавшись любой из стандартных математических программ. В частности для данных рис. 4 выигрыш по точности составляет 1.7 раза, что эквивалентно примерно трех кратному росту выборки с исходных 16 опытов до 46 опытов.

Библиографический список

1. ГОСТ Р 52633.5–2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа.

2. Иванов, А. И. Второй национальный стандарт России по быстрому автоматическому обучению больших искусственных нейронных сетей на малых выборках биометрических данных / А. И. Иванов, А. В. Безяев, Е. А. Малыгина, Ю. И. Серикова // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф. (г. Пенза, 24 апреля 2019 г.). – Пенза, 2019. – С. 174–177.

3. Серикова, Ю. И. Двойная регуляризация процедур обучения нейронов Махаланобиса за счет симметризации корреляционных связей и компенсации ошибок вычисления коэффициентов парной корреляции биометрических данных / Ю. И. Серикова // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф. (г. Пенза, 24 апреля 2019 г.). – Пенза, 2019. – С. 26–34.

4. Ложников, П. С. Биометрическая защита гибридного документооборота : монография / П. С. Ложников. – Новосибирск : Изд-во СО РАН, 2017. – 130 с.

5. Ложников, П. С. Методология защиты смешанного документооборота на основе многофакторной биометрической аутентификации с применением нейросетевых алгоритмов : автореф. дис. … д-ра техн. наук : 05.13.19 / Ложников П. С. – Омск, 2019. – 32 с.

6. Коллекция искусственных нейронов, эквивалентных статистическим критериям для их совместного применения при проверке гипотезы нормальности малых выборок биометрических данных / А. И. Иванов, А. Г. Банных, Е. Н. Куприянов, В. С. Лукин, К. А. Перфилов, К. Н. Савинов // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф. (г. Пенза, 24 апреля 2019 г.). – Пенза, 2019. – С. 156–164.

7. Сериков, А. В. Корреляционная молекула с эллиптическими квантователями для вычислений на малых обучающих выборках / А. В. Сериков, С. В. Качалин // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф. (г. Пенза, 24 апреля 2019 г.). – Пенза, 2019. – С. 123–129.

8. Волчихин, В. И. Квантовая суперпозиция дискретного спектра состояний математической молекулы корреляции для малых выборок биометрических данных / В. И. Волчихин, А. И. Иванов, А. В. Сериков, Ю. И. Серикова // Вестник Мордовского университета. – 2017. – Т. 27, № 2. – С. 230–243.

9. Волчихин, В. И. Тестирование аналогового и квантового оракулов линейной вычислительной сложности, предсказывающих значения коэффициента корреляции на малой выборке в 32 опыта / В. И. Волчихин, А. И. Иванов, А. В. Сериков, Ю. И. Серикова // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2017. – № 3. – С. 70–80.

Для цитирования:

Качалин, С. В. Минимальный функционал калькулятора, выполняющего нейросетевую регуляризацию вычисления коэффициентов корреляции на малых выборках биометрических данных / С. В. Качалин, К. Н. Савинов, Н. А. Иванова, Т. А. Золотарева // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 37–41.

УДК 612.087.1

Е. А. Малыгина

НОВАЯ ПАРАДИГМА ИСПОЛЬЗОВАНИЯ КВАДРАТИЧНЫХ НЕЙРОНОВ С МНОГОУРОВНЕВЫМИ КВАНТОВАТЕЛЯМИ

Аннотация. Рассматриваются преимущества использования квадратичных нейронов с многоуровневым квантованием данных по сравнению с линейными нейронами, обученными по ГОСТ Р 52633.5–2011, в высоконадежных системах аутентификации. Показано увеличение мощности квадратичных нейронов за счет снижения числа необходимых для преобразователя нейронов или снижения числа входов у квадратичных нейронов при сохранении их численности. Все это потенциально усиливает защищенность нового типа нейронных сетей к атакам, так как длина цепочки входных нейронов без общих связей оказывается от трех и более раз больше по сравнению с цепочкой таких же линейных нейронов.

E. A. Malygina

NEW PARADIGM FOR USING SQUARE NEURONS WITH TIERED QUANTIFIERS

Abstract. The benefits of using quadratic neurons with multi-level quantification of data are considered compared to linear neurons trained in GOST P 52633.5–2011, in highly reliable authentication systems. It is shown to increase the power of quadratic neurons by reducing the number of neurons needed for a converter or reducing the number of inputs in quadratic neurons while maintaining their numbers. All this potentially strengthens the protection of a new type of neural networks to attacks, as the length of the chain of input neurons without common connections is three or more times greater than the chain of the same linear neurons.

В настоящее время идет активная работа по использованию биометрических данных граждан при допуске их к электронным государственным и частным ресурсам, платежным системам, хранилищем данных и т.п., размещенным в открытой информационной среде, например, интернет. При этом хочется отметить, что появление больших систематизированных массивов данных, включающих персональные или иные данные граждан представляет лакомый кусок для злоумышленников и требует дополнительной защиты.

События, связанные с пандемией коронавирусной инфекции COVID-19, поставили новые проблемы перед мировым сообществом. Сейчас рассматриваются вопросы онлайн-обучения, которое

станет если не фундаментом общемирового образования, то его неотъемлемой и быстро развивающейся частью.

Работа на «удаленке» офисных сотрудников стала необходимой реальностью, и российские законодатели рассматривают внесение данного пункта в Трудовой Кодекс Российской Федерации [1].

По данным интернет-издания Forbes из-за пандемии в феврале 2020 г. 300 млн работающих жителей Китая вынуждены находиться дома. В марте 2020 г. в США зафиксировано 10 млн американцев, потерявших работу – это, как отмечает издание, больше, чем во время кризиса 2007–2009 гг.

Во всем мире перенесено или отменено несколько сотен значимых мероприятий. Издание отмечает, что в России после режима «самоизоляции» могут закрыться до 40 % предприятий.

Ускорившийся процесс онлайн-сервисов требует разработки новых механизмов защиты данных. Одними из которых являются высоконадежная биометрико-нейросетевая аутентификации личности и разграничения доступа в открытом информационном пространстве, безопасное хранение данных на «облачных» сервисах с использованием стыка биометрии и криптографии [2].

Ученых всегда привлекал вопрос создания искусственных нейронов, при этом делались попытки изучения и копирования работы естественных нейронных сетей живых организмов. Анализируя данные, приведенные в работе [3] можно сделать вывод, что процессы, происходящие в нейронах нервной системы живых организмов разнообразны, активность и количество нейронов во многом зависит от поставленной задачи и требуемой скорости ее решения, но при этом видно, что на выходе как отдельного нейрона или нейронной сети нет однозначного решения «Да/Нет». По всей вероятности, увеличивая число выходов нейрона и, все более вовлекая другие нейроны, также имеющие много выходов, решается любая многомерная задача с наибольшей точностью. При этом необходимо отметить, что данная задача разбивается на отдельные блоки, которые решаются отдельными нейронами параллельно, тем самым значительно увеличивая скорость ее решения.

Первая модель искусственного нейрона и модель нейронной сети впервые была изложена в работе американского нейрофизиолога Уоррена Мак-Каллока (Warren Sturgis McCulloch) и математика Уолтера Питтса (W. Pitts) «Логическое исчисление идей, относящихся к нервной деятельности» («A logical calculus of the ideas immanent in nervous activity»), которая была опубликована в 1943 г. [4].

У. Мак-Каллок и У. Питс создали модель, в которой нейроны рассматриваются как устройство, оперирующее двоичными числами. Авторы предположили, что нейронная сеть теоретически могла выполнять числовые или логические операции любой сложности [4].

В 1957 г. Фрэнк Розенблатт создает программную реализацию искусственного нейрона, названным им «персептроном» и предлагает из них искусственную нейронную сеть, предназначенную для решения задач по распознаванию образов. Он также создает первый нейрокомпьютер «Марк-1», способный обучаться и распознавать рукописные шрифты, что не было случайным выбором, так как во время холодной войны с Советским Союзом правительство США было заинтересовано в быстром переводе большого количества документов с русского языка. Работа Ноама Хомского по грамматике упрощала процесс перевода, однако ученые недооценили сложность лексики переводимого текста. Для перевода текста без ошибок, компьютер должен дополнительно понимать суть переводимого текста [5]. Вследствие слабых мощностей компьютеров и отсутствие эффективных алгоритмов обучения нейронных сетей интерес к технологии значительно снизился.

Только в 80-е гг. XX в. в связи с ростом вычислительной мощности компьютеров разработки новых алгоритмов по обучению нейросетей вновь возрос интерес к их использованию. В настоящее время нейросети, робототехника, компьютеризация стали частью «искусственного интеллекта».

Необходимость дальнейшего развития данной технологии связана с тем, что нейросети способны решать слабо формализованные задачи большой размерности [6]. Нет необходимости знать, как связаны между собой данные и какова сложность реально существующих связей. Вполне достаточно с помощью «Учителя» правильно указывать примеры распознаваемых образов «Свой» и образов «Чужие» [7]. Затем можно применить автомат или полуавтомат обучения, который выполняет следующие операции: выбирает структуру нейронной сети; задаѐт форму функций возбуждения нейронов; задаѐт связи между нейронами; обучает нейроны.

Широкое применение средств биометрической идентификации/аутентификации личности и систем разграничения доступа поставило задачу высоконадежного распознавания биометрических данных личности: папиллярного рисунка отпечатка пальца, сетчатки глаза, голоса, почерка и т.д. [8] На сегодня существует более

27 биометрических параметров распознавания личности и это не предел [9].

Разработанные в период с 2004 по 2013 гг. национальные стандарты серии ГОСТ Р 52633.0–6 [10–16], позволили России по праву занять ведущее место в мире по регламентированию использования линейных нейронных сетей большого размера в вопросах защиты информации. На сегодняшний день подобного рода стандартов не разработано не в одной из ведущих стран Евросоюза, США, Канаде и Китае.

В указанной выше группе российских стандартов, нейросетевые преобразователи биометрия-код строятся, исходя из применения в нейроне входного сумматора (линейного элемента) для обогащения входных данных перед их квантованием.

Это связано с относительной простотой и абсолютной устойчивостью используемого алгоритма обучений нейронов с линейными функционалами предварительного обогащения входных данных [17].

Однако из теории известно, что квадратичные функционалы повышения качества данных так же имеют простые и устойчивые алгоритмы обучения [18–20]. Другими словами, в ряде биометрических приложений вместо сетей нейронов вполне могут быть использованы сети квадратичных форм.

Мощность нейронных сетей, обученных по ГОСТ Р 52633.5–2011 [15] можно усилить за счет перехода от обычных нейронов к квадратичным нейронам, осуществляющим обогащение данных одновременно как в линейном, так и в квадратичном пространствах (рис. 1).

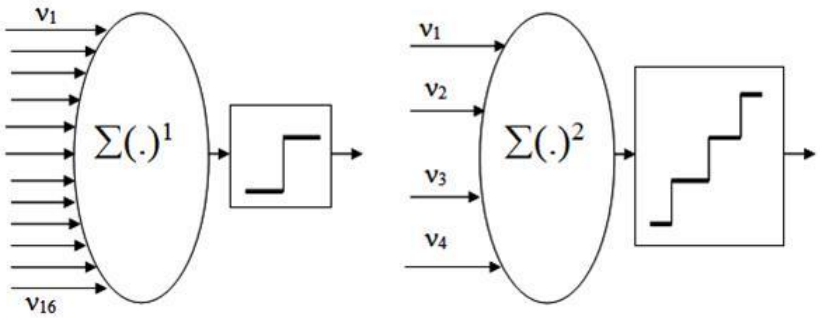


Рис. 1. Два нейрона с близкими вероятностями ошибок первого и второго рода: линейный нейрон – левая часть рисунка; квадратичный нейрон – правая часть рисунка

Рассмотрим возможные преимущества использования квадратичных нейронов в преобразователях биометрия-код. Один двухмерный квадратичный нейрон может быть заменен 4-мя двухмер-

ными линейными нейронами, каждый из которых имеет линию разделения в виде касательной к эллипсу квадратичного нейрона «Свой». При этом вычислительная сложность алгоритмов обучения квадратичных нейронов должна увеличиться с линейной до полиномиальной (квадратичной). Один квадратичный нейрон можно заменить 3, 4, 5 и более линейными нейронами, что свидетельствует о низкой относительной мощности линейных нейронов [21].

Как результат, переход от линейных нейронов к квадратичным нейронам всегда сопровождается снижением числа необходимых для преобразователя нейронов или снижением числа входов у квадратичных нейронов при сохранении их численности. Один квадратичный нейрон по своей эффективности эквивалентен 3, 4 нейронам с линейными разделяющими функциями. При этом вместо 16 входов у линейных нейронов достаточно 4 входов у квадратичных нейронов (рис. 1). Все это потенциально усиливает защищенность нового типа нейронных сетей к атакам, так как длина цепочки входных нейронов без общих связей оказывается от трех и более раз больше по сравнению с цепочкой таких же линейных нейронов.

Основным недостатком квадратичных нейросетевых функционалов является то, что выходной код для образов «Чужой» обладает низкой энтропией, связанной с отсутствием баланса состояний «0» и «1» в разрядах выходного кода. Данный недостаток можно устранить за счет применения выходного квантователя на выходе сумматоров нейронов с тремя и более выходными состояниями [22].

Предположительно, что через несколько лет будут проведены исследования новой схемы нейросетевых преобразований и для квадратичных нейронов будут разработаны специфические атаки и созданы механизмы защиты.

При этом следует отметить, что линейные и квадратичные нейроны должны использоваться совместно, так как полиномы всегда сильнее, чем каждая из компонент полинома при решении задач идентификации, аутентификации, аппроксимации [23, 24].

В настоящее время методы и алгоритмы обработки биометрической информации, разработанные в России в 2006–2012 гг., подвергаются анализу и переработке. Замена линейных нейронов на квадратичные, использование многоуровневых квантователей и механизмов криптографической защиты [2] позволит получать новые результаты, позволяющие обеспечить должный уровень защиты информации и разграничения доступа к ней, как локально, так дистанционно в открытой информационной среде интернет.

Библиографический список

1. О внесении изменений в Трудовой кодекс Российской Федерации в части регулирования дистанционной и удаленной работы : проект Федерального закона. – URL: https://rg.ru/2020/06/16/proekt-udalenka-site-dok.html

2. Техническая спецификация (проект ТК 26 «Криптографическая защита информации»). Защита нейросетевых биометрических контейнеров с использованием криптографических алгоритмов. – URL: https://tc26.ru/discussions/

3. Николлс, Джон. От нейрона к мозгу / Джон Николлс, Мартин Роберт, Валлас Брюс, Фукс Пол ; пер. с англ. П. М. Балабана, А. В. Галкина, Р. А. Гиниатуллина, Р. Н. Хазипова, Л. С. Хируга. – Москва : Едиториал УРСС, 2003. – 672 с.

4. Мак-Каллок, У. С. Логическое исчисление идей, относящихся к нервной активности / У. С. Мак-Каллок, В. Питтс // Автоматы : сб. / под ред. К. Э. Шеннона и Дж. Маккарти. – Москва : Изд-во иностр. лит., 1956. – С. 363–384.

5. Первая модель искусственного нейрона // Яндекс. Дзен. – URL: https://zen.yandex.ru/media/aiqcnt/pervaia-model-iskusstvennogo-neirona-5becfc485f5dcb00a99bd17b

6. Волчихин, В. И. Быстрые алгоритмы обучения нейросетевых механизмов биометрико-криптографической защиты информации / В. И. Волчихин, А. И. Иванов, В. А. Фунтиков. – Пенза : Изд-во ПГУ, 2005. – 276 с.

7. Иванов, А. И. Нейросетевые алгоритмы биометрической идентификации личности / А. И. Иванов. – Москва : Радиотехника, 2004. – 144 с. – (Сер.: Нейрокомпьютеры и их применение. Кн. 15).

8. Руководство по биометрии : пер. с англ / Р. М. Болл, Дж. Х. Коннел, Ш. Панканти, Н. К. Ратха, Э. У. Сеньор. – Москва : Техносфера, 2007. – 368 с.

9. Нейросетевая защита персональных биометрических данных / В. И. Волчихин, А. И. Иванов, И. Г. Назаров, В. А. Фунтиков, Ю. К. Язов. – Москва : Радиотехника, 2012. – 160 с.

10. ГОСТ Р 52633.0–2006. Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации. – Москва : Стандартинформ, 2007. – 25 с.

11. ГОСТ Р 52633.1–2009. Защита информации. Техника защиты информации. Требования к формированию баз естественных биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации. – Москва : Стандартинформ, 2010. – 24 с.

12. ГОСТ Р 52633.2–2010. Защита информации. Техника защиты информации. Требования к формированию синтетических биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации. – Москва : Стандартинформ, 2011. – 17 с.

13. ГОСТ Р 52633.3–2011. Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора. – Москва : Стандартинформ, 2018. – 12 с.

14. ГОСТ Р 52633.4–2011. Защита информации. Техника защиты информации. Интерфейсы взаимодействия с нейросетевыми преобразователями биометрия – код доступа. – Москва : Стандартинформ, 2019. – 42 с.

15. ГОСТ Р 52633.5–2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа. – Москва : Стандартинформ, 2012. – 20 с.

16. ГОСТ Р 52633.5–2011. Защита информации. Техника защиты информации. Защита информации. Техника защиты информации. Требования к индикации близости предъявленных биометрических данных образу «Свой». – Москва : Стандартинформ, 2018. – 20 с.

17. Иванов, А. И. Нейросетевые технологии биометрической аутентификации пользователей открытых систем : дис. … д-ра техн. наук : 05.13.01 / Иванов А. И. – Пенза, 2002. – 391 с.

18. Хайкин, С. Нейронные сети: полный курс / С. Хайкин. – Москва : Вильямс, 2006. – С. 1104.

19. Галушкин, А. И. Нейронные сети: история развития / А. И. Галушкин, Я. З. Цыпкин. – Москва : Радиотехника, 2001. – 840 с.

20. Волчихин, В. И. Перспективы использования искусственных нейронных сетей с многоуровневыми квантователями в технологии биометрико-нейросетевой аутентификации / В. И. Волчихин, А. И. Иванов, В. А. Фунтиков, Е. А. Малыгина // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2013. – № 4 (28). – С. 86–96.

21. Волчихин, В. И. Соотношение мощности нейронов с линейным и квадратичным обогатителями биометрических данных / В. И. Волчихин, А. И. Иванов, Е. А. Малыгина, А. П. Юнин // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2018. – № 1 (45). – С. 17–25. – DOI 10.21685/2072-3059-2018-1-2.

22. Малыгина, Е. А. Условие корректной оценки стойкости к атакам подбора преобразователей биометрия-код с нейронами, осуществляющими многоуровневое квантование / Е. А. Малыгина // Безопасность информационных технологий : сб. тр. науч.-техн. конф. кластера Пензенских предприятий, обеспечивающих безопасность информационных технологий. – Пенза : Изд-во ПГУ, 2014. – Т. 9. – С. 12, 13.

23. Иванов, А. И. Закон распределения значений критерия Крамера – фон Мизеса для проверки гипотезы нормальности малых выборок / А. И. Иванов, Е. А. Малыгина, С. Е. Вятчанин, С. В. Туреев // Электронные информационные системы. – 2019. – № 1 (20). – С. 97–105.

24. Быстрый алгоритм обучения больших сетей искусственных нейронов квадрата среднего геометрического плотностей распределения значений многомерных биометрических данных / В. И. Волчихин, А. И. Иванов, К. А. Перфилов, Е. А. Малыгина, Ю. И. Серикова // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2018. – № 3 (47). – С. 23–35. – DOI 10.21685/2072-3059-2018-3-3.

Для цитирования:

Малыгина, Е. А. Новая парадигма использования квадратичных нейронов с многоуровневыми квантователями / Е. А. Малыгина // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 42–48.

УДК 519.2

А. П. Юнин, А. И. Иванов, А. В. Строков, С. Р. Махсудов

НЕЙРОСЕТЕВОЕ ОБОБЩЕНИЕ ТРЕХ СТАНДАРТНЫХ ТЕСТОВ КОНТРОЛЯ КАЧЕСТВА «БЕЛОГО ШУМА», ПОЛУЧАЕМОГО ХЕШИРОВАНИЕМ СЛУЧАЙНОЙ ЧАСТИ БИОМЕТРИЧЕСКИХ ДАННЫХ

Аннотация. Рассматриваются нейросетевые обобщения трех тестов качества случайных последовательностей длиной 256 бит. Тестирование ориентировано на реализацию в доверенной вычислительной среде микро-ЭВМ с 32-разрядной двоичной системой. Как функция локального хэширования рассматривается генератор Парка – Миллера.

A. P. Yunin, A. I. Ivanov, A. V. Strokov, C. P. Makhsudov

NEURAL NETWORK GENERALIZATION OF THREE STANDARD QUALITY CONTROL TESTS OF "WHITE NOISE" OBTAINED BY HASHING A RANDOM PIECE OF BIOMETRIC DATA

Abstract. Neuronet summarys of three tests of the quality of random sequences length of 256 bits are considered. Testing is focused on the implementation in a trusted computing environment micro computer with 32 bit binary system. As a feature of local hashing is considered the Park-Miller generator.

Введение

Средства биометрической защиты цифровых прав граждан России могут быть эффективными только в случае выполнения биометрических и криптографических преобразований в доверенной вычислительной среде [1]. Например, в качестве доверенной вычислительной среды могут быть использованы 4-х битные процессоры RFID идентификационных карт, 8-ми битные процессоры SIM карт или микро SD карт. Как результат криптографические и биометрико-нейросетевые вычисления должны выполняться на 4-х, 8-ми, 16-ти, 32-х битных процессорах с малым потреблением и, соответственно, ограниченными вычислительными ресурсами.

Еще одним условием безопасности доверенных вычислений является создание криптографических ключей внутри доверенной

вычислительной среды. Для этой цели, например, может быть использован линейный регистр с обратной связью [2]. Для решения этой же цели может быть использован линейный конгруэнтный генератор псевдослучайной последовательности Парка-Миллера [3, 4]:

$$X_{k+1} = X_k \cdot 75 \cdot \mathrm{mod}(2^{31} - 1), \qquad (1)$$

где $(2^{31} - 1)$ – это длина псевдослучайной последовательности, X_0 - начальное состояние в виде 32 битного случайного числа (ключа для генератора).

Просто программный генератор псевдослучайных чисел (1) не вызывает доверия у криптографической общественности из-за своей простоты и предсказуемости.

Получение случайных начальных состояний генератора Парка-Миллера из нестабильной части биометрических данных

Для того, что бы сделать генератор Парка-Миллера действительно случайным достаточно ограничиться получением от него серии укороченных последовательностей, например, длинной 2^{10} при этом необходимо использовать для каждой укороченной последовательности свое значение начального состояния – X_0. Начальные состояния генератора могут быть получены:

1) от физического генератора «белого шума» например, теплового шума туннельного диода,

2) из случайной составляющей биометрических данных, например, состоящей из младшего бита каждого из 416 биометрических параметров динамики случайного рукописного образа из 5 букв [5, 6].

Гарантией случайности в первом случае является тепловой шум, во втором случае гарантией является случайная составляющая биометрических данных. Пользователь не способен точно воспроизвести даже один и тот же рукописный образ. Каждый из биометрических параметров имеет свою длину значимых разрядов (до старшего значимого разряда), последний младший бит является случайным.

На рис. 1 приведены примеры двух рукописных образов «*Пенза*», каждый из которых средой «БиоНейроАвтограф» [5] через вычисление двухмерного преобразования Фурье в 416 биометрических параметров, открыто хранящихся в файле «params.txt».

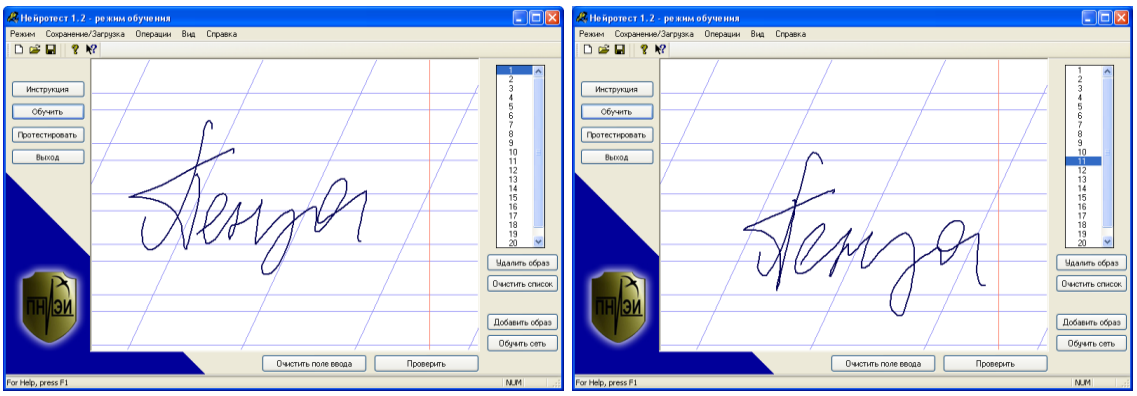


Рис. 1. Сбор биометрических данных в среде моделирования [4]

В табл. 1 приведены значения параметров, соответствующих разным примерам одного и того же образа «Пенза». В последних двух столбцах таблицы приведены математическое ожидание параметров и их стандартное отклонение.

Таблица 1

Параметры с 001-го по 416-тый для 4 разных примеров биометрического образа «*Пенза*»

№	Пример-1	Пример-2	Пример-3	Пример-4	.....	Мат. ож.	СКО
001	15.443	7.086	8.157	7.171	.....	10.21	2.882
002	–4.649	–2.065	–4.577	–4.827	.....	–1.491	2.245
003	–4.462	–3.628	–5.875	–3.847	.....	–4.368	2.173
004	–14.861	–14.174	–17.039	–17.394	.....	–16.21	1.664
.....							
416	3.211	8.002	4.777	1.622	.....	4.997	1.727

Данные табл. 1 приведены в десятичной форме для удобства чтения. Они подтверждают тезис о том, что младшие разряды 416 контролируемых параметров случайны, даже для одного биометрического образа. Повторить случайную последователь человек не может, даже при своем желании. Это означает, что биометрические данные действительно могут являться источником случайности.

То есть, случайные состояния младших разрядов биометрических данных могут быть дополнительно хешированы преобразованием Парка-Миллера (1) и, соответственно, мы можем получить серию криптографических ключей длинной 256 бит. В частности 2 младших разряда дают серию длинной $416 \times 2 = 832$ бит, что позволяет получить, как минимум, три не перекрывающихся ключа

длинной 256 бит (256 × 3 = 768 бит). На самом деле, мы можем получить множество ключей, например, сдвигом 256 бит ключа с шагом 16 бит. Тогда мы получим серию из 36 вариантов 256 битных ключей. Если же создать кольцо обратной связи, объединив конкатенацией начало и конец 832 битной последовательности, то мы получим очень длинную автосвертку этой последовательности с хэширующим преобразованием Парка-Миллера (1). Получается, что из нестабильной компоненты одного примера биометрического образа хэшированием Парка-Миллера мы можем получить серию из сотен вариантов потенциальных криптографических ключей.

Отбраковка части случайных последовательностей с использованием трех тестов NIST

Таким образом, открывается возможность сортировки сотен вариантов претендентов на использование в качестве криптографического ключа. Для этой цели, например, могут быть использованы тесты NISN (Национального института стандартизации США). Формально любой из 16 тестов NISN можно рассматривать как некоторый статистический критерий близости исследуемой 256 битной случайной последовательности к «идеальному» белому шуму.

С другой стороны, любому статистическому критерию может быть построен эквивалентный ему нейрон [7]. Это означает, что для всех статистических критериев (для всех тестов NIST) могут быть построены эквивалентные нейроны. Например, первый тест NIST сводится к подсчету единиц в 256 битном ключе. Математическое ожидание числа единиц в ключе составляет E(a1) = 128, стандартное отклонение σ(a1) = 8. Пример распределения данных приведен на рис. 2.

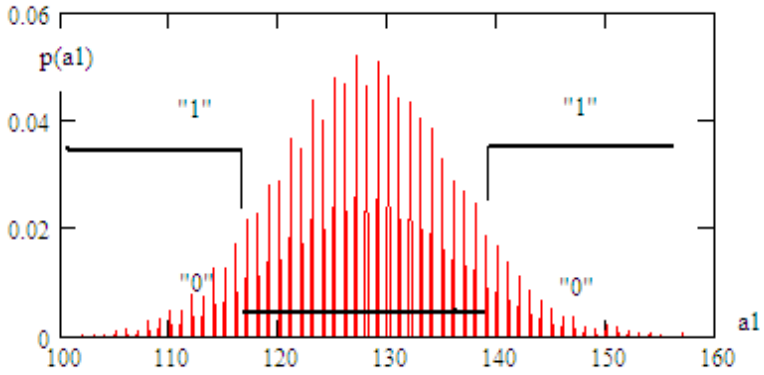


Рис. 2. Выходные состояния нейрона-a1, анализирующего число единиц в коде

Очевидно, что наиболее вероятные состояния белого шума будут попадать в цент распределения, отображенного на рис. 2. В связи с этим мы может выполнить выделение последовательностей наиболее похожих на «идеальный» белый шум. Получается формальный нейрон, дающий «0» расстояние до белого шума при числе единичных состояний от 118 до 138. Примерно 15 % последовательностей с меньшим или большим числом единиц отбрасываются как слишком далекие от «идеального» белого шума.

Еще одним критерием NIST является подсчет числа единиц в 8-ми не перекрывающихся фрагментах по 32 бита. Математическое ожидание числа бит в таком блоке $E(\ .\ ) = 16$, стандартное отклонение $\sigma(\ .\ ) = 2.82$. Распределение выходных состояний квадратичного нейрона оценки качества белого шума – a2, приведено на рис. 3.

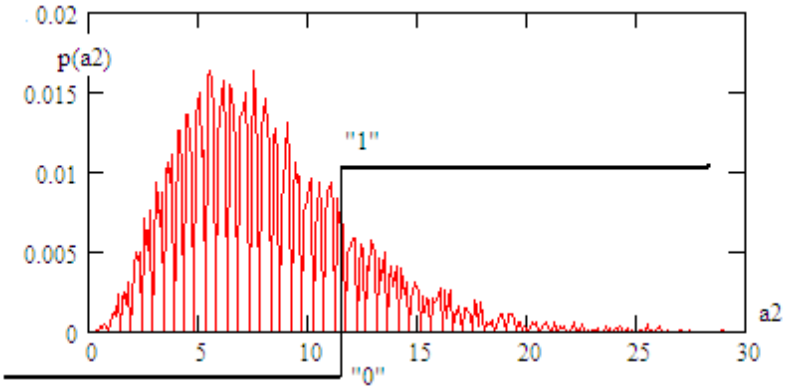


Рис. 3. Выходные состояния нейрона-a2, анализирующего число единиц в 8 фрагментах кода длиной 32 бита

Для квадратичного нейрона мода выходных состояний составляет значение $a2 \approx 6.0$. Наиболее далекие от белого шума последовательности, дают отклики сумматора нейрона со значениями $a2 > 12$. Это позволяет отбрасывать примерно 15 %, проверяемых последовательностей.

В качестве третьего теста (критерия) NIST рекомендует подсчитывать число одинаковых рядом стоящих состояний. Например, это могут быть последовательности из 1, 2, 3, …., 8 одинаковых состояний «000…0» или «111…1». Статистические моменты для этого теста близости белого шума к идеалу приведены в табл. 2.

На рис. 4 дано распределение значений выходных состояний сумматора соответствующего квадратичного нейрона с 8-ю входами.

Таблица 2

Соотношение первых двух статистических моментов белого шума длинной 256 бит для разного числа серии, обнаруженных одинаковых состояний

	Длина обнаруженных последовательностей из одинаковых состояний							
	1	2	3	4	5	6	7	8
E(.)	64	32	16	8	3.9	1.99	0.99	0.50
σ(.)	8.9	5.2	3.62	2.62	1.88	1.37	0.96	0.70
E(.)/σ(.)	7.19	6.15	4.42	3.05	2.07	1.45	1.03	0.71

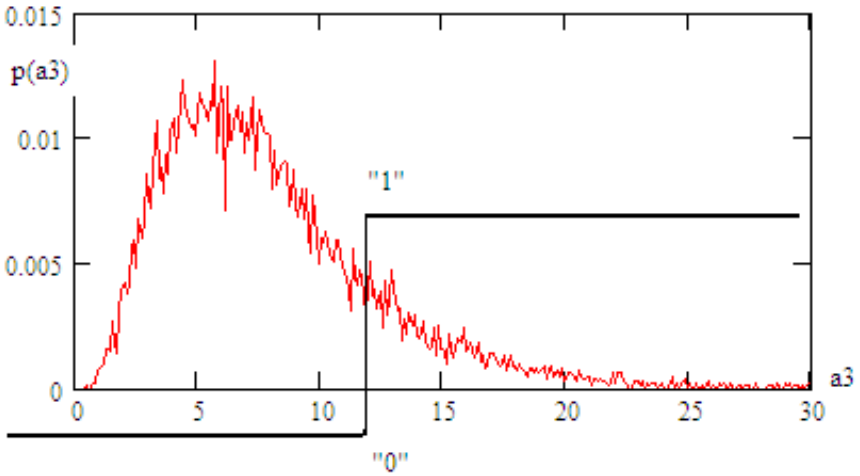


Рис. 4. Выходные состояния нейрона-a3, анализирующего частоту появления фрагментов кода с 1, 2, 3,…, 8 .одинаковыми состояниями разрядов

Если выбрать порог квантователя выходного нейрона a3 = 12, то такой нейрон будет отбрасывать (браковать) примерно 15 % случайных кодовых последовательностей.

В итоге получается, что мы построили три нейрона, каждый из которых отбраковывает примерно по 15 % реализаций белого шума, как далекие от идеала. При этом по результатам численного эксперимента мы можем оценить наблюдаемую корреляционную сцепленность выходных состояний этих трех нейронов:

$$corr(a0, a1) = 0.015 \quad corr(a0, a2) = 0.011 \quad corr(a1, a2) = 0.081.$$

В сравнении с данными о корреляционной сцепленности классических статистических критериев [7], мы получили очень хорошие результаты. Корреляционной сцепленностью такого уровня можно пренебречь, считая выходные состояния нейронов слабо зависимыми (не коррелированными).

Последнее означает, что рассматриваемое нейросетевое обобщение трех критериев NIST имеет коэффициент полезного действия $0.85 \cdot 0.85 \cdot 0.85 = 0.85^3 = 0.64$. Рассматриваемый нейросетевой классификатор бракует примерно 36 % проверяемых кодовых последовательностей.

Заключение

Любому из 16-ти тестов NIST может быть построен эквивалентный нейрон, бракующий порядка 15 % проверяемых последовательностей. То есть, построив нейросетевое обобщение всех 16-ти тестов NIST, мы сможем отбраковывать последовательности с вероятностью $1 - 0.85^{16} = 0.923$. Одним из путей дальнейшего повышения жесткости контроля, по нашему мнению, является переход к использованию нейросетевых обобщений сверток Хэмминга, вычисленных по модулям от 2 до 2048 [8, 9]. В этом случае в место 16 тестов NIST мы получаем несколько тысяч тестов, что теоретически позволяет многократно увеличивать жесткость выбраковки, в соответствии с требованиями ФСБ России, предъявляемых к качеству криптографических ключей.

Библиографический список

1. Гулов, В. П. Перспектива нейросетевой защиты облачных сервисов через биометрическое обезличивание персональной информации на примере медицинских электронных историй болезни / В. П. Гулов, А. И. Иванов, Ю. К. Язов, О. В. Корнеев // Вестник новых медицинских технологий. – 2017. – Т. 24, № 2 (июнь). – С. 220–225.

2. Безяев, А. В. Типовая схема защиты нейросетевых архивов биометрических данных не криптографическим хешированием через применение линейной рекуррентны подсчета контрольных сумм CRC-4 / А. В. Безяев, А. И. Иванов, О. В. Корнеев // Безопасность информационных технологий : тр. науч.-техн. конф. кластера пензенских предприятий, обеспечивающих безопасность информационных технологий. – Пенза, 2016. – Т. 10. – С. 15–20. – URL: http://пниэи.рф/activity/science/BIT/ T10-p15.pdf

3. Кельтон, В. Имитационное моделирование. Классика CS / В. Кельтон, А. Лоу. – 3-е изд. – Санкт-Петербург : Питер, 2004. – С. 465, 466.

4. Генератор псевдослучайных чисел // Википедия. – URL: https://ru.wikipedia.org/wiki/Генератор_псевдослучайных_чисел

5. Иванов, А. И. Среда моделирования «БиоНейроАвтограф». Программный продукт создан лабораторией биометрических и нейросетевых технологий, размещен с 2009 г. на сайте АО «ПНИЭИ» / А. И. Иванов, О. С. Захаров. – URL: http://пниэи.рф/activity/science/noc/bioneuroautograph.zip

6. Строков, А. В. Программное средство создания действительно случайных криптографических ключей из неоднозначной компоненты биометрических данных динамики рукописного почерка пользователя. / А. В. Строков, Е. И. Казанцев // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф. (г. Пенза, 24 апреля 2019 г.). – Пенза, 2019. – С. 139–144.

7. Коллекция искусственных нейронов эквивалентных статистическим критериям для их совместного применения при проверке гипотезы нормальности малых выборок биометрических данных / А. И. Иванов, А. Г. Банных, Е. Н. Куприянов, В. С. Лукин, К. А. Перфилов, К. Н. Савинов // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф. (г. Пенза, 24 апреля 2019 г.). – Пенза, 2019. – С. 156–164.

8. Юнин, А. П. Оценка качества «белого шума»: реализация теста «стаи обезьян» через множество сверток Хэмминга для разных систем счисления / А. П. Юнин, А. И. Иванов, К. А. Ратников // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф., (г. Пенза, 24 апреля 2019 г.). – Пенза, 2019. – С. 10–15.

9. Иванов, А. И. Оценка энтропии длинных кодовых слов на выходе нейросетевого преобразователя биометрии в пространствах множества сверток Хэмминга / А. И. Иванов, А. П. Юнин, М. А. Бояршинов // Интеллектуальные системы в производстве. – 2019. – Т. 17, № 2. – С. 30–36.

Для цитирования:

Юнин, А. П. Нейросетевое обобщение трех стандартных тестов контроля качества «белого шума», получаемого хешированием случайной части биометрических данных / А. П. Юнин, А. И. Иванов, А. В. Строков, С. Р. Махсудов // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 49–56.

УДК 519.2

Н. А. Иванова

ПРИБЛИЖЕННАЯ ОЦЕНКА ОШИБОК, ВОЗНИКАЮЩИХ ИЗ-ЗА МАЛОГО ОБЪЕМА ВЫБОРКИ ПРИ ВЫЧИСЛЕНИЯХ КОЭФФИЦИЕНТОВ КОРРЕЛЯЦИИ ПО ФОРМУЛЕ ПИРСОНА

Аннотация. При малых объемах выборки от 16 до 26 отсчетов двух переменных формула Пирсона дает значительные ошибки. Так как формула Пирсона часто используется, необходимо формализовать вычисление ошибок и сообщать о предсказанных значениях ее ошибок пользователям.

N. A. Ivanova

APPROXIMATE ESTIMATE OF ERRORS DUE TO THE SMALL SAMPLE SIZE IN CALCULATING PEARSON-BASED CORRELATION RATIOS

Abstract. With small sample volumes of 16 to 26 counts of two variables, Pearson's formula makes significant errors. Because Pearson formula is often used, it is necessary to formalize the calculation of errors and report the predicted values of its errors to users.

Значительный вклад в становление современной статистики внес Пирсон, предложивший в начале XX в. хи-квадрат критерий и формулу для вычисления коэффициента корреляции между двумя последовательностями случайных чисел:

$$r(x,y)=\frac{1}{n}\sum_{i=1}^{n}\frac{\left(E(x)-x_i\right)\cdot\left(E(y)-y_i\right)}{\sigma(x)\cdot\sigma(y)}, \qquad (1)$$

где $E(x), E(y)$ – математические ожидания двух переменных; $\sigma(x), \sigma(y)$ – стандартные отклонения двух случайных последовательностей.

При больших объемах выборки n = 16 000 формула Пирсона дает оценки коэффициентов корреляции с приемлемой для практики погрешностью $\Delta r \approx \pm 0.01$. К сожалению, получение выборок в несколько тысяч опытов далеко не всегда практически реализуемо. Так при обучении линейных нейронных сетей по ГОСТ Р 52633.5–2011 [1] используют от 16 до 26 примеров образа «Свой». Такой же

объем выборки предположительно будет применяться и при обучении квадратичных нейронов следующего перспективного стандарта [2]. То есть регуляризация обучения квадратичных нейронов Махаланобиса [3] и нейронов Байеса-Пирсона [4, 5] в будущем будет связана с вычислением коэффициентов корреляции на выборках в 16 опытов. К сожалению, ошибка вычисления коэффициента корреляции, при столь малой выборке, может достигать значений $\Delta r \approx \pm 0.70$. Столь значительная погрешность ослабляет эффективность операции регуляризации [3–5].

Из рис. 1 видно, что приемлемая для процедур регуляризации погрешность $\Delta r \approx \pm 0.10$ возникает с малой вероятностью 0.18. Значительно большая погрешность со значениями, попадающими в интервал от 0.1 до 0.7 по модулю возникает с вероятностью 0.82. При малой выборке в 16 опытов результаты с «плохой» точностью получаются примерно в 4 раза чаще, чем результаты с «приемлемой» точностью.

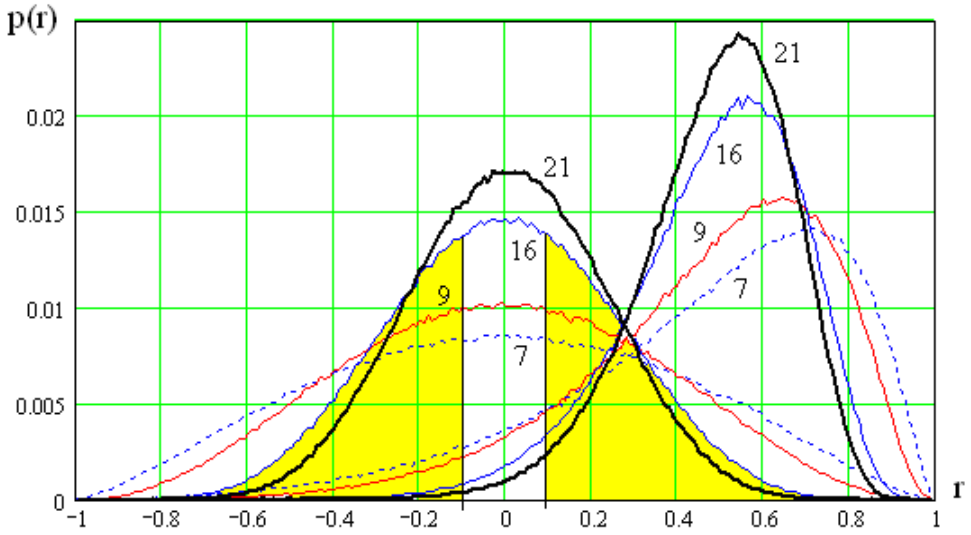


Рис. 1. Распределения значений коэффициентов корреляции при вычислении коэффициентов корреляции по формуле Пирсона для выборок из 7, 9, 16, 21 пар случайных отсчетов

Если увеличивать объем выборки, сохраняя значение коррелированности данных, то мы будем наблюдать монотонное снижение стандартного отклонения, как это показано на рис. 2.

Подобное монотонное снижение стандартного отклонения хорошо приближается гиперболой. Приближающие зависимость гиперболы на рис. 2 показаны пунктиром. Если считать все распределения значений ошибок формулы Пирсона нормальными, то ин-

тервал неопределенности близок к величине ±3σ(r). В рамках этой гипотезы интервал неопределенности ошибок вычислений описывается семейством гипербол, зависящих от двух переменных:

$$\Delta r(r,n) = \pm 3 \cdot \left(0.1 + \frac{1}{n \cdot 0.41 - 0.11}\right) \cdot \left(1 - r^2\right). \qquad (2)$$

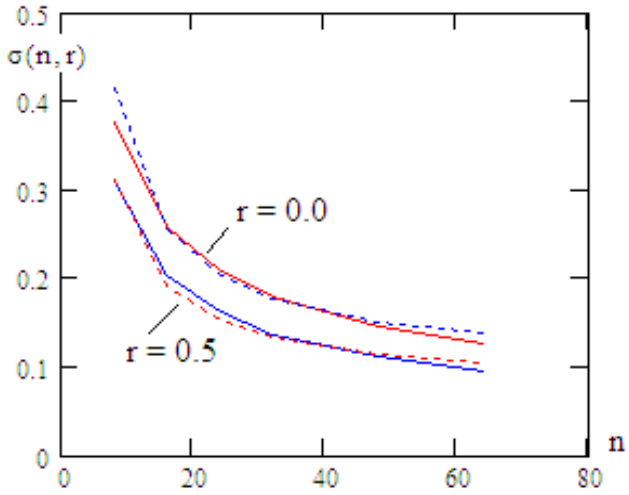


Рис. 2. Гиперболическое снижение стандартного отклонения вычисления коэффициентов корреляции по Пирсону с ростом размера выборки исходных данных

Как видно из рис. 2, приближение (2) достаточно точно описывает ситуацию для выборок объемом от 8 до 64 опытов. Опорные точки, по которым построено гиперболическое приближение (2), приведены в табл. 1.

Таблица 1

n	8	16	24	32	48	64
σ(n, r = 0.0)	0.379	0.258	0.208	0.180	0.147	0.127
σ(n, r = 0.1)	0.375	0.256	0.206	0.176	0.145	0.126
σ(n, r = 0.2)	0.368	0.247	0.201	0.170	0.140	0.121
σ(n, r = 0.3)	0.357	0.238	0.192	0.162	0.134	0.116
σ(n, r = 0.4)	0.331	0.223	0.178	0.151	0.123	0.106
σ(n, r = 0.5)	0.305	0.204	0.162	0.137	0.112	0.095
σ(n, r = 0.6)	0.269	0.173	0.139	0.117	0.096	0.082
σ(n, r = 0.7)	0.230	0.143	0.111	0.092	0.075	0.064
σ(n, r = 0.8)	0.178	0.104	0.081	0.067	0.054	0.046

Библиографический список

1. ГОСТ Р 52633.5–2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия – код доступа. – Москва : Стандартинформ, 2012. – 20 с.

2. Иванов, А. И. Второй национальный стандарт России по быстрому автоматическому обучению больших искусственных нейронных сетей на малых выборках биометрических данных / А. И. Иванов, А. В. Безяев, Е. А. Малыгина, Ю. И. Серикова // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф. (г. Пенза, 24 апреля 2019 г.). – Пенза, 2019. – С. 174–177.

3. Серикова, Ю. И. Двойная регуляризация процедур обучения нейронов Махаланобиса за счет симметризации корреляционных связей и компенсации ошибок вычисления коэффициентов парной корреляции биометрических данных / Ю. И. Серикова // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф. (г. Пенза, 24 апреля 2019 г.). – Пенза, 2019. – С. 26–34.

4. Иванов, А. И. Оценка надежности верификации автографа на основе искусственных нейронных сетей, сетей многомерных функционалов Байеса и сетей квадратичных форм / А. И. Иванов, П. С. Ложников, А. Е. Сулавко // Компьютерная оптика. – 2017. – Т. 41, № 5. – С. 765–774.

5. Ложников, П. С. Биометрическая защита гибридного документооборота / П. С. Ложников. – Новосибирск : Изд-во СО РАН, 2017. – 130 с.

Для цитирования:

Иванова, Н. А. Приближенная оценка ошибок, возникающих из-за малого объема выборки при вычислениях коэффициентов корреляции по формуле Пирсона / Н. А. Иванова // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 57–60.

УДК 519.2

И. А. Крохин

ПРОТИВОДЕЙСТВИЕ АТАКАМ МАРШАЛКО ИТЕРАЦИОННЫМ ДООБУЧЕНИЕМ НЕЙРОНОВ КАК СПОСОБ ПОВЫШЕНИЯ СТОЙКОСТИ БИОМЕТРИЧЕСКОЙ ЗАЩИТЫ ЛИЧНЫХ КРИПТОГРАФИЧЕСКИХ КЛЮЧЕЙ

Аннотация. Целью работы является анализ процесса обучения нейронной сети по стандарту ГОСТ Р 52633 для систем аутентификации пользователя на предмет возможности повышения защиты личных криптографических ключей пользователя при атаке Маршалко.

I. A. Krokhin

COUNTERING MARSHALKOATTACKS AS A WAY TO INCREASE THE DURABILITY OF BIOMETRIC PROTECTION OF PERSONAL CRYPTOGRAPHIC KEYS

Abstract. The aim of the work is to analyze the learning process of a neural network according to GOST R 52633 for user authentication systems, with a view to the possibility of increasing the protection of personal cryptographic keys of a user during an Marshalko attack.

В настоящее время, в связи с масштабной информатизацией современного общества, возникает необходимость обеспечения высоконадежной идентификации и аутентификации личности пользователя в информационном пространстве. Одно из наиболее перспективных направлений защиты личных данных, является биометрическое распознавание пользователя [2]. Данный метод аутентификации широко внедряется в повседневную жизнь, начиная от распознавания людей в потоке в ГУП «Московский метрополитен» [3] по изображению лица и идентификации клиентов ПАО «Сбербанк» по изображению лица и голосу до защиты данных пользователя, без необходимости запоминать длинные коды доступа, состоящие из последовательности случайный символов.

Один из наиболее эффективных методов безопасного хранения и восстановления личного криптографического ключа был предложен в России. Для хранения криптографических ключей ис-

пользуются нейросетевые преобразователи «биометрия-код». Преобразователь «биометрия-код» – преобразователь, способный преобразовывать вектор нечетких, неоднозначных биометрических параметров «Свой» в четкий однозначный код ключа (пароля). Преобразователь, откликающийся случайным выходным кодом на воздействие случайного входного вектора, не принадлежащего множеству образов «Свой» [4].

На рис. 1 представлена общая схема биометрической аутентификации на базе нейронной сети, описанной в стандарте ГОСТ Р 52633.

Рис. 1. Общая схема биометрической аутентификации на основе нейронной сети

Полученные биометрические данные пользователя, преобразуются в вектор биометрических параметров, после чего обрабатывается заранее обученной нейронной сетью. Параметры нейронной сети, такие как связи и веса нейронов хранятся в контейнере нейронной сети. Нейронная сеть преобразует вектор биометрических параметров в криптографический ключ, который предоставляется криптографическому протоколу аутентификации. Если данные верны, то пользователю предоставляется доступ к системе.

Блок-схема обучения одного искусственного нейрона представлена на рис. 2.

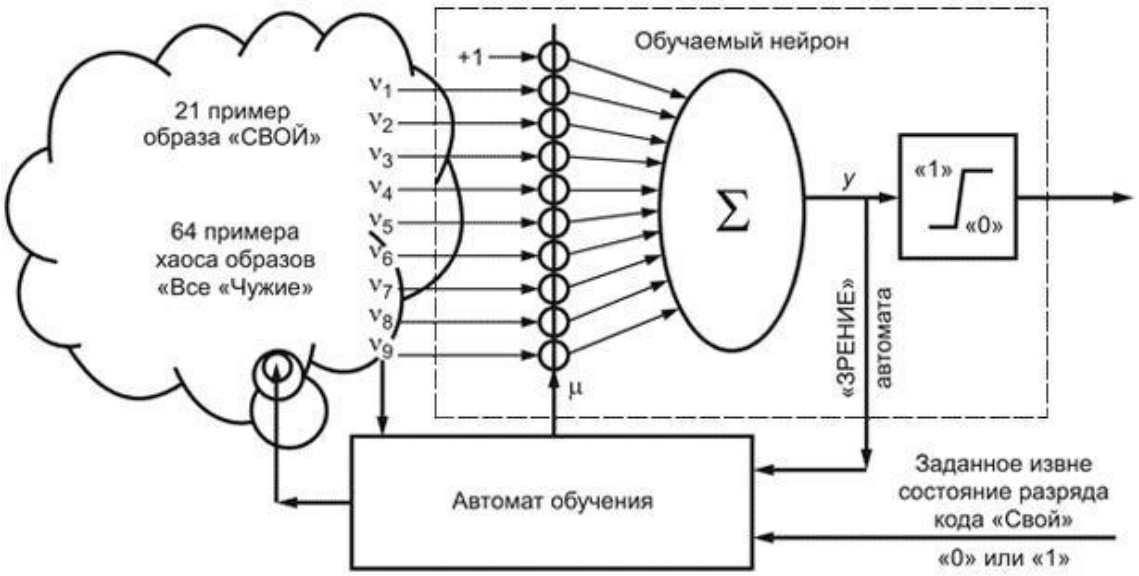


Рис. 2. Блок-схема обучения одного искусственного нейрона [4]

В ГОСТ Р 52633.5–2011 указано, что для обучения нейрона требуется не менее 21 примеров образа «Свой» и 64 независимых примеров образа «Чужой». В однослойных нейронных сетях один нейрон отвечает за один разряд выходного кода. Задача «автомата обучения» состоит в вычислении весовых коэффициентов наμ_j, которые смещают распределение откликов сумматора нейрона в заданное извне состояние его выходной нелинейности («0» или «1») [4].

Модули весов нейронов первого слоя вычисляются по формуле:

$$\mu_j = \frac{\left| \mathrm{M}_{\text{чужой}}\langle v_j \rangle - \mathrm{M}_{\text{свой}}\langle v_j \rangle \right|}{\sigma_{\text{чужой}}\langle v_j \rangle * \sigma_{\text{свой}}\langle v_j \rangle},$$

где $\mathrm{M}_{\text{свой}}\langle v_j \rangle$ – математическое ожидание значений j-го признака образа «Свой», $\sigma_{\text{свой}}\langle v_j \rangle$ – среднеквадратичное отклонение значений j-го признака образа «Свой», $\mathrm{M}_{\text{чужой}}\langle v_j \rangle$ и $\sigma_{\text{чужой}}\langle v_j \rangle$ – аналогичные показатели образа «Чужой»[4].

Выбор знака весового коэффициента осуществляется исходя из условия:

$$\begin{cases} \text{если «Свой»} \rightarrow \text{«1»}, \text{то } sign(\mu_j) = sign\left(\mathrm{M}_{\text{свой}}(v_j) - \mathrm{M}_{\text{чужой}}(v_j) \right) \\ \text{если «Свой»} \rightarrow \text{«0»}, \text{то } sign(\mu_j) = -sign\left(\mathrm{M}_{\text{свой}}(v_j) - \mathrm{M}_{\text{чужой}}(v_j) \right) \end{cases}$$

Иными словами, если для входного образа «Свой» нейрон настроен на выход единицы, при этом $M_{свой}\langle v_j \rangle > M_{чужой}\langle v_j \rangle$, выбирается положительный знак весового коэффициента «+», если условие не выполняется «–». Иначе, если нейрон настроен на нуль, знаки инвертируются.

Выход сумматора нейрона на этапе верификации, рассчитывается по формуле:

$$y_i = \sum_{j=1}^{m} \mu_j * v_j + \mu_0,$$

где v_j, – значение j-го входа i-го нейрона, ассоциированного с одним из признаков, m – число входов нейрона, μ_j – весовой коэффициент j-го входа i-го нейрона, μ_0 – нулевой вес [4].

Так же в стандарте определяются следующие угрозы безопасности:

– Возможность угадать входной вектор биометрических параметров.

– Получение секретной информации из структуры нейронной сети, таких как веса нейронов и таблица связей.

– Возможность угадать личный криптографический ключ.

В статье [1] описаны варианты атаки, подтверждающие, что методы защиты нейронных сетей, предложенные в стандарте, являются недостаточными. Так же автор приходит к выводу, что веса нейронов необходимо считать частью личного криптографического ключа в системе аутентификации.

Схема биометрической аутентификации в стандарте ГОСТ Р 52633, основанная на нейронной сети использует преобразование содержащие всю информацию об биометрических данных и секретном ключе пользователя. Поскольку нейронная сеть описывается системой линейных неравенств, сложность нахождения секретного ключа по известным таблицам весов нейронов эквивалентна сложности нахождения одного решения соответствующей системы линейных неравенств [1].

Атака Маршалко, строится на наблюдении большого числа выходов у незащищенных нейронов и поиска общих связей. Данная атака сокращает ключ с 30–50 бит до 15 бит. Схема нахождения общих связей нейронов при обучении по ГОСТ Р 52633 представлена на рис. 3.

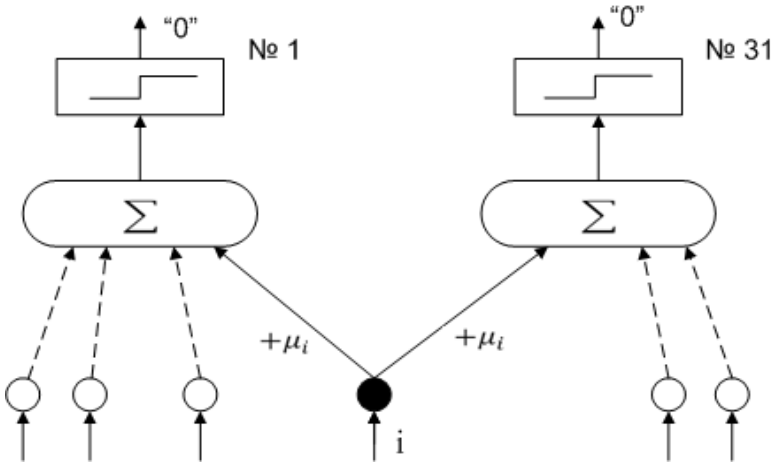


Рис. 3. Схема нахождения общих связей

Для защиты нейронной сети, ее необходимо обучить так, чтобы при атаке Маршалко не предоставлялось возможным точно определить общие связи. Атака работает если можно подобрать пары со 100 % вероятностью, если уменьшить вероятность до 90 % атака условно работает, на менее 50 %, атака не работает совсем, однако не стоит забывать о качестве работы преобразователя «биометрия-код».

Добиться повышения стойкости к атакам Маршалко возможно путем обычного итерационного дообучения нейронов с общими связями. При этом знак общей связи может быть принудительно изменен, тогда качество решения задачи нейроном падает, необходимо поднимать уровень качества за счет обычного итерационного дообучения нейрона с сохранением одного измененного знака.

Необходимо средствами имитационного моделирования оценить на сколько итерационное дообучение нейрона эффективно защищает знак общих связей.

Кроме того, если итерационно дообучить все нейроны сети, что должен появиться эффект маскирования модулей значений весовых коэффициентов нейронов.

Все эти меры являются предпосылками снижения эффективности атаки Маршалко и их эффективность должна быть подтверждена статистически.

Библиографический список

1. Marshalko, G. B. On the security of a neural network-based biometric authentication scheme / G. B. Marshalko // Математические вопросы криптографии. – 2014. – Т. 5, № 2. – С. 87–98.

2. Технология использования больших нейронных сетей для преобразования нечетких биометрических данных в код ключа доступа : монография / Б. Б. Ахметов, А. И. Иванов, В. А. Фунтиков, А. В. Безяев, Е. А. Малыгина. – Алматы : Изд-во LEM, 2014. – 144 с.

3. В Московском метрополитене запустили систему распознавания лиц. – 2020. – URL: https://mosmetro.ru/press/news/2722/

4. ГОСТ Р 52633.5–2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрии. – Москва : Стандартинформ, 2012. – 20 с.

5. Иванов, А. И. Оценка надежности верификации автографа на основе искусственных нейронных сетей, сетей многомерных функционалов Байеса и сетей квадратичных форм / А. И. Иванов, П. С. Ложников, А. Е. Сулавко // Компьютерная оптика. – 2017. – № 5 (41). – С. 765–774.

6. Иванов, А. И. Нейросетевая защита конфиденциальных биометрических образов гражданина и его личных криптографических ключей / А. И. Иванов. – Пенза : ПНИЭИ, 2014. – 57 с.

7. Сбербанк присоединился к проекту Центрального Банка по биометрической идентификации клиентов. – 2020. – URL: https://www.sberbank.ru/ru/press_center/all/article?newsID=7d31d4b3-6deb-4732-b5eb-37a2627d2bb5&blockID=1303&regionID=77&lang=ru/

Крохин, И. А. Противодействие атакам Маршалко итерационным дообучением нейронов как способ повышения стойкости биометрической защиты личных криптографических ключей / И. А. Крохин // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 61–66.

УДК 519.2

Е. Н. Куприянов, А. И. Иванов

ОРТОГОНАЛИЗАЦИЯ СТАТИСТИКО-НЕЙРОСЕТЕВОГО АНАЛИЗА МАЛЫХ ВЫБОРОК БИОМЕТРИЧЕСКИХ ДАННЫХ НА ПРИМЕРЕ ИСПОЛЬЗОВАНИЯ НЕЙРОНОВ ЛЕЖАНДРА В ПЕРВОМ СЛОЕ ДВУХСЛОЙНОЙ СЕТИ ИСКУССТВЕННЫХ НЕЙРОНОВ

Аннотация. Статистический нейросетевой анализ биометрических данных рассматривается как некоторое обобщение множества статистических критериев, созданных в прошлом веке. Отмечается, что наиболее изученными являются нейроны, обогащающие данные в линейном пространстве (персептроны) и нейроны, обогащающие входные данные в квадратичном пространстве. Нейросетевое обобщение классических статистических критериев обычно дает сильную корреляционную сцепленность выходных состояний нейронной сети. Одним из путей повышение мощности нейросетевых обобщений является их ортогонализация, например применением полиномиальных нейронов Лежандра в первом слое двухслойной сети искусственных нейронов.

E. N. Kupriyanov, A. I. Ivanov

ORTHOGONALIZATION OF STATISTICAL-NEURAL NETWORK ANALYSIS OF SMALL SAMPLES OF BIOMETRIC DATA ON THE EXAMPLE OF THE USE OF LEANDRE NEURONS IN THE FIRST LAYER OF THE TWO-LAYER NETWORK OF ARTIFICIAL NEURONS

Abstract. In the article, statistical neural network analysis of biometric data is considered as some generalization of many statistical criteria created in the last century. It is noted that the most studied are neurons enriching data in linear space (perseptrons) and neurons enriching input data in the square space. Neuronet summarization of classical statistical criteria usually gives a strong correlation of output states of the neural network. One way to increase the power of neural network generalizations is to orthogonize them, for example, by using Lezhander's polynomial neurons in the first layer of the two-layer network of artificial neurons.

Проблема статистического анализа малых выборок биометрических данных

Практика нейросетевой обработки малых выборок биометрических данных показала, что распределение биометрических пара-

метров образа «Свой» и образа «Чужой» описывается разными законами (хи-квадрат, нормальным, равномерным, бета распределением, …). Проблема состоит в том, что выборки биометрических данных малы от 16 до 21 состояний. Это не позволяет надежно оценивать соответствующие статистические гипотезы, применяя классические статистические критерии. Например, по стандартной методике [1, 2] проверку гипотезы нормальности с доверительной вероятностью 0.99 по критерию хи-квадрат можно выполнить только для выборок размером в 200 и более опытов.

Ослабить проблему малых выборок удается, если проверку гипотез выполнять параллельно, используя несколько статистических критериев, например, построив для каждого из классических статистических критериев эквивалентный ему искусственный нейрон [3–6], как это отображено на рис. 1.

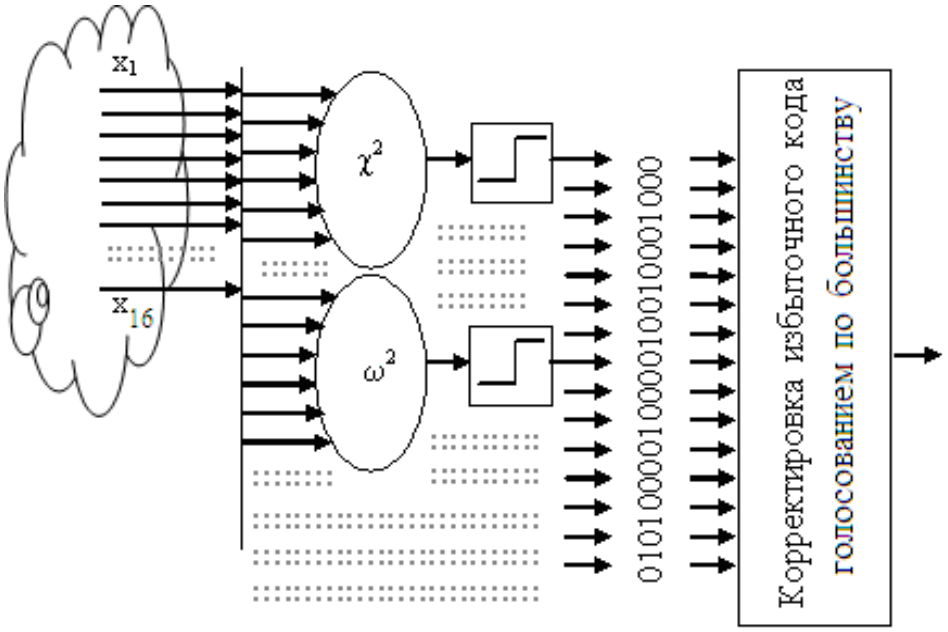


Рис. 1. Нейросетевое объединение множества известных статистических критериев, созданных в прошлом веке для проверки гипотезы нормальности

Такой подход к решению задачи был бы весьма и весьма эффективен, если бы выходные состояния искусственных нейронов были бы независимы (не коррелированны). К сожалению, этого не происходит. В табл. 1 даны данные о корреляционной связанности решений ряда классических статистических критериев.

Таблица 1

Коэффициенты корреляции между откликами классических статистических критериев

corr(χ2,KfM) = 0.478	corr(χ2,Fr) = 0.421 corr(KfM,Fr) = 0.94	corr(χ2,AD) = 0.357 corr(KfM,AD) = 0.654 corr(Fr,AD) = 0.608	corr(χ2,ADL) = 0.648 corr(KfM,ADL) = 0.831 corr(Fr,ADL) = 0.753 corr(AD,ADL) = 0.686

где χ^2- хи-квадрат критерий;
KfM - критерий Крамера фон-Мизеса;
Fr - критерий Фроцини;
AD - критерий Андерсона-Дарлинга;
ADL -логарифмическая форма AD критерия.

В итоге, нейросетевое обобщение множества статистических критериев рис. 1 работает, однако его эффективность была бы выше, если бы удалось снизить коэффициенты корреляции между параллельно работающими нейронами.

Двухслойная сеть искусственных нейронов, построенных с использованием ортогональных полиномов Лежандра в первом слое

Одним из известных способов устранения корреляционных связей является ортогонализация решаемой задачи. В частности, этим способом воспользовались Нейман и Девид, создавая свой статистический критерий [2]. Схема нейросетевой реализации их статистического критерия представлена на рис. 2.

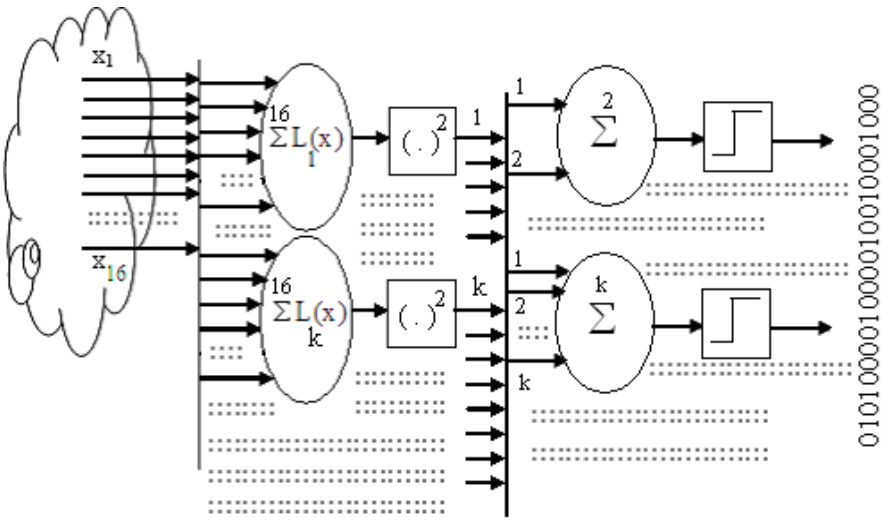


Рис. 2. Двухслойная сеть искусственных нейронов Неймана для выборки, состоящей из 16 опытов

При реализации полиномиальных нейронов первого слоя нейросети используются ортогональные полиномы Лежандра:

$$\begin{cases} L_1(x) = x; \\ L_2(x) = (3x^2 - 1)/2; \\ L_3(x) = (5x^3 - 3x)/2; \\ L_4(x) = (35x^4 - 30x^2 + 3)/8; \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \end{cases} \quad (1)$$

Так как полиномы Лежандра ортогональны на интервале от –1 до +1 необходимо выполнить приведение данных малой выборки к этому интервалу. Программная реализация нормировки дана на рис. 3.

```
x00 := | x00 ← sort(rbeta(16,6,6))
       | x00 ← x00 − x00_0
       | x00 ← 2x00 / x00_15 − 1
       | x00
```

$x00^T =$

	0	1	2	3	4	5	6	7
0	-1	-0.984	-0.798	-0.67	-0.651	-0.566	-0.381	-0.324

$x00^T =$

	4	5	6	7	8	9	10	11	12	13	14	15
0	-0.651	-0.566	-0.381	-0.324	-0.02	0.111	0.139	0.169	0.256	0.332	0.901	1

Рис. 3. Программная реализация процедуры приведения данных малой выборки к интервалу ортогональности полиномов Лежандра

Программная реализация каждого из полиномиальных нейронов Лежандра сводится к деформированию пространства накопления данных малой выборки одним из полиномов Лежандра (1). Далее все отсчеты выборки, деформированные тем или иным полиномом Лежандра, суммируются. Выходное состояние сумматора возводится в квадрат.

Второй слой нейронов выполняет суммирование 2, 3,.., k полиномиальных нейронов Лежандра, получая тем самым (k–1) статистику Неймана-Девида. В справочнике [2] на странице 333 даны значения порогов (квантилей доверительных вероятностей) для рассматриваемых статистик.

Заключение

На сегодняшний день хорошо изучены нейроны с накоплением данных в линейном пространстве (персептроны). Так же хорошо исследованы нейроны, осуществляющие обогащение данных в

квадратичных пространствах. Предположительно свойства полиномиальных ортогональных нейронов Лежандра так же должны быть исследованы применительно к биометрии в ближайшее время. Класс полиномиальных ортогональных нейронов фактически был создан в прошлом веке Нейманом, Девидом. На тот момент выделять математические конструкции в виде комбинации сумматоров и нелинейных преобразований, называя их искусственными нейронами было не принято. Интересно отметить, что идея ортогонализации преобразований была опубликована в 1937 г. [7] Нейманом, а второй слой нейронов появился в 1939 г. усилиями Девида [8].

Следует так же заметить, что ортогональные преобразования статистик Лежандра, выполненные под контроль выборок с равномерным распределением, оказались в прошлом веке единственными. Иные ортогональные полиномы Якоби, Чебышева, Эрмита, Лагерра никто не рассматривал. Очевидным является то, что замена полиномов Лежандра на любые другие ортогональные полиному приведет, как и в нашем случае к двухслойному нейросетевому аналогу. Все это должно стать предметом исследований в ближайшее время. Видимо ортогональные полиномиальные искусственные нейроны должны занять свое место среди других классов искусственных нейронов.

Библиографический список

1. Р 50.1.037–2002. Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Ч. I. Критерии типа $\chi 2$. Госстандарт России. – Москва, 2001. – 140 с.

2. Кобзарь, А. И. Прикладная математическая статистика. Для инженеров и научных работников / А. И. Кобзарь. – Москва : ФИЗМАТЛИТ, 2006. – 816 с.

3. Коллекция искусственных нейронов, эквивалентных статистическим критериям для их совместного применения при проверке гипотезы нормальности малых выборок биометрических данных / А. И. Иванов, А. Г. Банных, Е. Н. Куприянов, В. С. Лукин, К. А. Перфилов, К. Н. Савинов // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф. (г. Пенза, 24 апреля 2019 г.). – Пенза, 2019. – С. 156–164.

4. Иванов, А. И. Нейросетевое обобщение классических статистических критериев для обработки малых выборок биометрических данных / А. И. Иванов, Е. Н. Куприянов, С. В. Туреев // Надежность. – 2019. – № 2. – С. 22–27. – DOI 10.21683/1729-2646-2019-19-2-22-27.

5. Волчихин, В. И. Нейросетевой анализ малых выборок биометрических данных с использованием хи-квадрат критерия и критериев Андерсона –

Дарлинга / В. И. Волчихин, А. И. Иванов, А. В. Безяев, Е. Н. Куприянов // Инженерные технологии и системы. – 2019. – Т. 29, № 2. – С. 205–217. – DOI https: //doi.org/10.15507/2658-4123.029/2019.02.205-217.

6. Иванов, А. И. Нейросетевое обобщение семейства статистических критериев среднего геометрического и среднего гармонического для прецизионного анализа малых выборок биометрических данных / А. И. Иванов, К. А. Перфилов, В. С. Лукин // Информационно-управляющие телекоммуникационные системы, средства поражения и их техническое обеспечение : сб. науч. ст. Всерос. науч.-техн. конф. / под общ. ред. В. С. Безяева. – Пенза : АО «НПП "Рубин"», 2019. – С. 50–63.

7. Neyman, J. “Smooth” test for goodness-of-fit / J. Neyman // Scand. Aktuarietidsrift. – 1937. – Vol. 20. – P. 149–169.

8. David, F. N. On Neymans “smooth” test for goodness-of-fit / F. N. David // Biometrika. – 1939. – Vol. 31. – P. 191–199.

Для цитирования:

Куприянов, Е. Н. Ортогонализация статистико-нейросетевого анализа малых выборок биометрических данных на примере использования нейронов Лежандра в первом слое двухслойной сети искусственных нейронов / Е. Н. Куприянов, А. И. Иванов // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 67–72.

УДК 519.2

А. И. Иванов, К. А. Ратников

ИСПОЛЬЗОВАНИЕ СПЕКТРА КОЭФФИЦИЕНТОВ КОРРЕЛЯЦИИ ХЭММИНГА ДЛЯ КОРРЕКТНОГО СТАТИСТИЧЕСКОГО ОПИСАНИЯ ВЫХОДНЫХ КОДОВ НЕЙРОСЕТЕВЫХ МОЛЕКУЛ

Аннотация. Практикуемое сегодня вычисление показателя корреляционной сцепленности разрядов кода нейросетевой молекулы через коэффициенты корреляции Пирсона может быть дополнено оценкой парной корреляции разрядов через вычисление сверток Хэмминга. Дан пример расхождения двух видов оценок, корреляционной сцепленности для нейросетевой молекулы, обобщающей пять классических статистических критериев. Сделано предположение, что два вида оценки коэффициентов парной корреляции дополняют друг друга и позднее могут быть использованы для повышения точности предсказания энтропии выходных кодов нейросетевых молекул.

A. I. Ivanov, K. A. Ratnikov

USING THE HEMMING CORRELATION SPECTRUM TO CORRECTLY STATISTICALLY DESCRIBE THE OUTPUT CODES OF NEURAL NETWORK MOLECULES

Abstract. The current calculation of the correlation of neural network molecule code discharges through Pearson correlation ratios can be supplemented by an assessment of the pair correlation of discharges through the calculation of Hemming's bundles. An example of the divergence between the two types of assessments, correlational clutch for the neural network molecule, summarizes five classical statistical criteria. It has been suggested that two types of pair correlation scores complement each other and can later be used to improve the accuracy of the entropy prediction of output codes of neural network molecules.

Проблема нейросетевого обобщения классических статистических критериев проверки гипотезы нормальности на малых выборках

Классический статистический χ^2 критерий дает надежные оценки для выборок в 200 и более опытов. Если применять этот критерий для выборки в 16 опытов, то решение по разделению равномерных и нормальных данных принимается с низкой доверительной вероятностью – 0.65. Эта ситуация отображена на рис. 1.

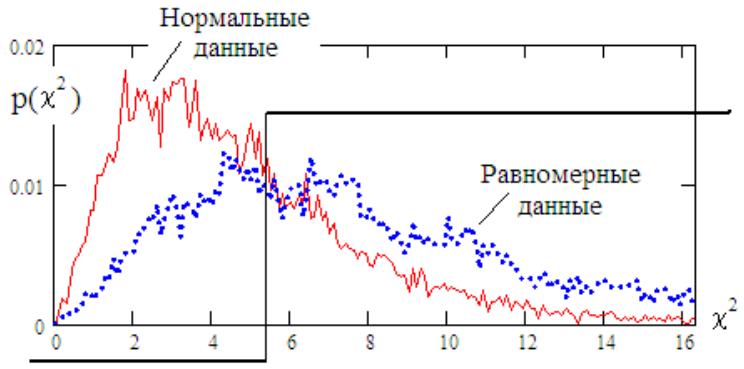


Рис. 1. Использование хи-квадрат критерия для разделения выборок в 16 опытов нормальных данных и равномерных данных

Так как доверительная вероятность принятия нейросетевого решения по χ^2 критерию недостаточна, необходимо добавить еще один статистический критерий, например, критерий Крамера-фон Мизеса или омега-квадрат критерий – ω^2. На рис. 2 приведена иллюстрация работы омега квадрат критерия. Использование критерия Крамера-фон Мизеса дает достаточно низкую доверительную вероятность – 0.599. Получается, что χ^2 критерий и ω^2 критерий сопоставимы по их качеству и, если следовать господствующей сегодня идеологии выбора наиболее информативных компонент (так называемая бритва Оккама), то мы должны отбросить один из двух почти одинаковых параметра.

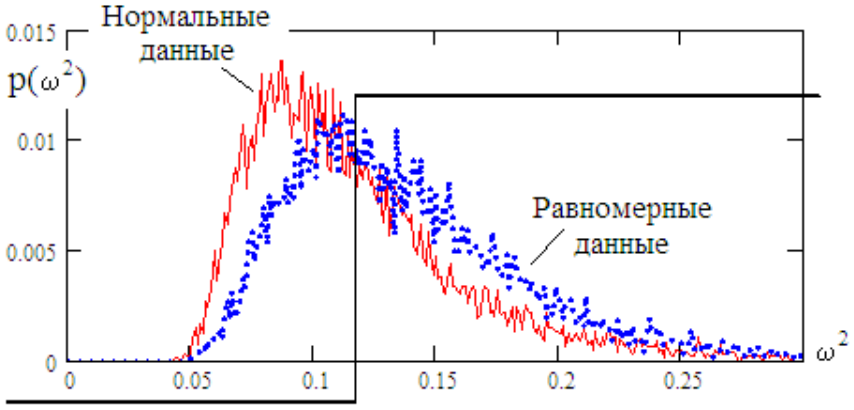


Рис. 2. Использование ω^2-квадрат критерия для разделения выборок в 16 опытов нормальных данных и равномерных данных

На самом деле бритва Оккама – это пережиток средневековья (Оккам – это английский монах XII в., занимавшийся проблемой числа ангелов способных разместиться на острие одной иглы). Сопоставимые по качеству параметры (в частности биометрические

параметры) нельзя отбрасывать, надо их совместно использовать. В частности – это нормальная практика нейросетевой биометрии, выполненной в соответствии с пакетом из 7 национальных биометрических стандартов России серии ГОСТ Р 52633.xx-20xx.

Нейросетевое обобщение большого числа классических статистических критериев

Теоретически под любой статистический критерий можно построить эквивалентный ему искусственный нейрон [1–6]. Обучение каждого из нейронов выполняется путем обнаружения точки совпадения вероятностей ошибок первого и второго рода $P_1 \approx P_2 \approx P_{EE}$. Точка равновероятных ошибок первого и второго рода фактически является точкой, где должен срабатывать каждый из квантователей нейрона, эмитирующего конкретный статистический критерий. То есть, задача настройки (обучения) множества искусственных нейронов, воспроизводящих работу множества статистических критериев должна решаться однократно.

Формально в XX в. было создано порядка 200 разных статистических критериев [7], то есть мы имеем техническую возможность в место использования одного статистического χ^2 критерия или одного ω^2 критерия использовать параллельно порядка 200 разных статистических критериев. Схема параллельного использования множества статистических критериев приведена на рис. 3.

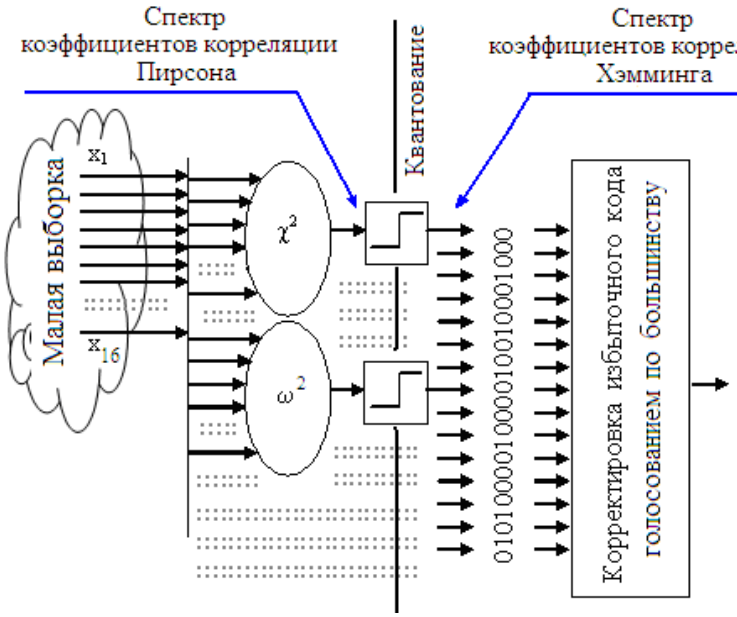


Рис. 3. Нейросетевая статистическая молекула, полученная обобщением множества классических статистических критериев

Как видно из рисунка при параллельном использовании 200 статистических критериев возникает 200 кратная кодовая избыточность, которую необходимо устранить. Самым простым способом является устранение кодовой избыточности по большинству кодовых состояний. В идеале, когда все нейроны принимают решение о наблюдении нормально распределенных данных в малой выборке, то мы должны наблюдать выходной длинный код, состоящий только из нулей «0000…..000». И наоборот, когда мы подадим на входы нейронной сети выборку из 16-ти равномерно распределенных данных, то мы в идеале должны получить выходной код, состоящий только из единиц «1111….111».

Все это означает, что мы можем построить простейший кодовый корректор, который работает, подсчитывая число нулей в коде. Если число нулей больше чем единиц, то принимается итоговое решение «0». При большинстве в выходном коде нейросети единиц, итоговое решение нейросети должно быть «1».

Нейросетевое объединения пяти классических статистических критериев

За период с 2017 по 2020 г. проведены работы по нейросетевому объединению примерно 16 разных статистических критериев [1–3]. Проиллюстрируем разработанную технологию на примере нейросетевого обобщения пяти следующих общеизвестных статистических критериев:

1. χ^2 – статистический хи-квадрат критерий Пирсона (1900);
2. ω^2 – статистический критерий Крамера-фон Мизеса (1928);
3. SKfM – статистический критерий Смирнова-Крамера-фон Мизеса (1936);
4. AD – статистический критерий Андерсона-Дарлинга (1952);
5. MT – статистический критерий Муроты-Такеучи (1982).

Следует отметить, что одной из основных проблем нейросетевого объединения статистических критериев является их существенные взаимные корреляционные связи. Данные о коэффициентах корреляции, пяти рассматриваемых критериев приведены в табл. 1.

Таблица 1

$P_1 \approx P_2 \approx P_{EE}$	Критерии	χ^2	ω^2	SKfM	AD	MT
0.350	χ^2	**1.0**	0.514	0.019	0.195	0.148
0.401	ω^2	0.514	**1.0**	0.028	0.079	0.380
0.04	SKfM	0.019	0.028	**1.0**	–0.112	–0.589
0.469	AD	0.195	0.079	–0.112	**1.0**	–0.374
0.155	MT	0.148	0.380	–0.589	–0.374	**1.0**

Если бы корреляция отсутствовала, то прогнозирование вероятности итогового результата работы нейросети было бы очень простой задачей. Для этой цели достаточно было бы перемножить между собой равновероятные ошибки первого и второго рода (данные первого столбца таблицы 1). В итоге мы бы получили очень оптимистичную результирующую оценку:

$$\tilde{P}_{EE} \approx 0.35 \cdot 0.401 \cdot 0.04 \cdot 0.469 \cdot 0.155 = 0.0004. \qquad (1)$$

Крайне оптимистичная оценка (1) показывает то, как важно, стремиться к устранению корреляционных связей между откликами искусственных нейронов, например, через их ортогонализацию. С другой стороны, эта же оценка показывает необходимость корректного учета влияния корреляционных связей, например, через процедуру их симметризации [8].

Заметим, что выходные коды, рассматриваемой нейросетевой молекулы, могут иметь $2^5 = 32$ состояния. Если же мы в будущем сможем построить нейросетевую молекулу с 200 нейронами, то нам придется иметь дело с анализом кодов с 2200 состояниями, что технически сложно. Упростить задачу удается, если перейти от анализа обычных кодов к анализу расстояний Хэмминга (сверток Хемминга) всех кодов с их предельными значениями «00000» и «11111».

$$\begin{cases} h("00000") = \sum_{i=0}^{4} ("x_i") \oplus ("0") = 5 \cdot E\{("x_i") \oplus ("0")\} \\ h("11111") = \sum_{i=0}^{4} ("x_i") \oplus ("1") = 5 \cdot E\{("x_i") \oplus ("1")\} \end{cases}. \qquad (2)$$

В этом случае спектр рассматриваемой молекулы из 5 нейронов будет иметь всего шесть спектральных линий, как это показано на рис. 4.

Важно отметить то, что, заранее зная амплитуды вероятности спектральных линий сверток Хемминга, мы можем сообщать пользователю не только принятое нейронной сетью решение, но и доверительную вероятность к этому решению. Так если мы обнаружим выходное состояние «00000», то гипотеза нормальности выборки подтверждается с доверительной вероятностью 0.9941. Если в коде обнаружена одна единица, нормальное распределение подтверждается с доверительной вероятностью 0.9724. Обнаружение двух единиц дает доверительную вероятность 0.8738. Обнаружение трех единиц приводит к отвержению гипотезы нормального распределения в пользу гипотезы равномерного распределения с доверительной вероятностью 0.421.

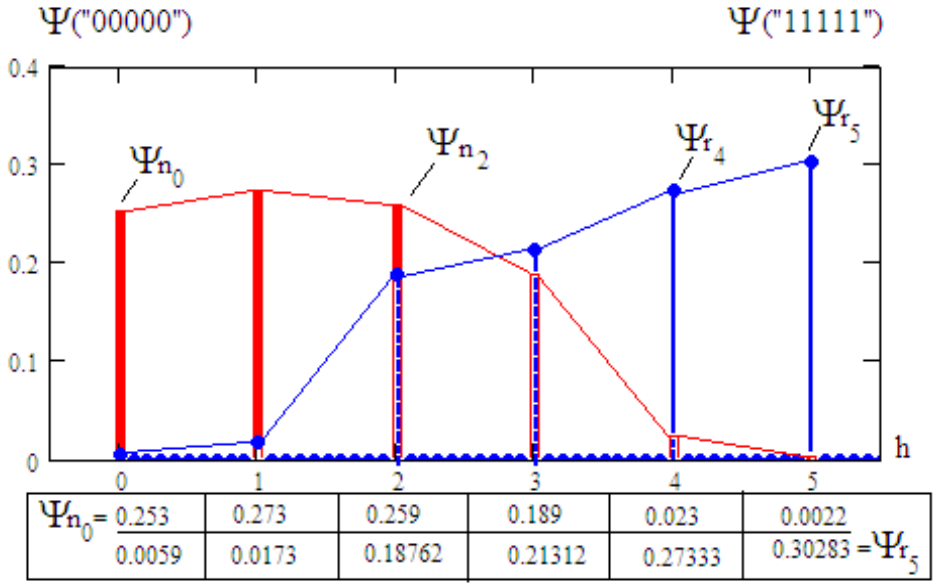


Рис. 4. Спектр линий расстояний Хэмминга для нормальных данных –$\Psi n_{(h)}$ и равномерных данных – $\Psi r_{(h)}$, а также их огибающие

Таким образом, переход в пространство сверток Хэмминга дает не только упрощение задачи, но открывает дополнительные полезные возможности статистического анализа.

Оценка показателей корреляционной сцепленности выходных кодовых состояний нейросетевой молекулы

По аналогии с коэффициентами корреляции Пирсона в непрерывном пространстве построим коэффициенты корреляции Хэмминга в дискретном пространстве выходных кодов нейросетевой молекулы:

$$\begin{cases} r("x_i","x_k") = \dfrac{1}{n}\cdot\sum_{i=0}^{n}("x_i")\oplus("x_k") = E\{("x_i")\oplus("x_k")\} \\ r("x_i",\neg"x_k") = \dfrac{1}{n}\cdot\sum_{i=0}^{n}("x_i")\oplus\neg("x_k") = E\{("x_i")\oplus\neg("x_k")\} \end{cases} \quad (3),$$

где n – число принятых во внимание выходных кодовых состояний, исследуемой нейросетевой молекулы, $\neg$ – символ операции инвертирования ($\neg$«0»= «1»; $\neg$«1»= «0»).

Проведенный численный эксперимент показал, что первое уравнение системы (3) дает корреляционную матрицу с нулевыми элементами на диагонали. Результаты вычислений приведены в табл. 2.

Таблица 2

$P_1 \approx P_2 \approx P_{EE}$	Критерии	«χ^2»	«ω^2»	«SKfM»	«AD»	«MT»
0.350	«χ^2»	**0.0**	0.317	0.382	0.374	0.433
0.401	«ω^2»	0.317	**0.0**	0.403	0.413	0.499
0.04	«SKfM»	0.382	0.403	**0.0**	0.496	0.151
0.469	«AD»	0.374	0.413	0.496	**0.0**	0.406
0.155	«MT»	0.433	0.499	0.151	0.406	**0.0**

Результаты вычислений по второму уравнению системы (3) приведены в табл. 3.

Таблица 3

$P_1 \approx P_2 \approx P_{EE}$	Критерии	«χ^2»	«ω^2»	«SKfM»	«AD»	«MT»
0.350	«χ^2»	**1.0**	0.683	0.618	0.626	0.567
0.401	«ω^2»	0.683	**1.0**	0.597	0.588	0.501
0.04	«SKfM»	0.618	0.597	**1.0**	0.504	0.849
0.469	«AD»	0.626	0.588	0.504	**1.0**	0.594
0.155	«MT»	0.567	0.501	0.849	0.594	**1.0**

Новые преобразования (3) дают значения коэффициентов корреляции Хэмминга, попадающие в интервал от 0 до 1. Это не привычно, так как коэффициенты корреляции Пирсона изменяются в интервале от –1 до +1. Люди привыкли к коэффициентам корреляции Пирсона, в связи с этим выполним масштабное преобразование коэффициентов корреляции Хэмминга, приведя их к привычному интервалу вариаций:

$$\begin{cases} rr("x_i","x_k") = 2 \cdot \left\{ E\left\{ ("x_i") \oplus ("x_k") \right\} - 0.5 \right\} \\ rr("x_i",\neg"x_k") = 2 \cdot \left\{ E\left\{ ("x_i") \oplus \neg("x_k") \right\} - 0.5 \right\} \end{cases}. \qquad (4)$$

Данные, соответствующие обычному интервалу вариаций коэффициентов корреляции приведены в табл. 4 и 5. В обоих табл. 4 и 5 модули коэффициентов корреляции одинаковые, а знаки у коэффициентов противоположные:

$$r("x_i","x_k") = -r("x_i",\neg"x_k") = r(\neg"x_i","x_k"). \qquad (5)$$

Таблица 4

$P_1 \approx P_2 \approx P_{EE}$	Критерии	«χ^2»	«ω^2»	«SKfM»	«AD»	«MT»
0.350	«χ^2»	**–1.0**	0.366	0.235	0.252	0.135
0.401	«ω^2»	0.366	**–1.0**	0.194	0.175	0.002
0.04	«SKfM»	0.235	0.194	**–1.0**	0.0083	0.698
0.469	«AD»	0.252	0.175	0.0083	**–1.0**	0.188
0.155	«MT»	0.135	0.002	0.698	0.188	**–1.0**

Таблица 5

$P_1 \approx P_2 \approx P_{EE}$	Критерии	«χ^2»	«ω^2»	«SKfM»	«AD»	«MT»
0.350	«χ^2»	**1.0**	–0.366	–0.235	–0.252	–0.135
0.401	«ω^2»	–0.366	**1.0**	–0.194	–0.175	–0.002
0.04	«SKfM»	–0.235	–0.194	**1.0**	–0.0083	–0.698
0.469	«AD»	–0.252	–0.175	–0.0083	**1.0**	–0.188
0.155	«MT»	–0.135	–0.002	–0.698	–0.188	**1.0**

Обе рассмотренные формы вычисления спектра коэффициентов корреляции Хэмминга равнозначны. Получается, что параллельно с привычными всем корреляционными матрицами Пирсона в дискретном пространстве кодов нейросетевой молекулы, обобщающей статистические критерии, могут быть построены аналогичные по физическому смыслу корреляционные матрицы Хэмминга. Видимо, корреляционные матрицы Пирсона и корреляционные матрицы Хэмминга дополняют друг друга и их совместное использование позволит увеличить достоверность статистического описания нейросетевых конструкций.

Библиографический список

1. Коллекция искусственных нейронов, эквивалентных статистическим критериям для их совместного применения при проверке гипотезы нормальности малых выборок биометрических данных / А. И. Иванов, А. Г. Банных, Е. Н. Куприянов, В. С. Лукин, К. А. Перфилов, К. Н. Савинов // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф. (г. Пенза, 24 апреля 2019 г.). – Пенза, 2019. – С. 156–164.

2. Иванов, А. И. Искусственный нейрон для контроля по критерию вариаций коэффициентов эксцесса малых выборок биометрических данных с нормальным распределением / А. И. Иванов, А. Г. Банных // Информационно-управляющие телекоммуникационные системы, средства поражения и их техническое обеспечение : сб. науч. ст. Всерос. науч.-техн. конф. / под общ. ред. В. С. Безяева. – Пенза : АО «НПП "Рубин"», 2019 . – С. 84–94.

3. Иванов, А. И. Нейросетевое обобщение семейства статистических критериев среднего геометрического и среднего гармонического для прецизионного анализа малых выборок биометрических данных / А. И. Иванов, К. А. Перфилов, В. С. Лукин // Информационно-управляющие телекоммуникационные системы, средства поражения и их техническое обеспечение : сб. науч. ст. Всерос. науч.-техн. конф. / под общ. ред. В. С. Безяева. – Пенза : АО «НПП "Рубин"», 2019. – С. 50–63.

4. Волчихин, В. И. Нейросетевой анализ малых выборок биометрических данных с использованием хи-квадрат критерия и критериев Андерсона – Дарлинга / В. И. Волчихин, А. И. Иванов, А. В. Безяев, Е. Н. Куприянов //

Инженерные технологии и системы. – 2019. – Т. 29, № 2. – С. 205–217. – DOI https: //doi.org/10.15507/2658-4123.029/2019.02.205-217

5. Иванов, А. И. Нейросетевое обобщение классических статистических критериев для обработки малых выборок биометрических данных / А. И. Иванов, Е. Н. Куприянов, С. В. Туреев // Надежность. – 2019. – № 2. – С. 22–27. – DOI 10.21683/1729-2646-2019-19-2-22-27.

6. Иванов, А. И. Искусственные математические молекулы: повышение точности статистических оценок на малых выборках (программы на языке MathCAD) / А. И. Иванов. – Пенза : Изд-во ПГУ, 2020. – 36 с.

7. Кобзарь, А. И. Прикладная математическая статистика. Для инженеров и научных работников / А. И. Кобзарь. – Москва : ФИЗМАТЛИТ, 2006. – 816 с.

8. Иванов, А. И. Учет влияния корреляционных связей через их усреднение по модулю при нейросетевом обобщении статистических критериев для малых выборок / А. И. Иванов, А. Г. Банных, Ю. И. Серикова // Надежность. – 2020. – № 20 (2). – С. 28–34. – URL: https://doi.org/10.21683/1729-2646-2020-20-2-28-34

Для цитирования:

Иванов, А. И. Использование спектра коэффициентов корреляции Хэмминга для корректного статистического описания выходных кодов нейросетевых молекул / А. И. Иванов, К. А. Ратников // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 73–81.

УДК 519.2, 004.622, 004.056.53

А. П. Карпов, А. П. Юнин

РЕГУЛЯРИЗАЦИЯ ВЫЧИСЛЕНИЯ ЭНТРОПИИ ЛЕГКО ЗАПОМИНАЕМЫХ ДЛИННЫХ ОСМЫСЛЕННЫХ ПАРОЛЕЙ НА РУССКОМ И АНГЛИЙСКОМ ЯЗЫКАХ В ПРОСТРАНСТВЕ СВЕРТОК ХЭММИНГА ПО МОДУЛЮ 256

Аннотация. Рассматривается задача оценки стойкости осмысленных парольных фраз на русском и английском языке, посредством вычисления энтропии кодов этих фраз. Приводятся методы повышения корректности вычисления энтропии кодов осмысленных парольных фраз. Предлагается использовать ряд методов повышения корректности вычисления энтропии кодов осмысленных парольных фраз. Делается вывод, что наиболее корректная оценка осмысленных парольных фраз на русском и английском языке может быть получена с использованием методов регуляризации вычисления энтропии кодов парольных фраз.

A. P. Karpov, A. P. Yunin

REGULARIZATION OF THE CALCULATION OF THE ENTROPY OF EASILY REMEMBERED LONG MEANINGFUL PASSWORDS IN RUSSIAN AND ENGLISH IN THE SPACE OF HAMMING CONVOLUTIONS MODULO 256

Abstract. The problem of assessing the persistence of meaningful password phrases in Russian and English is considered by calculating the entropy of the codes of these phrases. Methods for increasing the correctness of calculating the entropy of codes of meaningful password phrases are given. It is proposed to use a number of methods to increase the accuracy of calculating the entropy of codes of meaningful password phrases. It is concluded that the most correct assessment of meaningful password phrases in Russian and English can be obtained using regularization methods for calculating the entropy of password phrase codes.

Цифровизация всех областей деятельности государства и общества является актуальной тенденцией современного общества. Одновременно с этим, растет число угроз информационной безопасности, риски и величина ущерба от реализации этих угроз. С целью снижения рисков и нейтрализации угроз информационной безопасности необходимо применять комплекс организационно-

технических мер защиты, при этом реализация мер защиты существенно отличается в зависимости от различных параметров конкретной сетевой инфраструктуры.

При этом, базовыми защитными мерами, применяющимися в большинстве информационных систем, является система разграничения доступа и защиты от несанкционированного доступа. Такие системы основаны на процедурах проверки подлинности штатных пользователей – процедурах аутентификации. Методы аутентификации пользователей можно разделить на методы, основанные на знании пользователем некой информации (пароля), методы, основанные на обладании пользователем неким предметом (токен, смарт-карта), и биометрические методы аутентификации, основанные на уникальности различных характеристик пользователя.

В настоящее время наиболее распространѐнным методом аутентификации является аутентификация, основанная на знании пользователем пароля. При этом, оптимальным паролем является длинная псевдослучайная последовательность букв и цифр с использованием знаков различных языков и регистров. Однако применение данных методов сопряжено с человеческим фактором – пользователь не в состоянии запоминать «хорошие» пароли и вынужден записывать и сохранять их, создавая тем самым уязвимости системы аутентификации, которыми может воспользоваться злоумышленник для реализации угроз информационной безопасности.

Оптимальным решением с точки зрения удобства пользователя и качества построения системы аутентификации является использование пользователем длинных осмысленных парольных фраз – легких для запоминания, но трудных для атак злоумышленника. Согласно обновленным рекомендациям Национального института стандартов и технологий, данное решение является наиболее предпочтительным при разработке и эксплуатации систем аутентификации [1].

Для систем аутентификации пользователей, построенных на легко запоминаемых, осмысленных парольных фразах одной из актуальных задач является оценка стойкости парольной фразы – необходимо определить некое пороговое значение стойкости и проверять пароли пользователей на соответствие данному значению.

Решение этой задачи задачей можно получить, оценивая энтропию парольных фраз по Шеннону. Так, для парольной фразой длины 32 символа (надежный пароль, надежный пароль даже с учетом стремительно растущих вычислительных мощностей зло-

умышленников), энтропию можно рассчитать, используя следующую формулу

$$H("x_1, x_2, ..., x_{256}") = -\sum_{i=1}^{2^{256}} P_i \cdot \log_2(P_i), \quad (1)$$

где P_i – вероятность появления одного из 2^{256} состояний восьмибитного кода, тестируемой парольной фразы.

Однако, на практике оценка энтропии по Шеннону неприменимо. Так, для реализации вычислений по формуле (1), необходимо определять вероятности появления очень редких событий, и оценка энтропии на обычной вычислительной машине оказывается технически не выполнима даже для паролей, состоящих из 8 букв (64 бита).

На практике вычислить энтропию длинных кодов можно с использованием рекомендаций ГОСТ Р 52633.3 [2]. В данном стандарте описаны методы оценки энтропии посредством перехода к вычислению расстояний Хэмминга по модулю два:

$$h_2 = 256 - \sum_{i=1}^{256} ("x_i") \oplus ("T_i"), \quad (2)$$

где $\oplus$ – операция сложения по модулю два; «x_i» – разряды тестируемой последовательности; «T_i» – разряды тестовой последовательности, например, осмысленных фраз русского языка в 8-ми битной ASCII кодировке.

При этом ожидание редких событий по Шеннону замещается на предсказание вероятности редких событий в пространстве расстояний Хэмминга. Это возможно за счет нормального распределения расстояний Хэмминга для кодов с длиной более 32 бит, что позволяет, в частности, оценивать энтропию сильно коррелированных откликов нейронной сети на биометрические образы [3–5].

При условии, что объем тестовой выборки должен примерно на порядок быть больше, чем обратная величина экспериментально оцениваемой вероятности, для оценки расстояния 256 битной последовательности к эталонному тексту по Шеннону (1) требуется 2^{256+4} случайных кодов, в то время как при переходе в пространство расстояний Хэмминга, для вычисления математического ожидания – $E(h)$ и стандартного отклонения – $\sigma(h)$ достаточно выборки в сто опытов ($100 \approx 2^7$). Таким образом, сокращение объемов тестовой выборки при вычислении энтропии в рассматриваемом случае может достигать величины – 2^{253}.

С учетом того, что свертка Хэмминга может быть выполнена не только по модулю два [3–5], можно повысить точность вычис-

лений, учитывая структуру кодов парольных фраз. Для сопоставления результатов регуляризации, нормируем интервал, в котором могут меняться расстояния Хэмминга:

$$\tilde{h} = \frac{h}{\max(h)}, \qquad (3)$$

На рис. 1 приведены распределения нормированных усредненных расстояний Хэмминга для ста парольных фраз на русском языке длинной 32 символа и эталонных текстов на русском и английском языках длинной 32 000 символов. В качестве эталонных текстов были выбраны литературные произведения конца двадцатого – начала двадцать первого века. Осмысленные парольные фразы выбирались случайным образом из аналогичных произведений.

Рис. 1. Распределения расстояний Хэмминга при свертывании усредненных кодов длинных осмысленных паролей с эталонными текстами на русском и английском языках

Если попытаться осуществить атаку, подбирая пароль на русском языке английскими фразами, то получим очень большую оценку энтропии:

$$-\log\left(pnorm\left(\frac{1}{256}, 0.439, 0.027\right), 2\right) = 192.661 \text{ бит}$$

Полученные оценки являются слишком оптимистичными, что обусловлено тем, что при вычислениях не принимается в расчет восьмибитная кодировка символов ASCII. При использовании в качестве кодировки другие кодировки, используемые в современных системах – UTF, KOI8-R и др., проблема оптимистичности сохраняется. Это связано с тем, что в любой из кодировок коды символов английского и русского алфавита при свертке парольной фразы с эталонным текстом будут иметь не более 8 значимых информационных бит. Учитывая структуру кодов, получим следующее выражение для расчета энтропии длинных осмысленных парольных фраз:

$$h_8 = 256 \cdot 32 - \sum_{i=1}^{32} \left("c_i, c_{i+1}, \ldots, c_{i+8}"\right) \oplus_8 \left("x_i, x_{i+1}, \ldots, x_{i+8}"\right). \qquad (4)$$

Полученные распределения расстояний Хэмминга с учетом структуры кодов ASCII приведены на рис. 2.

Рис. 2. Распределения расстояний Хэмминга при свертывании усредненных кодов длинных осмысленных паролей с эталонными текстами на русском и английском языках с учетом кодировки символов

Оценивая энтропию осмысленного пароля на русском при его тестировании на русском и английском языках с учетом структуры кодов, получаем гораздо более реалистичные данные:

$$-\log\left(pnorm\left(\frac{1}{256\times 32}, 0.27, 0.034\right), 2\right) = 28 \text{ бит}$$
$$-\log\left(pnorm\left(\frac{1}{256\times 32}, 0.544, 0.013\right), 2\right) = 482 \text{ бит.}$$

На обоих рисунках, можно заметить «хвост» распределения расстояний Хэмминга, обусловленный, связанный с существенным расстоянием между группами кодов знаков препинания (коды 32–46) и кодов букв латиницы (коды 65–122) кириллицы (коды 192–255). Процедуры вычисления сверток Хэмминга можно сделать более устойчивыми, перекодировав символы алфавита, сокращая расстояния между кодами группы «кириллица» и знаков препинания для текстов на русском и отсортировав эти значения по возрастанию/убыванию вероятности встречи символа в тексте. Пример подобной кодировки показан на рис. 3

Символ	Код	Вероятность
,	0	0,011065769
…	1	0,001189843
"	2	2,5346E-05
:	3	0,001546144
	4	0,139647644
-	5	0,001504824
.	6	0,024704492
ё	7	6,05987E-05
А	8	0,001908975
Б	9	0,000941134
В	10	0,003555366
Г	11	0,000426946
Д	12	0,001494951
Е	13	0,000580543
Ж	14	0,000152863
З	15	0,001437133

Кодировка ASCII с устраненными разрывам между символов

Символ	Код	Вероятность
	0	0,137721253
о	1	0,096626212
а	2	0,087388314
е	3	0,092833245
и	4	0,088318646
	5	0,087634134
	6	0,09035751
	7	0,085120575
	8	0,091806405
л	9	0,099913548
в	10	0,083493414
к	11	0,080470095
м	12	0,087512211
д	13	0,086626818
у	14	0,090447353
.	15	0,089776632

Кодировка, учитывающая вероятность встречи символа

Рис. 3. Часть таблицы перекодировки символов

Пример вычисления энтропии усредненных кодов осмысленного пароля на русском при его тестировании на русском языке в кодировке без разрывов между кодами символов приведена на рис. 4.

Рис. 4. Распределения расстояний Хэмминга при свертывании усредненных кодов длинных осмысленных паролей с эталонными текстами на русском языке после перекодировки символов

Руководствуясь полученными результатами, можно заключить, что в системах аутентификации оценку стойкости длинных легко запоминаемых парольных фраз можно выполнять, используя вычисления энтропии по Хэммингу. При этом более корректные результаты таких оценок можно получить при использовании в качестве эталонного текста текст на языке носителя парольной фразы, а также за счет перекодировки символов текста, которая устраняет разрывы между символами и учитывает вероятности встречи символов в тексте.

Библиографический список

1. NIST Special Publication 800-63B. Digital Identity Guidelines. Authentication and Lifecycle Management / P. A. Grassi, J. L. Fenton, E. M. Newton, R. A. Perlner, A. R. Regenscheid, W. E. Burr, J. P. Richer. – URL: https://nvlpubs.nist.gov/nistpubs/SpecialPublications/NIST.SP.800-63b.pdf

2. ГОСТ Р 52633.3–2011. Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора. – Москва : Стандартинформ, 2018. – 12 с.

3. Иванов, А. И. Оценка усиления стойкости коротких цифровых паролей (PIN-кодов) при их рукописном воспроизведении / А. И. Иванов, О. В. Ефимов, В. А. Фунтиков // Защита информации. INSIDE. – 2006. – № 1. – С. 55–57.

4. Малыгин, А. Ю. Быстрые алгоритмы тестирования нейросетевых механизмов биометрико-криптографической защиты информации / А. Ю. Малыгин, В. И. Волчихин, А. И. Иванов, В. А. Фунтиков. – Пенза, 2006 : Изд-во ПГУ. – 161 с.

5. Иванов, А. И. Многомерная нейросетевая обработка биометрических данных с программным воспроизведением эффектов квантовой суперпозиции : монография / А. И. Иванов. – Пенза : Изд-во АО «ПНИЭИ», 2016. – 133 с. – URL: http://пниэи.рф/activity/science/BOOK16.pdf

6. Волчихин, В. И. Условия корректного вычисления энтропии осмысленных длинных паролей в пространстве сверток Хэмминга с эталонными текстами на русском и английском языках / В. И. Волчихин, А. И. Иванов, А. П. Карпов, А. П. Юнин : сб. ст. Всерос. науч.-техн. конф., посвящ. 100-летию со дня рождения одного из основоположников советской вычислительной техники Б. И. Рамеева. – URL: http://rosoperator.ru/rameev100

Для цитирования:

Карпов, А. П. Регуляризация вычисления энтропии легко запоминаемых длинных осмысленных паролей на русском и английском языках в пространстве сверток Хэмминга по модулю 256 / А. П. Карпов, А. П. Юнин // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 82–89.

УДК 519.2, 004.622

А. И. Иванов, А. В. Безяев, Е. И. Качайкин, А. В. Елфимов

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ: АВТОМАТИЗИРОВАННЫЙ НЕЙРОСЕТЕВОЙ АНАЛИЗ «МЕРТВОЙ» ПОДПИСИ ПОД ДОКУМЕНТАМИ НА БУМАЖНЫХ НОСИТЕЛЯХ

Аннотация. Дан обзор исследований, проведенных в России по нейросетевому анализу «мертвых» автографов на бумаге в период с 2009 г. по настоящее время с целью автоматизации почерковедческой экспертизы. Выделены два направления исследований. Дается обоснование необходимости разработки национального стандарта России, регламентирующего подготовку данных при анализе авторства «мертвых» автографов на бумаге большими искусственными нейронными сетями. Оценивается снижение стоимости и сокращение времени проведения предварительной полностью автоматизированной нейросетевой экспертизы.

A. I. Ivanov, A. V. Bezyaev, E. I. Kachaikin, A. V. Elfimov

ARTIFICIAL INTELLIGENCE: AUTOMATED NEURAL NETWORK ANALYSIS OF "DEAD" SIGNATURES UNDER DOCUMENTS ON PAPER MEDIA

Abstract. An overview of research conducted in Russia on the neural network analysis of "dead" autographs on paper from 2009 to the present is given in order to automate handwriting examination. Two areas of research have been identified. The justification for the need to develop a national standard of Russia, regulating the preparation of data when analyzing the authorship of "dead" autographs on paper by large artificial neural networks. The cost reduction and reduction of the time of pre-fully automated neural network examination is estimated.

Общие положения технологии создания приложений нейросетевой биометрии в контексте развития искусственно интеллекта

Президент России В. В. Путин своим указом № 460 от 10.10.2019 «О развитии искусственного интеллекта в Российской Федерации» задал вектор движения отечественной науки на ближайшее время. Очевидно, что указ президента подразумевает разработку новых технологий, создаваемых в интересах того или ино-

го ведомства. Естественно, что разработка новых технологий не может начинаться с нуля, должен учитываться уже имеющийся в России научно технический потенциал. В данной работе будет предпринята попытка инвентаризации отечественного научно-технического задела по анализу динамики «живых» рукописных автографов, восстановления псевдо-динамики «мертвых» автографов под документами на бумаге, преобразования «мертвых» подписей как статических изображений для последующего высокоразмерного нейросетевого анализа в интересах экспертизы их подлинности.

Необходимо отметить, что анализ «мертвых» подписей под документами является одной из востребованных ветвей криминалистики. Именно из-за востребованности обществом этого направления исследований сегодня сложились определенные показатели затрат на труд людей-экспертов. Судя по объявлениям в сети Интернет, экспертиза подлинности «мертвой» подписи под документом стоит от 3000 рублей до 10000 рублей и занимает от 3 до 7 дней. Рынок сформировал расценки на услугу того или иного качества и приемлемые для потребителя интервалы времени выполнения работ.

Столь значительная стоимость услуги и времени ее исполнения обусловлена уровнем загрузки высококвалифицированных людей-экспертов, способных выполнять подобную работу. С другой стороны, столь высокая стоимость услуг и столь длительный срок выполнения экспертизы делают ее низко востребованной практикой поддержки приемлемого для общества уровня правопорядка. Частные охранные агентства, службы безопасности банков и иных предприятий, нотариусы редко прибегают к глубокому экспертному анализу подлинности автографов под документами.

Очевидно, что снижение стоимости предварительного анализа подлинности «мертвого» автографа в документе до уровня стоимости месячных SMS услуг (от 30 до 90 рублей) и снижение времени выполнения экспертизы с трех суток до 3 минут должны положительно повлиять на востребованность структурами поддержания правопорядка этой биометрической услуги.

Современные технологические возможности по анализу динамики воспроизведения «живой» рукописной подписи

Последнее десятилетие прошлого XX в. и первые двадцать лет нынешнего XXI в. были отмечены бурным развитием биометрических технологий. В частности был разработан и введен в действие международный стандарт ISO/IEC sc37 (Биометрия),

который в России гармонизован как стандарт [1]. Так как достоверные биометрические данные недоступны в России в 2009 г. было создано специальное приложение [2] «БиоНейроАвтограф» свободно распространяемое и позволяющее любому лицу получит данные своей динамики рукописного почерка. Для этой цели пользователь должен написать на любом графическом планшете рукописный слова своим почерком, как это показано на рис. 1.

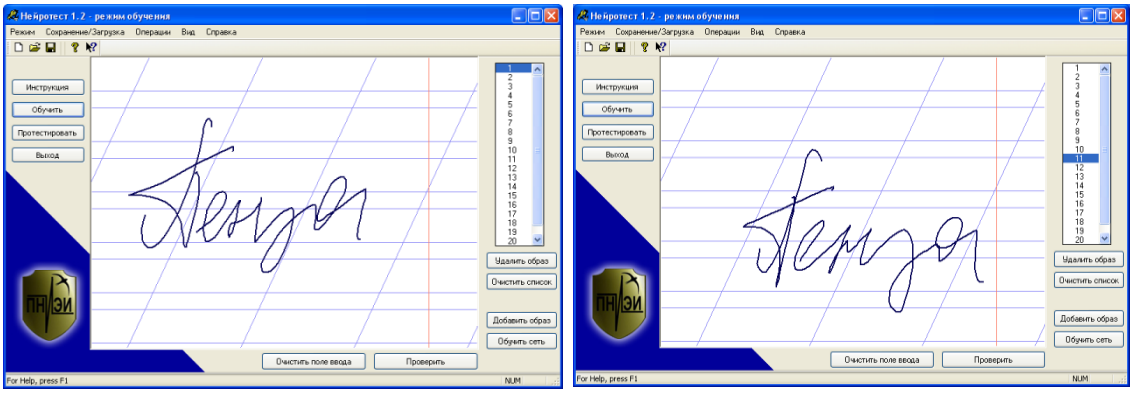


Рис. 1. Сбор биометрических данных «Свой» в среде моделирования «БиоНейроАвтограф» [2]

В программный продукт встроена нейронная сеть, автоматически обучаемая алгоритмом ГОСТ Р 52633.5–2011 [3] на 20 примерах образа «Свой». На рис. 1 отражена ситуация, когда вводится 20 раз рукописный образ «*Пенза*» перед автоматическим обучением нейросети.

После обучения нейронной сети выполняется ее тестирование по ГОСТ Р 52633.3–2011 [4]. Для тестирования необходимо сформировать тестовую базу образов «Чужой», что отражено на рис. 2. Для корректного тестирования необходима тестовая база, состоящая примерно из 100 образов «Чужой».

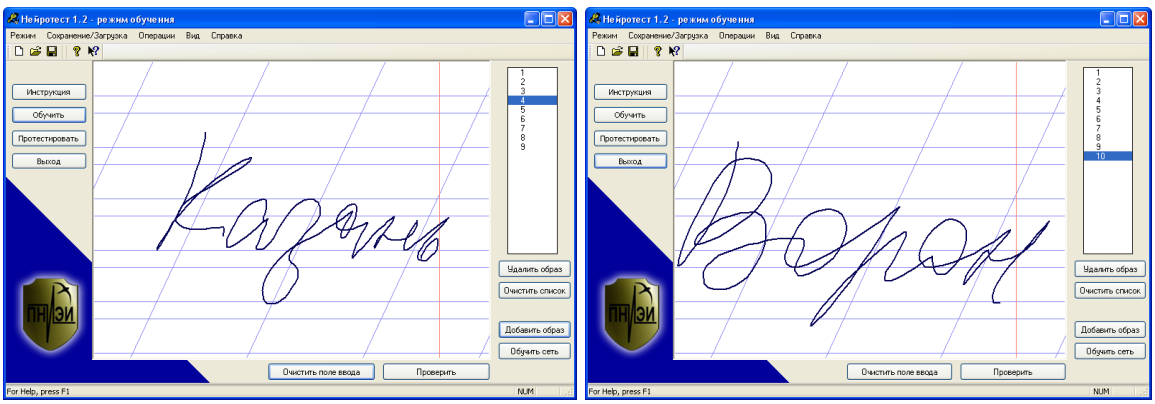


Рис. 2. Сбор биометрических данных тестовой базы образов «Чужой» в среде моделирования «БиоНейроАвтограф» [2]

Программный продукт обеспечивает доступ ко входным данным нейросети (контролируется 416 биометрических параметров) и выходным кодам (256 бит ключа, задаваемого при обучении). Выводятся так же весовые коэффициенты всех 256 искусственных нейронов, как получить доступ к информации и как выполнять лабораторные работы, описано в учебном пособии [5].

Получение биометрических данных из «мертвой» подписи, отсканированной с документа на бумаге

Анализ динамики рукописного почерка большими искусственными нейронными сетями дают доверительную вероятность решения 0.95 при обучении нейросети на 20 примерах (1–1/20) = 0.95, вероятность перепутывания образа «Свой» и случайного образа «Чужой» для рукописного пароля из 5 букв составляет от 0.001 до 0.0000001. Эти цифры могут контролироваться тестированием процедурами ГОСТ Р 52633.3–2011 [4]. Эксперты –люди не могут указать вероятности ошибок принимаемых ими решений. К тому же, они анализируют при экспертизе малое число параметров (16 параметров в место 416 параметров динамики автографа). То есть люди-эксперты анализируют значительно меньше биометрических параметров и потому, скорее всего, их решения имеют большее значение вероятностей ошибок первого и второго рода.

Предположительно, если людям экспертам дать больше информации о динамике рукописного почерка, то и они после дообучения будут принимать более точные решения. Полагаясь на этот тезис целесообразно анализировать статические рукописные образы, извлекая из них биометрические параметры несколькими методами:

- обход фрагментов подписи, воспроизводящих наиболее вероятную траекторию движения пера [6];
- оконтуривание образа автографа без учета замкнутых фрагментов;
- выделение центров массы замкнутых фрагментов рукописного образа;
- выделение особых точек рукописной надписи (точек обрывов, точек пересечения линий, точек смены знака производных траектории по двум координатам движения пера;
- двухмерное Фурье преобразование, фрагментов подписи и подписи в целом;
- ……

Приведенный выше перечень методов не полон, возможно, его дополнение в будущем. Расширение списка возможно при проведении самостоятельных исследований, конкурирующих между собой на рынке подобных услуг производителей. Однако на сегодняшний день можно считать, что работоспособная технология извлечения биометрических параметров из «мертвой» подписи уже создана и апробирована [7]. Созданы технические предпосылки для автоматизации почерковедческой экспертизы [8].

Возможность независимого тестирования качества программных продуктов, представленных на рынке услуг искусственного интеллекта

Так же как продукт «БиоНейроАвтограф», его аналоги для анализа «мертвой» подписи продукты должны иметь встроенный режим самотестирования вероятности ошибок первого и второго рода [9, 10]. Кроме того, должна быть создана общедоступная база тестовых образов «Свой» достаточного объема и создана достаточного объема база тестовых образов «Чужой». Последнее позволит потребителям самостоятельно проводить тестирование того или иного программного продукта разных производителей. Должны так же появиться сертифицированные тестовые лаборатории, способные подтвердить или опровергнуть заявленные производителем статистические показатели его продукта.

Требования к обучению нейронных сетей и к выбору структур используемых нейронных сетей

Очевидно, что все подобные программные продукты должны иметь только автоматическое обучение [11]. Последнее означает, что алгоритм ГОСТ Р 52633.5–2011 [5] обучения сети из 256 нейронов с обогащением данных в линейном пространстве вполне может использоваться для подобных программных продуктов.

Кроме того, одни и те же исходные данные параллельно могут обрабатываться и другими типами искусственных нейронных сетей, для которых на сегодняшний день построены алгоритмы полностью автоматического обучения. В частности могут быть использованы сети квадратичных форм, состоящие из нейронов Махалонобиса [12], радиальных нейронов или нейронов Пирсона [13–15]. Принципиально важным является то, что квадратичные нейроны выполняют обогащение данных в квадратичных пространствах, а не в линейных пространствах, как это делают персептроны.

Как результат выходные коды нейронной сети персептронов и выходные коды сети квадратичных форм имеют существенную не коррелированную компоненту. Выходные коды сетей персептронов и сетей квадратичных форм дополняют друг друга и по этой причине их параллельное применение целесообразно (длина анализируемого выходного кода 2-х типов нейросетей удваивается).

Кроме того, в реальных программных продуктах должны дополнительно использоваться сети нейронов Байеса [12, 16], выполняющие обогащение (накопление) данных в гиперболических пространствах. Обогащение (накопление) данных в гиперболических пространствах дает слабо коррелированные решения по отношению к накоплению данных в линейном и квадратичном пространстве. В итоге мы имеем реальную возможность утроить длину выходного кода нейронных сетей, что положительно сказывается на достижимом соотношении вероятностей ошибок первого и второго рода, проводимой почерковедческой экспертизы нейросетевым искусственным интеллектом.

Библиографический список

1. ГОСТ Р ИСО/МЭК 19794-7–2006. Автоматическая идентификация. Идентификация биометрическая. Форматы обмена биометрическими данными. Ч. 7. Данные динамики подписи. – Москва : Стандартинформ, 2009.

2. Иванов, А. И. Среда моделирования «БиоНейроАвтограф». Программный продукт создан лабораторией биометрических и нейросетевых технологий, размещен с 2009 г. на сайте АО «ПНИЭИ» / А. И. Иванов, О. С. Захаров. – URL: http://пниэи.рф/activity/science/noc/bioneuroautograph.zip

3. ГОСТ Р 52633.5–2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия – код доступа. – Москва : Стандартинформ, 2012. – 20 с.

4. ГОСТ Р 52633.3–2011. Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора. – Москва : Стандартинформ, 2018. – 12 с.

5. Иванов, А. И. Автоматическое обучение больших искусственных нейронных сетей в биометрических приложениях : учеб. пособие / А. И. Иванов. – Пенза, 2013. – 30 с. – URL: http://пниэи.рф/activity/science/noc/tm_IvanovAI.pdf

6. Пат. 2390843 Российская Федерация, МКИ: G06K 9/62. Способ распознавания знаков / Иванов А. И., Андреев Д. Ю., Воячек С. А., Елфимов А. В. – № 2008117180/09 ; заявл. от 29.04.2008 ; опубл. от 27.05.2010, Бюл. № 15.

7. Качайкин, Е. И. Получение биометрических параметров высокого качества из статического изображения рукописной подписи / Е. И. Качайкин, С. В. Куликов // Инфокоммуникационные технологии. – 2015. – № 4. – С. 45–49.

8. Иванов, А. И. Автоматизация почерковедческой экспертизы, построенная на обучении больших искусственных нейронных сетей / А. И. Иванов, А. И. Газин, Е. И. Качайкин, Д. Ю. Андреев // Модели, системы, сети в экономике, технике, природе и обществе. – 2016. – № 1 (17). – С. 249–257.

9. Качайкин, Е. И. Оценка достоверности нейросетевой автоматизированной экспертизы рукописного почерка / Е. И. Качайкин, А. И. Иванов, А. В. Безяев, К. А. Перфилов // Вопросы кибербезопасности. – 2015. – № 2 (10). – С. 43–48.

10. Качайкин, Е. И. Оценка качества результатов почерковедческой экспертизы, осуществляемой нечетким экстрактором. Евразийский Союз Ученых (ЕСУ) / Е. И. Качайкин // Технические науки. – 2015. – № 4 (13). – С. 62–64.

11. Елфимов, А. В. Обучение нейросетевого идентификатора авторства рукописных текстов / А. В. Елфимов, С. А. Воячек, Е. И. Качайкин, С. В. Куликов // Нейрокомпьютеры: разработка, применение. – 2009. – № 6. – С. 17–21.

12. Иванов, А. И. Идентификация подлинности рукописных автографов сетями Байеса – Хэмминга и сетями квадратичных форм / А. И. Иванов, П. С. Ложников, Е. И. Качайкин // Вопросы защиты информации. – 2015. – № 2. – С. 28–34.

13. Качайкин, Е. И. Идентификация авторства рукописных образов с использованием нейросетевого эмулятора квадратичных форм высокой размерности / Е. И. Качайкин, А. И. Иванов // Вопросы кибербезопасности. – 2015. – № 4 (12). – С. 42–47.

14. Гильмутдинов, А. Х. Автоматизированная почерковедческая экспертиза / А. Х. Гильмутдинов, Е. И. Качайкин, А. И. Иванов, А. В. Безяев // Физика и технические приложения волновых процессов : XIII Междунар. науч.-техн. конф. Сек. 12, Фракталы и детерминированный хаос (г. Казань, 21–25 сентября 2015 г.). – Казань : Казан. нац. исслед. техн. ун-т им. А. Н. Туполева, 2015. – С. 305–307.

15. Гильмутдинов, А. Х. Почерковедческая экспертиза на основе радиально-базисных нейронных сетей Пирсона – Хэмминга / А. Х. Гильмутдинов, Е. И. Качайкин, А. И. Иванов, А. В. Безяев // Нелинейный мир. – 2017. – Т. 15, № 3. – С. 3–10.

16. Иванов, А. И. Биометрическая идентификация рукописных образов с использованием корреляционного аналога правила Байеса / А. И. Иванов, П. С. Ложников, Е. И. Качайкин, А. Е. Сулавко // Вопросы защиты информации. – 2015. – № 3. – С. 48–54.

Для цитирования:

Иванов, А. И. Искусственный интеллект: автоматизированный нейросетевой анализ «мертвой» подписи под документами на бумажных носителях / А. И. Иванов, А. В. Безяев, Е. И. Качайкин, А. В. Елфимов // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 90–96.

УДК 004.056.52

А. Е. Боршевников, Ю. В. Добржинский

О КОРРЕКТНОСТИ МОДЕЛИ СИСТЕМЫ ВЫСОКОНАДЕЖНОЙ БИОМЕТРИЧЕСКОЙ АУТЕНТИФИКАЦИИ С ИСПОЛЬЗОВАНИЕМ ЭЛЕКТРОЭНЦЕФАЛОГРАММЫ НА ОСНОВЕ СТАНДАРТОВ ГОСТ Р 52633

Аннотация. Рассматривается корректность математической модели нейросетевого преобразователя биометрия – код доступа на основе электроэнцефалограммы. Описывается модель нейросетевого преобразователя биометрия – код доступа. Проведена проверка требований корректности математической модели преобразователя на примере использования потенциала Р300.

A. E. Borshevnikov, Yu. V. Dobrzhinskii

CORRECTNESS OF A MODEL OF HIGH-RELIABLE BIOMETRIC AUTHENTICATION SYSTEM USING ELECTROENCEPHALOGRAM BASED ON STANDARDS GOST R 52633

Abstract. The article discusses the correctness of the mathematical model of a neural net transformer biometrics – access code based on the electroencephalogram. The model of the neural net transformer biometrics – access code is described. The verification of the correctness requirements of the mathematical model of the transformed was performed using the example of the P300.

Растущая потребность общества в обеспечении информационной безопасности порождает запрос в средствах защиты, обеспечивающих высокий уровень безопасности. Например, такой уровень безопасности необходимо обеспечивать в критически важных объектах.

Отдельную нишу среди всех средств защиты занимают технологии биометрической аутентификации. Рассматриваемый высокий уровень безопасности могут обеспечить средства высоконадежной биометрической аутентификации. Обычно реализация таких систем сводится к восстановлению некоторой парольной строки (криптографического ключа). Одним из подходов для построения подобных систем является подход, основанный на использовании боль-

ших и сверхбольших нейронных сетей совместно со специальными алгоритмами обучения [1–4].

Также для обеспечения высокого уровня безопасности важна конфиденциальность биометрической характеристики, используемой для аутентификации. С этой точки зрения одной из характеристик, обладающей высоким уровнем конфиденциальности, является электроэнцефалограмма (ЭЭГ).

Для исследования подобных систем используется их математическое моделирование, в результате которого получают характеристики данных систем. Наиболее важными характеристиками систем биометрической аутентификации являются ошибки первого и второго рода. Также выходными значениями, которые позволяют оценить качество построенной системы, могут быть и другие параметры.

При разработке математической модели важно, чтобы модель была корректной, то есть результаты вычислений, производимых моделью должны соответствовать результатам работы, описываемого объекта в реальном мире (или результатам, которые должен показывать разрабатываемый объект). Можно говорить о корректности построенной модели только при выполнении определенных требований. Выделяют следующие требования к математическим моделям [5]:

1. Свойство полноты.
2. Свойство точности.
3. Свойство адекватности.
4. Свойство продуктивности.
5. Свойство экономичности.
6. Свойство робастности.
7. Свойство наглядности.

Для принятия корректности математической модели необходимо выполнение большей части перечисленных требований.

Приведем описание модели нейросетевого преобразователя и покажем корректность, описанной модели. Входными значениями модели будут:

1. I – общее количество электродов электроэнцефалографа.

2. $S_{Свой}, S_{Чужой}$ – базы биометрических данных легитимного пользователя (Свой) и злоумышленника (Чужой).

3. $N_{Свой}, N_{Чужой}$ – количество образцов биометрических данных в базах «Свой» и «Чужой».

4. J – количество выбираемых для анализа значений биометрической характеристики.

5. L, R – количество нейронов в слоях первого и второго слоев нейросетевого преобразователя.

ЭЭГ представляет собой нечеткий сигнал, снимаемый по нескольким каналам и изменяющийся во времени, который сложно описывается математически. По этой причине мы будем считать, что сигнал ЭЭГ это набор функций с каналов электроэнцефалографа, зависящих от времени. Наборы этих функций, снятые с легитимных пользователей системы и злоумышленников, составляют базы электроэнцефалограмм "Свой" и "Чужой" соответственно.

В случае описания нейросетевого преобразователя необходимо описать нейронную сеть. В общем случае это можно сделать следующим образом:

$$NET(a_i, \overline{w_l}, \overline{W_r}, \overline{net_l}, \Delta_r) = \overline{K}_{rest},\ 1 \le i \le I,\ 1 \le l \le L,\ 1 \le r \le R, \quad (1)$$

где $\overline{net_l}$ – вектор связей нейрона l; $\overline{w_i}$ – вектор весовых коэффициентов первого слоя; $\overline{W_r}$ – вектор весовых коэффициентов второго слоя; Δ_r – коэффициент использования; $\overline{K}_{rest}$ – восстанавливаемый преобразователем ключ.

Для нейросетевого преобразователя рассчитываются следующие характеристики: неполнота базы естественных биометрических образов "Чужой" φ, математическое ожидание расстояния Хэмминга $\overline{H}(\overline{K}_{rest})$, среднее значение модулей коэффициентов парной корреляции r_{xy}, стабильность выходного кода $P_{0,i}, P_{1,i}$, вероятность ошибки первого рода P_1, вероятность ошибки второго рода P_2.

Приведем данные для полученной модели и проверим ее корректность. Для проведения проверки было решено использовать нейросетевой преобразователь, описанный в работе [6]. Данная конструкция преобразователя была выбрана из-за лучших показателей по сравнению с другими исследуемыми преобразователями. При тестировании использовались следующие параметры преобразователя. Количество электродов – 14. Количество выбираемых коэффициентов Фурье для одного электрода – 15. Количество нейронов первого слоя – 320. Размер восстанавливаемого ключа был выбран – 256, что означает использование во втором слое нейронной сети 256 нейронов. Количество входов на нейрон было взято 4. Преобразователь обучался по стандарту ГОСТ Р 52633.5 [4].

Для проведения проверки была использована синтетическая база образов «Чужой», в которой содержится 10^4 биометрических образцов. Данная база была создана с помощью методов, изложенных в стандарте ГОСТ Р 52633.2 [3].

Для тестирования работы модели проводилось обучение преобразователя с использованием дополненной базы образов «Чужой» и без нее. Для эксперимента один естественный образ был выбран в качестве образа «Свой», остальные девять естественных образов сформировали базу «Чужой».

Был проведен опыт по возможности получения злоумышленником секретного ключа при условии знания весовых коэффициентов и «мысленного пароля» легитимного пользователя.

Стоит отметить, что при проведении экспериментов по восстановлению ключа для образа «Свой», ключ безошибочно восстанавливался. Для случаев, когда тестирующая выборка является небольшой, и ошибка первого рода не была выбрана, данную ошибку можно вычислить по следующей формуле [7]:

$$P_1 \approx \int_1^{\infty} \frac{1}{2^{\frac{\Omega}{2}} \cdot \Gamma\left(\frac{\Omega}{2}\right)} \cdot x^{\frac{\Omega}{2}-1} \cdot e^{-\frac{x^2}{2}} \cdot dx, \qquad (2)$$

где Ω – количество степеней свободы в распределении X^2.

В случае, когда в проведенной серии испытаний по предъявлению биометрической характеристики образа «Свой», состоящей из m опытов, не обнаружен факт отказа в доступе, число степеней свободы в распределении вычисляется по формуле:

$$\Omega = \frac{1}{m+1}. \qquad (3)$$

Для $m = 20$ получим ошибку первого рода $P_1 = 7 \cdot 10^{-3}$.

Значения $\overline{H}(\overline{K}_{rest})$ принимают следующие величины:

– 68, 81, 74 для преобразователя, обученного на базе, состоящей исключительно из естественных образов.

– 130, 122, 150 для преобразователя, обученного на базе, дополненной синтетическими образами.

Приведем схему приблизительной оценки ошибки второго рода на основе полученных результатов.

Вероятность ошибки второго рода P_2 можно вычислить приближенно, исходя из гипотезы нормального закона распределения значений вероятности ошибок, по формуле [2]:

$$P_2 \approx \frac{1}{\sqrt{2\pi}} \int_{\sqrt{n}E(q(v))}^{\infty} e^{-\frac{x^2}{2}} dx, \qquad (4)$$

где n – число учитываемых преобразователем биометрических параметров; $E(q(v))$ – среднее качество всех учитываемых преобразователем биометрических параметров.

В построенном преобразователе использовались 210 параметров. Производя вычисления, получим приблизительную оценку вероятности ошибки второго рода $P_2 \le 10^{-12}$.

Для проведенных экспериментов с недополненной и дополненной синтетическими образами базами показатель неполноценности составил $\frac{1}{6}$ и $\frac{1}{3}$ соответственно.

Стабильность выходного кода $P_{0,i}, P_{1,i}$ составила 0,5.

Среднее значение модулей коэффициентов парной корреляции $r_{xy} = 0,06$.

Покажем корректность, построенной модели.

Требование точности выполняется частично. Данное требование контролируется по нормам, устанавливаемым стандартом ГОСТ Р 52633.0 [1]. Пусть $\varepsilon()$ – погрешность выходной величины. Тогда можно получить следующие величины погрешности для выходных характеристик модели:

– $\varepsilon(P_1) \le 0$.

– $\varepsilon(P_2) \le 0$.

– $\varepsilon(\overline{H}(\overline{K}_{rest})) \le |8|$.

– $\varepsilon(\varphi) \ge 0$.

– $\varepsilon(P_{0,i}, P_{1,i}) \le |0,1|$.

– $\varepsilon(r_{xy}) \le |0,15|$.

Несмотря на то, что некоторые характеристики не удовлетворяют требованиям точности можно говорить о выполнении данного требования, так как для значений ошибок первого и второго рода данные требования выполняются.

Требования полноты, адекватности, продуктивности, экономичности, робастности и наглядности за счет точной постановки условий для модели.

Таким образом, построенная модель нейросетевого преобразователя «Биометрия – код доступа» на основе электроэнцефало-

граммы в целом является корректной. Невыполнение некоторых требований точности в основном можно связать с малым размером, исследуемой базы данных ЭЭГ. Для более точной оценки точности важными задачами для последующих исследований являются:

– Расширение базы естественных образов "Свой" и "Чужой".

– Расширение базы синтетических образов "Чужой".

Поставленные задачи также позволят проверить масштабируемость, построенной системы.

Библиографический список

1. ГОСТ Р 52633.0–2006. Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации. – Москва : Стандартинформ, 2007. – 25 с.

2. ГОСТ Р 52633.1–2009. Защита информации. Техника защиты информации. Требования к формированию баз естественных биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации. – Москва : Стандартинформ, 2010. – 24 с.

3. ГОСТ Р 52633.2–2010. Защита информации. Техника защиты информации. Требования к формированию синтетических биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации. – Москва : Стандартинформ, 2011. – 17 с.

4. ГОСТ Р 52633.5–2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия – код доступа. – Москва : Стандартинформ, 2012. – 20 с.

5. Маркелов, Г. Е. Основные принципы построения математических моделей / Г. Е. Маркелов // Вестник МГТУ им. Н. Э. Баумана. Сер.: Естественные науки. – 2005. – № 4. – С. 59–70.

6. Гончаров, С. М. Использование вейвлет-преобразования для выделения биометрических характеристик потенциала Р300 в задачах высоконадежной биометрической аутентификации / С. М. Гончаров, А. Е. Боршевников // Информация и безопасность. – 2016. – Т. 19, ч. 4. – С. 527–530.

7. Ахметов, Б. С. Оценка вероятностей появления ошибок нейросетевых преобразователей биометрия-код на основе малых выборок / Б. С. Ахметов, А. И. Иванов, А. Ю. Малыгин, Т. С. Картбаев // Высокие технологии – залог устойчивого развития : тр. II Междунар. науч. конф. – Алматы, 2013. – Т. 1. – С. 234–237.

Для цитирования:

Боршевников, А. Е. О корректности модели системы высоконадежной биометрической аутентификации с использованием электроэнцефа-лограммы на основе стандартов ГОСТ Р 52633 / А. Е. Боршевников, Ю. В. Добржинский // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 97–102.

УДК 519.2, 004.622

А. Е. Сулавко

РАЗНОСТНЫЕ НЕЙРОНЫ БАЙЕСА С МНОЖЕСТВОМ КВАНТОВАТЕЛЕЙ ДЛЯ ВЫСОКОНАДЕЖНОЙ АУТЕНТИФИКАЦИИ И ЗАЩИЩЕННОГО ИСПОЛНЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Аннотация. Рассмотрена проблема защиты биометрических эталонов и личных криптографических ключей (длинных паролей) пользователя при создании преобразователей биометрия-код на основе гибридных нейронных сетей из классических нейронов и разностных нейронов Байеса. Нейросетевые преобразователи биометрия-код являются идеологической основой для серии стандартов ГОСТ Р 52633, которые могут быть востребованы при разработке средств высоконадежной биометрической аутентификации и электронной подписи с биометрической активацией. Однако подобный подход может быть применен при создании систем искусственного интеллекта в защищенном исполнении. Рассмотрена гибридная модель преобразователя образов, не компрометирующего биометрический эталон и ключ (пароль) пользователя и устойчивого к атакам извлечения знаний из параметров решающего правила.

A. E. Sulavko

BAYES' DIVERGENT NEURONS WITH MANY QUANTIFIERS FOR HIGHLY RELIABLE AUTHENTICATION AND PROTECTED EXECUTION OF ARTIFICIAL INTELLIGENCE

Abstract. The problem of protecting biometric references and personal cryptographic keys (long passwords) of the user in the creation of biometric code converters based on hybrid neural networks from classic neurons and difference neurons of Bayes is considered. Neuronet converters biometric code are the ideological basis for a series of standards GOST P 52633, which can be in demand in the development of highly reliable biometric authentication and electronic signature with biometric activation. However, this approach can be applied to the creation of artificial intelligence systems in a secure execution. The hybrid model of the image converter, which does not compromise the biometric reference and key (password) of the user, and resistant to attacks extracting knowledge from the parameters of the decisive rule is considered.

Один из основных мировых трендов на сегодняшний день связан с развитием технологий искусственного интеллекта (ИИ).

Под этим термином подразумевается способность программ выполнять задачи, которые считаются прерогативой человека – классификация, кластеризация, регрессия. В соответствии с Указом Президента РФ № 490 «О развитии искусственного интеллекта в Российской Федерации» до 2030 г. ставится задача поддержки научных исследований, направленных на развитие отечественных методов и технологий ИИ.

ИИ должен быть исполняться в защищенном режиме. Под «защищенным исполнением» понимается невозможность совершения следующих действий любым неавторизованным лицом или субъектом (процессом, пользователем, злоумышленником):

– анализа операций, совершаемых ИИ (алгоритма работы ИИ, суть преобразований);

– управления ИИ (с помощью изменения алгоритма работы, подмены данных ИИ и т.д.);

– извлечения знаний ИИ.

Любое несанкционированное вмешательство в работу ИИ может повлечь за собой последствия – материальный ущерб, нарушение информационной безопасности, угрозу жизни, здоровья граждан, технологический сбои или катастрофы и т.д. Все зависит от назначения конкретной реализации ИИ и возможностей, которыми данный экземпляр ИИ обладает.

Важным аспектом, касающимся алгоритмов обучения ИИ, является их устойчивость и способность «настроить» ИИС или ИНС, даже если объем обучающей выборки ограничен. Алгоритм обучения должен быть робастным, при этом после обучения ИИС (ИНС) должна быть способна давать высоконадежные, качественные решения поставленных задач. На сегодняшний день стандартизованы процедуры полностью автоматического обучения и защиты параметров искусственных нейронных сетей, осуществляющих обогащение бедных входных данных путем их накопления в линейном пространстве (ГОСТ Р 52633.5–2011).

Серия стандартов ГОСТ Р 52633 относится к построению нейросетевых преобразователей биометрия-код (НПБК) [1], на базе которых должны строится средства высоконадежной биометрической аутентификации. НПБК настраивается на выдачу ключа (пароля) пользователя при предъявлении его биометрического образа. При предъявлении образа любого другого субъекта НПБК должен формировать случайный бинарный код, близкий по информационной энтропии к «белому шуму». Однако подобные решения воз-

можно использовать для других задач интеллектуального анализа данных (пока только распознавания образов, но в перспективе для кластеризации и регрессии).

Результаты последних исследований указывают на то, что гибридные нейронные сети [2] также способны к робастной настройке, при этом могут давать более высокоточные решения. Рассмотрим способ защиты параметров, обученных гибридных нейро-Байесовских сетей путем применения обратимых и необратимых преобразований, аналогичных тем, которые применяются для защиты искусственных нейронных сетей, обученных по ГОСТ Р 52633.5 [3].

Классический нейрон теряет мощность при усилении корреляционных связей между признаками. По этой причине количество информации на входе классического нейрона всегда меньше, чем сумма собственной информации всех признаков. Абсолютно иначе дело обстоит, если вместо функционала взвешенного суммирования использовать разностный Байесовский функционал (1) с функцией активации (2):

$$y = \sum_{j=1}^{n} \left| \frac{a_t}{\sigma(a_t)} - \frac{a_j}{\sigma(a_j)} \right|, j \neq t \tag{1}$$

$$f(y) = \begin{cases} 0, if \;\; y < \mu_0 \\ 1, if \;\; y > \mu_0 \end{cases} \tag{2}$$

где y – отклик нейрона на образ “Свой” или “Чужой”, $f(y)$ – ответ нейрона, a_j – значение j-го признака (входа нейрона), $m(a_j)$ и $\sigma(a_j)$ – математическое ожидание и среднеквадратичное отклонение значений j-го признака для образа “Свой”, μ_0 – порог активации нейрона, n – количество входов нейрона.

Многомерный разностный функционал Байеса (1) дает тем меньшее значение, чем выше коэффициент корреляции признака под номером t с признаками под номерами j. Соответственно, чем больше n и выше корреляция между признаками r, тем меньше вероятность ошибок распознавания образа (рис. 1). Убедиться в этом несложно, достаточно сгенерировать описания абстрактных классов образов в двух пространствах признаков с нормальным законом распределения: независимых и зависимых, после чего необходимо построить соответствующие функции плотностей вероятностей (рис. 1).

На основе метрики (1) возможна нейросетевая обработка коррелированных сочетаний признаков. Отметим, что для ее корректной работы требуется хранить только параметры $\sigma(a)$. Это могут

быть как среднеквадратичные отклонения признаков для класса образов «Свой», так и для класса «Чужие», что также допустимо. Во втором случае $\sigma_s(a)$ вообще не компрометирует биометрический эталон пользователя.

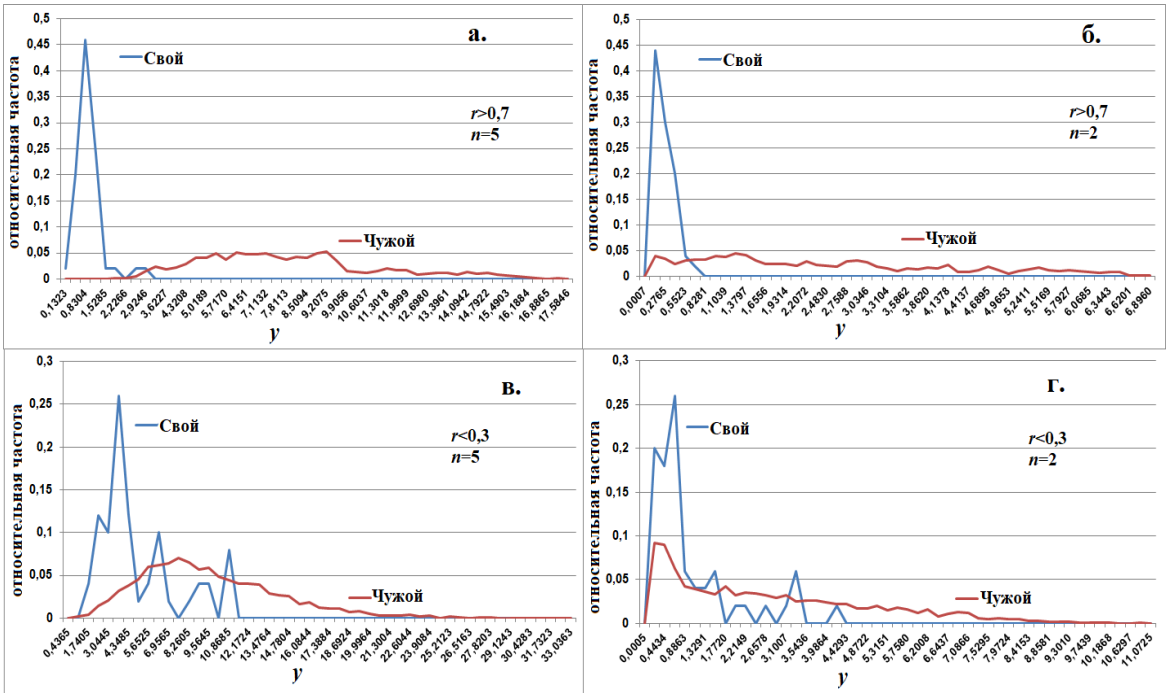


Рис. 1. Эмпирические плотности вероятности откликов разностных нейронов Байеса (1) при распознавании абстрактных образов: в случае зависимых признаков (r > 0,7): (а) n = 5 (б) n = 2; в случае независимых признаков (r < 0,3): (в) n = 5, (г) n = 2

Порог срабатывания функции активации (2) для разностного нейрона Байеса вычисляется по следующему правилу:

$$\mu_0 = m_o(d_t) + \sigma_o(d_t)\cdot\beta,$$

где β – коэффициент баланса FRR и FAR для нейронов Байеса.

Нейрон на основе метрики (1) можно модифицировать, применив функцию активации с несколькими квантователями, чем больше квантователей, тем выше хеширующие свойства разностного Байесовского нейрона. Трехуровневое квантование соответствует функциям активации:

$$f(y)=\begin{cases}"01", if\ \ y<\mu_0\\ "10", if\ \ \mu_0<y<\mu_1\\ "11", if\ \ y>\mu_1\end{cases} \qquad f(y)=\begin{cases}"00", if\ \ y<\mu_0\\ "10", if\ \ \mu_0<y<\mu_1\\ "11", if\ \ y>\mu_1\end{cases}$$

$$f(y)=\begin{cases}"00", if\ \ y<\mu_0\\ "01", if\ \ \mu_0<y<\mu_1\\ "11", if\ \ y>\mu_1\end{cases}\qquad f(y)=\begin{cases}"00", if\ \ y<\mu_0\\ "10", if\ \ \mu_0<y<\mu_1\\ "01", if\ \ y>\mu_1\end{cases}$$

Существует всего 4 варианта ее реализации (к разным нейронам можно применить различные активационные функции в рамках одного слоя сети). На выходе этих функций будет два бинарных значения, при этом для каждой из них существует только 3 возможных двухбитных состояния. Четырехуровневое квантование однозначно определяется функцией активации с тремя порогами:

$$f(y)=\begin{cases}"00", if\ \ y<\mu_0\\ "10", if\ \ \mu_0<y<\mu_1\\ "01", if\ \ \mu_1<y<\mu_2\\ "11", if\ \ y>\mu_2\end{cases}$$

Чтобы сеть из разностных нейронов Байеса (1) обладала хорошими хэширующими свойствами, необходимо, чтобы состояния каждого нейрона были равновероятны (по аналогии с соответствующим требованием для двухуровневого линейного нейрона в ГОСТ Р 52633.5-2011): для трехуровневых нейронов – 0,333, для четырехуровневых – 0,25 и т.д.

Возникает вопрос: зачем нужны множественные квантователи, если можно использовать несколько нейронов с одним квантователем и двумя бинарными выходными состояниями? В качестве ответа можно привести два аргумента:

1. Решения всех нейронов в той или иной степени коррелированы. Даже, если на входах двух нейронов различные признаки, их решения все равно могут коррелировать, так как сами признаки могут быть коррелированными.

2. Чем больше нейронов, тем выше вычислительная нагрузка.

Таким образом, один «многоуровневый» нейрон работает как несколько независимых нейронов с одним квантователем.

Если объединить классические нейроны и разносные нейроны Байеса в гибридную сеть, можно снизить показатели FRR и FAR, а также повысить энтропию ответов ПБК. При построении гибридного ПБК из классических нейронов и нейронов Байеса необходимо учитывать тот факт, что разностные нейроны Байеса компрометирует биты ключа пользователя. Каждый нейрон должен давать на выходе несколько бит ключа (верных или нет – зависит от преодоления соответствующих порогов нейрона), которые необходимо хранить.

Энтропия ответов ПБК на образы «Чужие» является важным показателем, так как она связана с FAR: чем ниже FAR, тем выше энтропия [1]. Приблизительную оценку энтропии можно получить, вычислив собственную информацию события «ложного допуска”:

$$E(\mathrm{FAR}) \approx -\log_2 \mathrm{FAR}.$$

Из рис. 2,а видно, что, контролируя допустимое число ошибочных бит в ответе «широкой» сети можно балансировать FRR и FAR (например, применяя корректирующие коды или второй слой нейронов для исправления нескольких неверных бит). Однако эта возможность одновременно является уязвимостью. Хакер может собрать большую базу примеров произвольных паролей, воспроизведенных различными «Чужими», и оценить среднюю стабильность ответов ПБК для каждого «Чужого» по формуле (3). Стабильность ответов НПБК при предъявлении образов «Чужих» возрастает, если ответы близки (в метрике Хемминга) к ключу пользователя или его инверсии. Данное нежелательно свойство НПБК позволяет совершать направленный перебор биометрических образов (осуществляя одновременно поиск наиболее близкого и наиболее дальнего образа «Чужого» относительно образа «Свой»).

$$\gamma_k = \sum_{l=1}^{L} 2\left|P_l(1) - 0{,}5\right|, \qquad (3)$$

где k – номер «Чужого», L – количество нейронов, l – номер нейрона, $P_l(1)$ – вероятность (или относительная частота) появления «единицы» (можно заменить на $P_l(0)$) в l-м разряде ответа ПБК (в выходе l-го нейрона) на примеры образа k-го «Чужого».

В своих работах Иванов А.И. и соавторы [1] предлагают защищать нейросетевые контейнеры путем применения обратимых и необратимых преобразований. Каждый нейрон имеет таблицы связей и весов. Для защиты таблиц нейросетевых функционалов нужно применять механизм защищенного нейросетевого контейнера (ЗНК). Нейроны можно выстроить в цепочку, как показано на рис. 3а. После обучения ПБК таблицы каждого нейрона шифруются наложением гаммы, представляющей собой контрольную сумму выходов всех предыдущих нейронов в цепочке:

$$tables'_l = XOR(tables_l, hash(pass, bit_1, ..., bit_{l-1})),$$

где bit_l – выход, на который настраивается l-й нейрон в цепочке, $hash()$ – криптографическая хеш-функция (например, md5), $tables_l$ – таблицы параметров соответствующего нейрона, $pass$ – пароль,

который является опциональным и служит для дополнительной (2-х факторной) защиты.

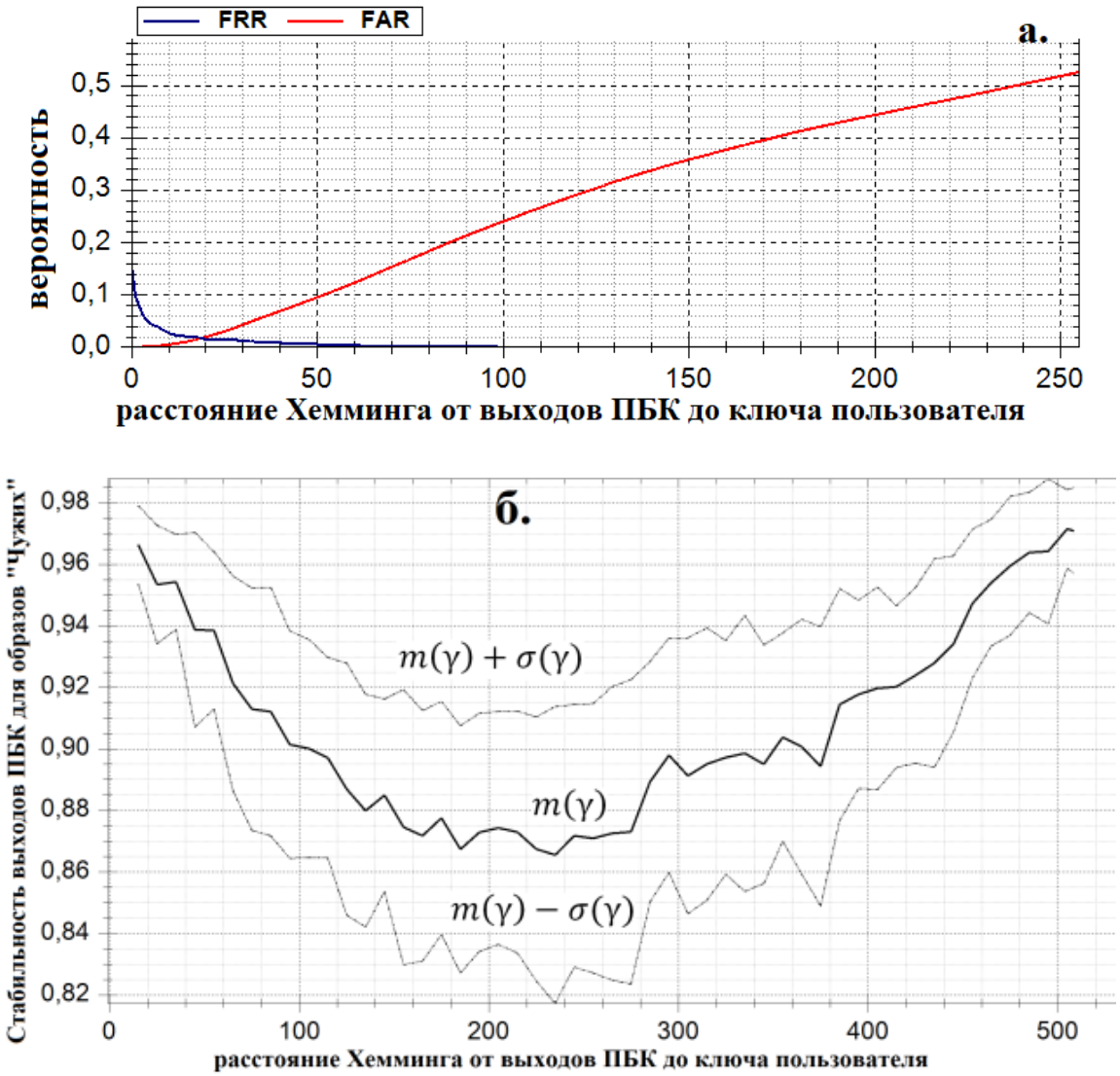


Рис. 2. Результаты тестирования классических НПБК: (а) Вероятности ошибок в зависимости от порога принятия, (б) Стабильность ответов НПБК на образы «Чужих» в зависимости от количества ошибочных бит (порога принятия)

При обработке биометрического образа нейросетевым ПБК в режиме ЗНК происходит процесс «распаковки» нейронов – параметры каждого следующего нейрона в цепочке дешифруются. Для получения на выходе ПБК верного ключа пользователя требуется, чтобы все нейроны «проголосовали» правильно. Если хотя бы один нейрон в цепочке выдаст ошибочный бит, это повлечет неверную дешифровку параметров всех последующих нейронов. В свою очередь последующие нейроны будут давать случайные выходы и

возникнет эффект хеширования биометрического образа «Чужого». Классические нейроны следует размещать в начале цепочки, а Байесовские нейроны – в конце (в силу того, что последние в незащищенном виде компрометируют часть ключа). Нейроны Байеса будут надѐжно защищены, если классических нейронов будет много (достаточно 256).

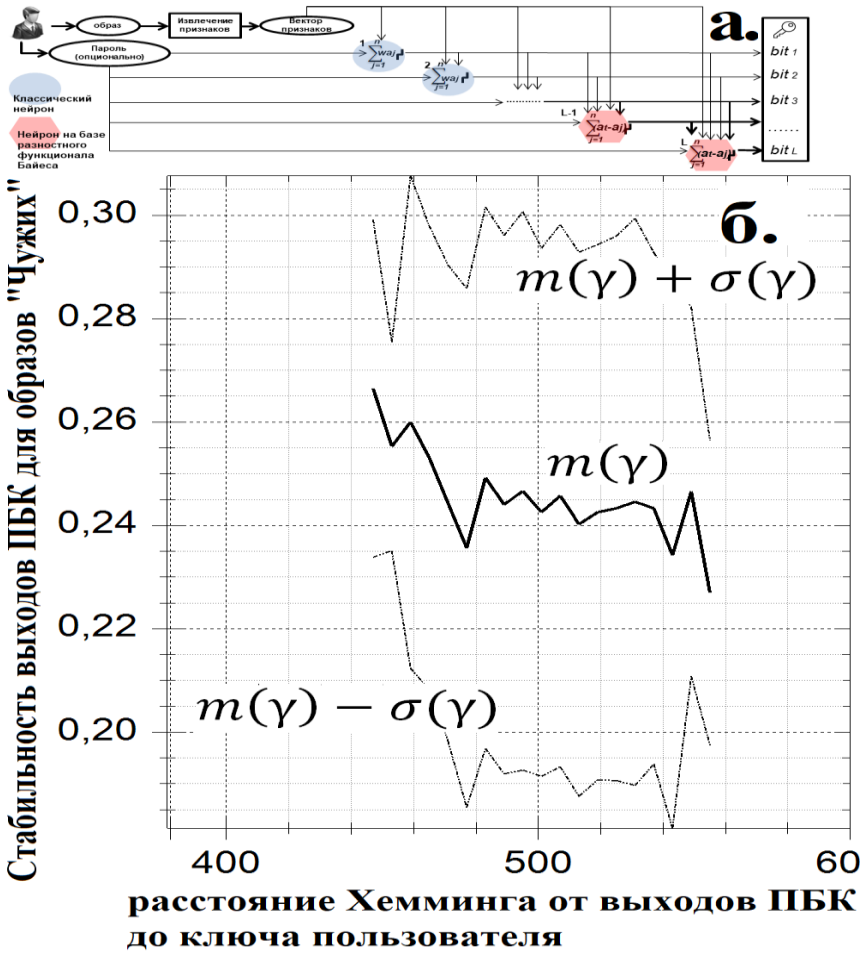


Рис. 3. Механизм ЗНК применительно гибридному НПБК на базе классических и разностных нейронов Байеса: а – Схема ЗНК, б – Стабильность ответов гибридных ПБК при поступлении образов «Чужих» в зависимости от числа ошибочных бит

Классические нейроны показывают хороший результат, если признаки достаточно информативны и слабо коррелированы ($r_{t,j} < 0.5$). Разностные нейроны Байеса обладают почти противоположными свойствами: они ориентированы на обработку сильно зависимых признаков ($r_{t,j} > 0.5$), так как неявно извлекают дополнительную информацию из корреляционной матрицы признаков.

Многоуровневые квантователи позволяют существенно повысить хеширующие свойства разностных Байесовских нейронов. Механизм ЗНК можно успешно применять в гибридных нейронных сетях, сочетая классические нейроны с другими мерами близости. В совокупности эти техники образуют мощный метод верификации образов, при этом данные обученных гибридных нейронных сетей будут надежно защищены от компрометации без применения сторонних криптографических средств (гибридных НПБК может самостоятельно «запаковывать» и «распаковывать» собственные «знания»).

На данный момент не видится значительных препятствий применения описанных методов в другой предметной области, в том числе, их использования для построения ИИ в защищенном исполнении.

Библиографический список

1. Технология использования больших нейронных сетей для преобразования нечетких биометрических данных в код ключа доступа : монография / Б. С. Ахметов, А. И. Иванов, В. А. Фунтиков, А. В. Безяев, Е. А. Малыгина. – Алматы : Изд-во LEM, 2014. – 144 с.

2. Сулавко, А. Е. Высоконадежная двухфакторная биометрическая аутентификация по рукописным и голосовым паролям на основе гибких нейронных сетей / А. Е. Сулавко // Компьютерная оптика. – 2020. – Т. 44, № 1. – С. 82–91. – DOI 10.18287/2412-6179-CO-567.

3. Иванов, А. И. Нейросетевая защита конфиденциальных биометрических образов гражданина и его личных криптографических ключей : монография / А. И. Иванов. – Пенза, 2014. – 57 с.

Для цитирования:

Сулавко, А. Е. Разностные нейроны Байеса с множеством квантователей для высоконадежной аутентификации и защищенного исполнения искусственного интеллекта / А. Е. Сулавко // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 103–111.

УДК 612.087.1; 519.7; 519.66

Е. А. Малыгина

УСОВЕРШЕНСТВОВАНИЕ АЛГОРИТМА ОБУЧЕНИЯ НЕЙРОННОЙ СЕТИ ЗА СЧЕТ ПОДБОРА БЛИЗКИХ СОСТОЯНИЙ ВЕСОВЫХ КОЭФФИЦИЕНТОВ ОБУЧАЕМЫХ НЕЙРОНОВ

Аннотация. Показано, что зная область неопределенности математических ожиданий и среднеквадратических отклонений, можно перебрать возможные близкие состояния весовых коэффициентов обучаемых нейронов, получив тем самым улучшенный вариант стандартного алгоритма обучения.

E. A. Malygina

IMPROVING THE NEURAL NETWORK LEARNING ALGORITHM BY SELECTING CLOSE STATES OF WEIGHTS OF TRAINED NEURONS

Abstract. Knowing the area of uncertainty of mathematical expectations and mid-quadratic deviations, it is possible to sort out the possible close states of weight coefficients of trained neurons, thus obtaining an improved version of the standard learning algorithm.

Известно, что все биометрические параметры личности существенно связаны между собой [1, 2], существующие между контролируемыми параметрами коэффициенты парной корреляции имеют плотности распределения значений, пример одной из которых отображен на рис. 1 [3].

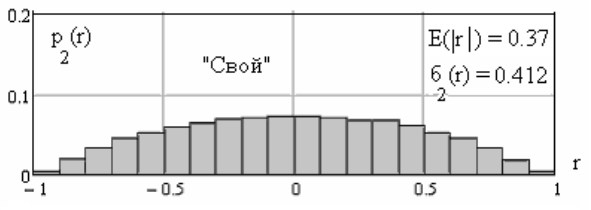


Рис. 1. Пример плотности распределения значений коэффициентов парной корреляции между биометрическими данными образа «Свой»

Из данных рис. 1 видно, что математическое ожидание модулей коэффициентов корреляции составляет 0,37.

Воспользуемся гипотезой нормального распределения многомерных зависимых данных, согласно которой эти данные должны описываться объемом некоторого многомерного гиперэллипса [3].

Отобразить гиперэллипс на плоской бумаге нельзя, в связи с этим рассмотрим двухмерный случай. Пример некоторого двухмерного сечения гиперэллипса дан на рис. 2.

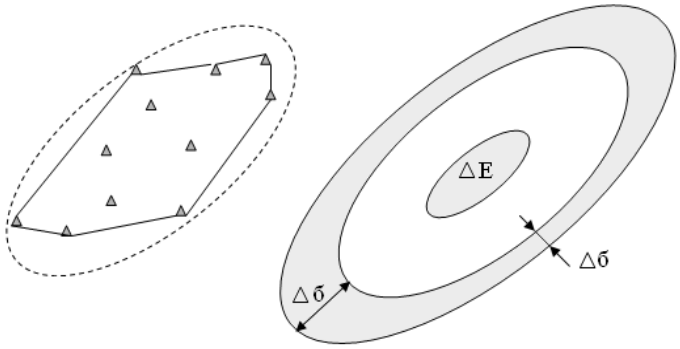


Рис. 2. Одно из сечений гиперэллипса при его дискретном и аналитическом представлении

В левой части рис. 2 представлено одиннадцать проекций 11 примеров обучающей выборки на рассматриваемую гиперплоскость. Очевидно, что попытка оконтурить примеры в обучающей выборке будет давать значительную ошибку приближения реальных данных. Возникает ошибка дискретизации, обусловленная малым числом примеров в представленной выборке.

Очевидно, желание уйти от ошибок дискретизации за счет интегральных процедур вычисления вектора математических ожиданий биометрических параметров $E(\nu_i)$ и их вектора их среднеквадратических отклонений $\sigma(\nu_i)$.

Зная вектора этих статистических моментов, можно рассчитать коэффициенты парной корреляции между нужными нам биометрическими параметрами $r_{i,j}$.

В свою очередь от коэффициента корреляции мы можем перейти к соотношению большой полуоси – a и малой полуоси – b, эллипса, оказавшегося в рассматриваемом сечении:

$$\frac{1+r(\nu_i,\nu_j)}{1-r(\nu_i,\nu_j)}=\frac{a}{b}. \qquad (1)$$

Получается, что, располагая информацией о математическом ожидании двух био-параметров, их среднеквадратических отклонениях и о корреляции между ними, мы можем построить соответ-

ствующий эллипс двухмерного распределения данных образа «Свой» (пунктирная линия в левой части рис. 2).

Данные вне эллипсоида будут соответствовать данным образов «Чужой», а данные внутри эллипсоида будут соответствовать данным образа «Свой».

Если необходимо описать решающее правило многомерного статистического анализа, то необходимо использовать соответствующую квадратичную форму:

$$D^2 = (E(\bar{\nu}) - \bar{\nu})^T \cdot [\rho]^{-1} \cdot (E(\bar{\nu}) - \bar{\nu}), \qquad (2)$$

где $[\rho]^{-1}$ – обратная матрица коэффициентов ковариации, каждый из ее элементов может быть определен через соответствующий коэффициент корреляции:

$$\rho(\nu_i, \nu_j) = \sigma(\nu_i) \cdot \sigma(\nu_j) \cdot r(\nu_i, \nu_j). \qquad (3)$$

Квадратичная форма (2) имеет важное теоретическое значение, однако для практических расчетов она мало применима. Проблема состоит в плохой обусловленности процедуры обращения ковариационных матриц.

На малых выборках из 20 примеров биометрических данных младшие статистические моменты (математические ожидания, среднеквадратические отклонения, коэффициенты корреляции) оцениваются с большими ошибками. Именно по этой причине вместо высокоразмерных классических матричных преобразований (2), необходимо использовать большие искусственные нейронные сети, обучение которых менее чувствительно к ошибкам из-за недостаточного объема исходных данных [4].

Очевидно, что чем больше реальных данных в обрабатываемой выборке, тем точнее получится результат.

То есть путем численного моделирования можно спрогнозировать величину интервала относительной ошибки вычисления математического ожидания $\frac{\Delta E(\nu_i)}{\sigma(\nu_i)}$ как функцию от числа данных в используемой выборке для нормального закона распределения значений. Построенные зависимости при разном значении коэффициента доверия отображены на рис. 3.

Аналогичным образом получено относительное значение ожидаемого интервала ошибки вычисления среднеквадратического отклонения. Кривые связи ошибки среднеквадратического отклонения и размеров исследуемой выборки приведены на рис. 4.

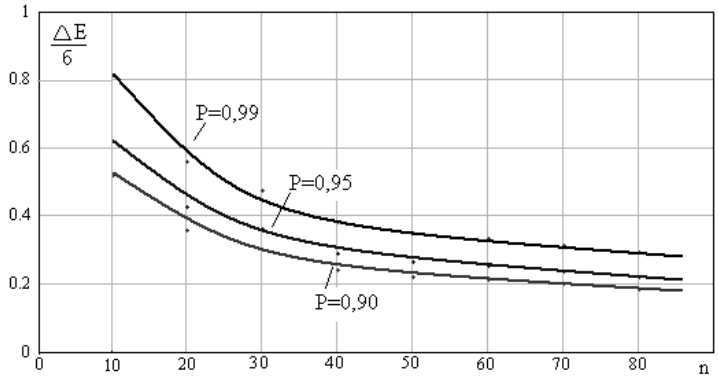


Рис. 3. Номограммы нормированной по среднеквадратическому отклонению интервала ошибок, установленный с доверительной вероятностью 0,99, 0,95 и 0,90

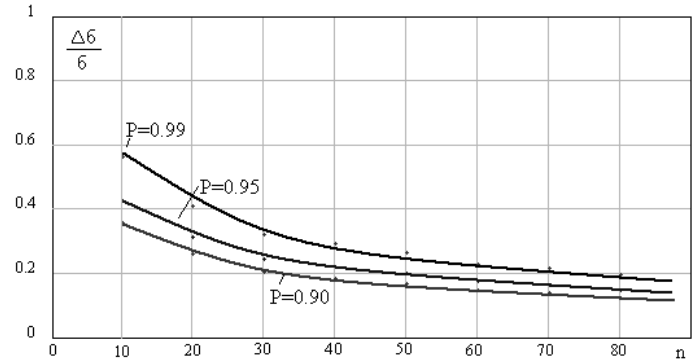


Рис. 4. Номограммы относительной ошибки оценки среднеквадратического отклонения для интервалов, соответствующих доверительной вероятности 0,99, 0,95 и 0,90

Третьим из младших статистических моментов являются коэффициенты корреляции. Связь относительной ошибки оценки значения коэффициентов корреляции биометрических параметров с объемом исходных данных дана на рис. 5.

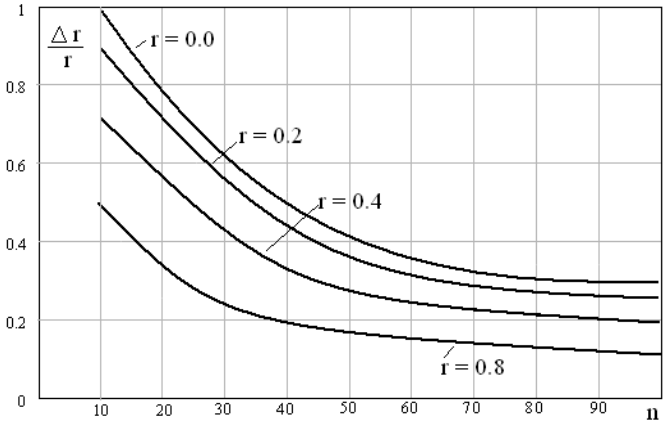


Рис. 5. Номограммы относительной ошибки вычисления коэффициентов корреляции, обусловленная конечным числом примеров в тестовой выборке при доверительной вероятности 0,99

Из данных приведенных выше рисунков видно, что ошибка измерения младших статистических моментов монотонно снижается, однако при малых объемах тестовой (обучающей) выборки ошибки оказываются значительными и составляют порядка 50 % от оцениваемого параметра. Столь значительная величина ошибок делает практически невозможным использование классических квадратичных форм даже относительно низкого порядка.

Например, если попытаться оценить число обусловленности для корреляционных матриц с одинаковыми коэффициентами корреляции r = 0,4 (эта величина близка к среднему значению модулей коэффициентов корреляции, смотри рис. 5), то получим следующие значения последовательности чисел обусловленности:

$$\mathrm{cond}\begin{pmatrix} 1 & 0.4 \\ 0.4 & 1 \end{pmatrix} = 2.762, \qquad \mathrm{cond}\begin{pmatrix} 1 & 0.4 & 0.4 \\ 0.4 & 1 & 0.4 \\ 0.4 & 0.4 & 1 \end{pmatrix} = 4.819, \qquad \mathrm{cond}\begin{pmatrix} 1 & 0.4 & 0.4 & 0.4 \\ 0.4 & 1 & 0.4 & 0.4 \\ 0.4 & 0.4 & 1 & 0.4 \\ 0.4 & 0.4 & 0.4 & 1 \end{pmatrix} = 7.11, \ldots.$$

По мере роста размерности корреляционной матрицы монотонно растет число ее обусловленности. Известно, что коэффициент обусловленности можно рассматривать как коэффициент усиления ошибок исходных данных, то есть даже при попытках решать двухмерные задачи должны получаться результаты с ошибкой 50 % × 2,76 ≈ 138 %.

Именно по этой причине в биометрии нельзя использовать простую и понятную линейную алгебру. Даже низкоразмерные квадратичные формы линейной алгебры начинают эффективно работать при нескольких сотнях примеров, используемых для обучения (обращения матрицы) или для тестирования.

Таким образом, зная область неопределенности математических ожиданий и среднеквадратических отклонений, можно перебрать возможные близкие состояния весовых коэффициентов обучаемых нейронов, получив тем самым улучшенный вариант стандартного алгоритма обучения [5].

Библиографический список

1. Волчихин, В. И. Быстрые алгоритмы обучения нейросетевых механизмов биометрико-криптографической защиты информации / В. И. Волчихин, А. И. Иванов, В. А. Фунтиков. – Пенза : Изд-во ПГУ, 2005. – 276 с.

2. ГОСТ Р 52633.0–2006. Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации. – Москва : Стандартинформ, 2007. – 25 с.

3. Akhmetov, B. Assessing the Level of Uncertainty of small samples of Multidimensional Biological and Biometric Data / B. Akhmetov, A. Ivanov, E. Malygina, S. Kachalin, N. Serikova // International journal of engineering sciences & research technology. – 2014. – № 3 (7). – P. 284–288.

4. Нейросетевая защита персональных биометрических данных / Ю. К. Язов, В. И. Волчихин, А. И. Иванов, В. А. Фунтиков, И. Г. Назаров. – Москва : Радиотехника, 2012. – 157 с.

5. ГОСТ Р 52633.5–2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия – код доступа. – Москва : Стандартинформ, 2012. – 20 с.

Для цитирования:

Малыгина, Е. А. Усовершенствование алгоритма обучения нейронной сети за счет подбора близких состояний весовых коэффициентов обучаемых нейронов / Е. А. Малыгина // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 112–117.

УДК 519.2, 004.622

А. Г. Банных

ВОСЬМИБИТНЫЕ ТАБЛИЦЫ СВЯЗЫВАНИЯ ЭНТРОПИИ 256-БИТНЫХ КОДОВ С МАТЕМАТИЧЕСКИМ ОЖИДАНИЕМ И СТАНДАРТНЫМ ОТКЛОНЕНИЕМ РАССТОЯНИЙ ХЭММИНГА

Аннотация. Рассматриваются таблицы, ориентированные на использование в доверенной вычислительной среде эмбриона доверенного искусственного интеллекта биометрико-нейросетевой аутентификации. Таблицы связывают математическое ожидание и стандартное отклонение расстояний Хэмминга при тестировании обученного нейросетевого преобразователя.

A. G. Banny

EIGHT-BIT ENTROPY BINDING TABLES 256 BIT CODES WITH MATHEMATICAL EXPECTATION AND STANDARD DEVIATION OF HEMMING DISTANCES

Abstract. Tables are considered, focused on the use in a trusted computing environment of the embryo of trusted artificial intelligence biometric-neural network authentication. Tables link mathematical expectation and standard deviation of Hamming distances when testing a trained neural network converter.

Введение

Как правило, после получения длинного личного ключа пользователя или его длинного пароля доступа, запускается некоторый криптографический протокол выполнения аутентификации. Криптографические протоколы принято считать надежными, если энтропия блока шифротекста длиной в 256 бит будет составлять ровно 256 бит. Если энтропия оказывается больше или меньше, то криптосхема защиты информации может оказаться дефектной. Эти соображения с некоторой натяжкой можно перенести на защиту криптографического ключа «нечеткими экстракторами» или размещением его данных в параметрах обученной нейронной сети. Знание энтропии состояний кода ключа оказывается эффективным контрольным параметром при анализе уровня защиты от попыток его подбора.

В случае, если разрядность кода доступа мала, то расчет энтропии этих кодов можно выполнить по Шеннону. В частности, при кодах длинной 16 бит «нечеткого экстрактора» [1, 2] энтропию следует вычислять по следующей формуле:

$$H(x_1, x_2, ..., x_{16}) = -\sum_{i=1}^{65536} p_i \cdot \log_2(p_i), \quad (1)$$

где p_i – вероятность появления одного из 2^{16} = 65536 состояний кодов.

Для того, чтобы оценить вероятность появления 2^{16} состояний кода необходимо иметь базу из 2^{16+4} папиллярных рисунков отпечатков пальцев «Чужой». Собрать базу из почти миллиона рисунков отпечатков пальцев сложно, но технически возможно. По этой причине процедуры вычисления энтропии по Шеннону для «нечетких экстракторов» вполне применимы.

Положение коренным образом меняется, если мы переходим к кодам длинной в 256 бит:

$$H(x_1, x_2, ..., x_{256}) = -\sum_{i=1}^{2^{256}} p_i \cdot \log_2(p_i). \quad (2)$$

Для прямых оценок очень малых вероятностей появления 2^{256} состояний кода потребуется использовать базу из 2^{256+4} биометрических образов «Чужой». Технически невозможно создать и использовать столь большую тестовую базу. Столь значительный объем тестовых примеров нельзя разместить в ограниченном объема памяти малогабаритной малопотребляющей доверенной вычислительной среды.

Переход при вычислениях в пространство расстояний Хэмминга

Для решения проблемы больших тестовых баз и сложных вычислений энтропии по Шеннону в России разработан стандарт ГОСТ Р 52633.5. По рекомендациям этого стандарта требуется применение малых тестовых баз образов «Чужой» объемом от 21 до 64 примеров. Процедура тестирования иллюстрируется рис. 1.

Применение стандарта дает экспоненциальное снижение объема тестовой выборки. Этот выигрыш обусловлен тем, что стандарт рекомендует переходить от анализа обычных кодов в пространство расстояний Хэмминга:

$$h = \sum_{i=1}^{256} \begin{bmatrix} "c_i" \\ \oplus \\ "x_i" \end{bmatrix}, \quad (3)$$

где "c_i" – дискретное состояние i-го разряда кода «Свой»; "x_i" дискретное состояние i-го разряда случайного кода образа «Чужой»; $\oplus$ – операция сложения по модулю 2.

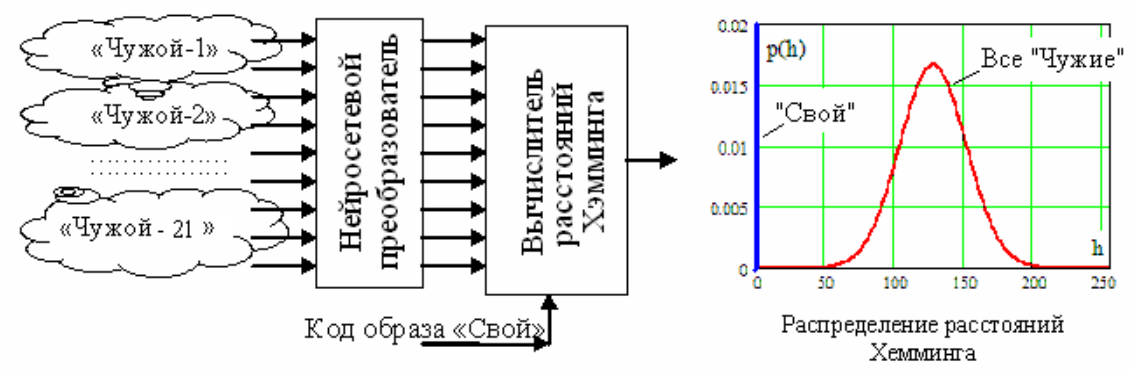


Рис. 1. Тестирование преобразователя биометрия-код при переходе в пространство расстояний Хэмминга

Далее по выборке в 21 опыт мы можем вычислить математическое ожидание – $E(h)$ и стандартное отклонение – $\sigma(h)$. Знание о значениях этих двух статистических моментов позволяет оценить вероятность угадывания кода «Свой» – P_2. Оценка вероятности выполняется в рамках гипотезы нормальности:

$$P_2 \approx \frac{1}{\sigma(h)\sqrt{2\pi}} \int_{-\infty}^{1} \exp\left\{ \frac{-\left(E(h)-u\right)^2}{2\left(\sigma(h)\right)^2} \right\} \cdot du\,. \qquad (4)$$

В этом случае энтропия нейросетевого преобразователя оценивается следующим образом:

$$H("x_1, x_2, ..., x_{256}") \approx -\log_2(P_2)\,. \qquad (5)$$

Применение тройки преобразований (3), (4), (5) позволяет уйти от экспоненциальной вычислительной сложности оценок энтропии длинных кодов по Шеннону [3–5].

Замена вычислений вероятностных интегралов на табличные вычисления

Следует отметить, что вычисление вероятностных интегралов вида (4) требует использования процессоров 64 разрядной сеткой, что трудно реализовать при использовании низкоразрядных и, соответственно, мало потребляющих микроконтроллеров с низкой стоимостью.

Решить проблему можно, если зафиксировать стандартное отклонение расстояний Хэмминга σ(h) и изменять, только математическое ожидание расстояний Хэмминга E(h), то мы получим почти линейную связь с оцениваемой энтропией (рис. 2).

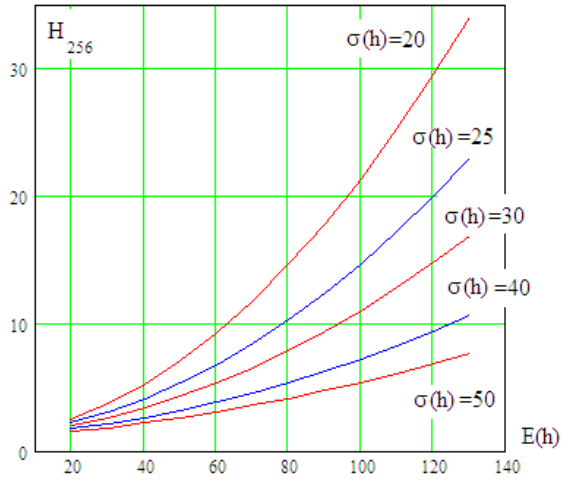


Рис. 2. Почти линейная связь энтропии с математическим ожиданием расстояний Хэмминга при фиксированном стандартном отклонении

Получается, что вычисления можно вычислять таблично. Первая часть таблицы приведена на рис. 3, вторая часть таблицы приведена на рис. 4.

№	E(h)	σ =												
		20	22	24	26	28	30	32	34	36	38	40	42	44
0	30	3.8	3.4	3.1	2.9	2.7	2.6	2.5	2.3	2.2	2.2	2.1	2	2
1	35	4.5	4	3.7	3.4	3.2	3	2.8	2.7	2.5	2.4	2.3	2.3	2.2
2	40	5.3	4.7	4.3	3.9	3.6	3.4	3.2	3	2.8	2.7	2.6	2.5	2.4
3	45	6.2	5.5	4.9	4.5	4.1	3.8	3.6	3.4	3.2	3	2.9	2.8	2.7
4	50	7.1	6.3	5.6	5.1	4.6	4.3	4	3.7	3.5	3.3	3.2	3	2.9
5	55	8.2	7.1	6.4	5.7	5.2	4.8	4.4	4.2	3.9	3.7	3.5	3.3	3.2
6	60	9.3	8.1	7.2	6.4	5.8	5.3	4.9	4.6	4.3	4.1	3.8	3.6	3.5
7	65	10.5	9.1	8	7.2	6.5	5.9	5.5	5.1	4.7	4.4	4.2	4	3.8
8	70	11.8	10.2	9	8	7.2	6.5	6	5.6	5.2	4.8	4.6	4.3	4.1
9	75	13.2	11.3	9.9	8.8	7.9	7.2	6.6	6.1	5.7	5.3	5	4.7	4.4
10	80	14.6	12.6	11	9.7	8.7	7.9	7.2	6.6	6.1	5.7	5.4	5.1	4.8
11	85	16.2	13.9	12.1	10.7	9.5	8.6	7.9	7.2	6.7	6.2	5.8	5.5	5.2
12	90	17.8	15.2	13.2	11.7	10.4	9.4	8.5	7.8	7.2	6.7	6.3	5.9	5.5
13	95	19.6	16.7	14.4	12.7	11.3	10.2	9.2	8.5	7.8	7.2	6.7	6.3	5.9
14	100	21.4	18.2	15.7	13.8	12.3	11	10	9.1	8.4	7.8	7.2	6.8	6.4
15	105	23.3	19.7	17.1	14.9	13.3	11.9	10.8	9.8	9	8.3	7.7	7.2	6.8
16	110	25.2	21.4	18.5	16.1	14.3	12.8	11.6	10.5	9.7	8.9	8.3	7.7	7.2
17	115	27.3	23.1	19.9	17.4	15.4	13.8	12.4	11.3	10.3	9.5	8.8	8.2	7.7
18	120	29.5	24.9	21.4	18.7	16.5	14.7	13.3	12.1	11	10.2	9.4	8.8	8.2
19	125	31.7	26.8	23	20	17.7	15.8	14.2	12.9	11.8	10.8	10	9.3	8.7
20	130	34.1	28.7	24.6	21.4	18.9	16.8	15.1	13.7	12.5	11.5	10.6	9.9	9.2
21	135	36.5	30.7	26.3	22.9	20.2	17.9	16.1	14.6	13.3	12.2	11.3	10.5	9.7
22	140	39	32.8	28.1	24.4	21.5	19.1	17.1	15.5	14.1	12.9	11.9	11.1	10.3
23	145	41.6	35	29.9	26	22.8	20.3	18.2	16.4	14.9	13.7	12.6	11.7	10.9
24	150	44.3	37.2	31.8	27.6	24.2	21.5	19.2	17.4	15.8	14.5	13.3	12.3	11.5

Рис. 3. Первая часть таблицы связи значений энтропии с математического ожидания расстояний Хэмминга E(h) и стандартным отклонением σ(h)

№	E(h)	σ =													
		46	48	50	52	54	56	58	60	62	64	66	68	70	72
0	30	1.9	1.9	1.8	1.8	1.8	1.7	1.7	1.7	1.6	1.6	1.6	1.9	1.6	1.6
1	35	2.1	2.1	2	2	1.9	1.9	1.8	1.8	1.8	1.7	1.7	2.1	1.7	1.7
2	40	2.3	2.3	2.2	2.1	2.1	2	2	2	1.9	1.9	1.8	2.3	1.8	1.8
3	45	2.6	2.5	2.4	2.3	2.3	2.2	2.2	2.1	2.1	2	2	2.6	2	1.9
4	50	2.8	2.7	2.6	2.5	2.5	2.4	2.3	2.3	2.2	2.2	2.1	2.8	2.1	2
5	55	3.1	2.9	2.8	2.7	2.7	2.6	2.5	2.4	2.4	2.3	2.2	3.1	2.2	2.2
6	60	3.3	3.2	3.1	3	2.9	2.8	2.7	2.6	2.6	2.5	2.4	3.3	2.4	2.3
7	65	3.6	3.5	3.3	3.2	3.1	3	2.9	2.8	2.7	2.7	2.5	3.6	2.5	2.5
8	70	3.9	3.7	3.6	3.4	3.3	3.2	3.1	3	2.9	2.8	2.7	3.9	2.7	2.6
9	75	4.2	4	3.8	3.7	3.6	3.4	3.3	3.2	3.1	3	2.9	4.2	2.9	2.8
10	80	4.5	4.3	4.1	4	3.8	3.7	3.5	3.4	3.3	3.2	3	4.5	3	2.9
11	85	4.9	4.6	4.4	4.2	4.1	3.9	3.8	3.6	3.5	3.4	3.2	4.9	3.2	3.1
12	90	5.2	5	4.7	4.5	4.3	4.2	4	3.9	3.7	3.6	3.4	5.2	3.4	3.3
13	95	5.6	5.3	5.1	4.8	4.6	4.4	4.3	4.1	3.9	3.8	3.6	5.6	3.6	3.5
14	100	6	5.7	5.4	5.1	4.9	4.7	4.5	4.3	4.2	4	3.8	6	3.8	3.7
15	105	6.4	6	5.7	5.5	5.2	5	4.8	4.6	4.4	4.3	4	6.4	4	3.9
16	110	6.8	6.4	6.1	5.8	5.5	5.3	5.1	4.9	4.7	4.5	4.2	6.8	4.2	4.1
17	115	7.2	6.8	6.5	6.1	5.8	5.6	5.3	5.1	4.9	4.7	4.4	7.2	4.4	4.3
18	120	7.7	7.2	6.9	6.5	6.2	5.9	5.6	5.4	5.2	5	4.6	7.7	4.6	4.5
19	125	8.2	7.7	7.3	6.9	6.5	6.2	5.9	5.7	5.5	5.2	4.9	8.2	4.9	4.7
20	130	8.6	8.1	7.7	7.3	6.9	6.6	6.3	6	5.7	5.5	5.1	8.6	5.1	4.9
21	135	9.1	8.6	8.1	7.6	7.3	6.9	6.6	6.3	6	5.8	5.4	9.1	5.4	5.2
22	140	9.6	9	8.5	8.1	7.6	7.3	6.9	6.6	6.3	6.1	5.6	9.6	5.6	5.4
23	145	10.2	9.5	9	8.5	8	7.6	7.3	6.9	6.6	6.4	5.9	10.2	5.9	5.7
24	150	10.7	10	9.4	8.9	8.4	8	7.6	7.3	6.9	6.7	6.1	10.7	6.1	5.9

Рис. 4. Вторая часть таблицы связи значений энтропии
с математического ожидания расстояний Хэмминга E(h)
и стандартным отклонением σ(h)

К сожалению, правая часть таблицы рис. 4 имеет значительные ошибки из-за того, что гипотеза нормальности (4) для этих данных не выполняется. Соответствующие распределения для разных значений математических ожиданий и стандартных отклонений приведены на рис. 5.

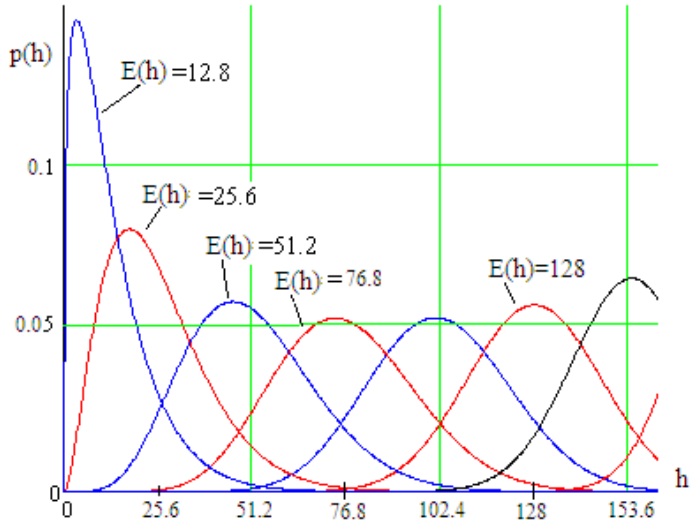


Рис. 5. Распределение расстояний Хэмминга для одинаково коррелированных данных $\tilde{r} = 0,33$ и монотонно изменяющихся значениях математического ожидания

Библиографический список

1. Juels, A. A Fuzzy Commitment Scheme / A. Juels, M. Wattenberg // Proc. ACM Conf. Computer and Communications Security (Singapore, 1–4 November, 1999). – Singapore, 1999. – P. 28–36.

2. Ramírez-Ruiz, J. Cryptographic Keys Generation Using FingerCodes / J. Ramírez-Ruiz, C. Pfeiffer, J. Nolazco-Flores // Advances in Artificial Intelligence – IBERAMIA-SBIA 2006 : Ibero-American Conf. on Artificial Intelligence (Ribeirão Preto, October 23–27, 2006). – Ribeirão Preto, Brazil, 2006. – P. 178–187.

3. Малыгин, А. Ю. Быстрые алгоритмы тестирования нейросетевых механизмов биометрико-криптографической защиты информации / А. Ю. Малыгин, В. И. Волчихин, А. И. Иванов, В. А. Фунтиков. – Пенза : Изд-во ПГУ, 2006. – 161 с.

4. Язов, Ю. К. Нейросетевая защита персональных биометрических данных / Ю. К. Язов, В. И. Волчихин, А. И. Иванов, В. А. Фунтиков, И. Г. Назаров. – Москва : Радиотехника, 2012 . – 157 с.

5. Ахметов, Б. С. Алгоритмы тестирования биометрико-нейросетевых механизмов защиты информации : монография / Б. С. Ахметов, В. И. Волчихин, А. И. Иванов, А. Ю. Малыгин. – Москва : А. Ю. Малыгин. – Казахстан, Алматы : КазНТУ им. Сатпаева, 2013. – 152.

Для цитирования:

Банных, А. Г. Восьмибитные таблицы связывания энтропии 256-битных кодов с математическим ожиданием и стандартным отклонением расстояний Хэмминга / А. Г. Банных // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 118–123.

УДК 004.056.55

Н. А. Постников

ОБЗОР МЕТОДОВ ОПТИМИЗАЦИИ КРИПТОГРАФИЧЕСКОГО ПРОГРАММНОГО МОДУЛЯ

Аннотация. Рассматриваются методы оптимизации криптографического программного модуля, которые позволяют увеличить скорость выполнения криптографических преобразований. Актуальность работы заключается в возросшем объеме обрабатываемых данных, снижении скорости обработки данных и практическом применении методов оптимизации при использовании различных криптографических алгоритмов.

N. A. Postnikov

OVERVIEW OF HOW TO OPTIMIZE A CRYPTOGRAPHIC SOFTWARE MODULE

Abstract. The article is dedicated to reviewing the methods of optimization of the cryptographic software module, which allow to increase the speed of cryptographic transformations. The relevance of the work lies in the increased amount of data being processed, the reduction in data processing speed and the practical application of optimization methods using different cryptographic algorithms.

Сокращения:
- ОС – Операционная система;
- ПО – Программное обеспечение;
- ЭВМ – Электронно-вычислительная машина;
- ОЗУ – Оперативное запоминающее устройство;
- ПЗУ – Постоянное запоминающее устройство.

Введение

Несмотря на ежегодное увеличение производительности центральных процессоров [1], существует проблема достижения на ЭВМ требуемой скорости выполнения криптографических преобразований данных большого объема актуальна. Устранить проблему могут методы оптимизации, рассматриваемые в статье.

Методы оптимизации

Оптимизация ПО – это обработка, связанная с переупорядочиванием и изменением операций (в том числе и криптографических) в компилируемой программе с целью получения более эффективной результирующей объектной программы. Оптимизация

выполняется на этапах подготовки к генерации и непосредственно при генерации объектного кода [2].

Оптимизация ПО может проводиться, как и вручную, так и автоматизировано. В последнем случае оптимизатором может быть, как отдельное программное средство, так и используемый компилятор, который преобразует исходный код в набор машинных команд. Кроме того, следует отметить, что современные процессоры могут оптимизировать порядок выполнения инструкций кода [3].

Существуют такие понятия, как высокоуровневая и низкоуровневая оптимизация. Высокоуровневые оптимизации в большинстве проводятся разработчиком, который, оперируя абстрактными сущностями (функциями, процедурами, классами и т.д.) и представляя себе общую модель решения задачи, может изменить архитектуру ПО [4]. Низкоуровневая оптимизация производится на этапе превращения исходного кода в набор машинных команд, и зачастую именно этот этап подвергается автоматизации.

При оптимизации кода вручную необходимо знать в каком именно участке ПО (адресное пространство, функции, циклы и алгоритмические операции) проводить оптимизации. Использование профилировщиков (профайлеров), которые динамически исследуют характер поведения ПО, выявляют частоты и продолжительность вызовов функций, а также создают граф вызовов исследуемых функций, могут выявить участки ПО, требующие оптимизации. Пример графа вызовов профайлера приведен на рис. 1.

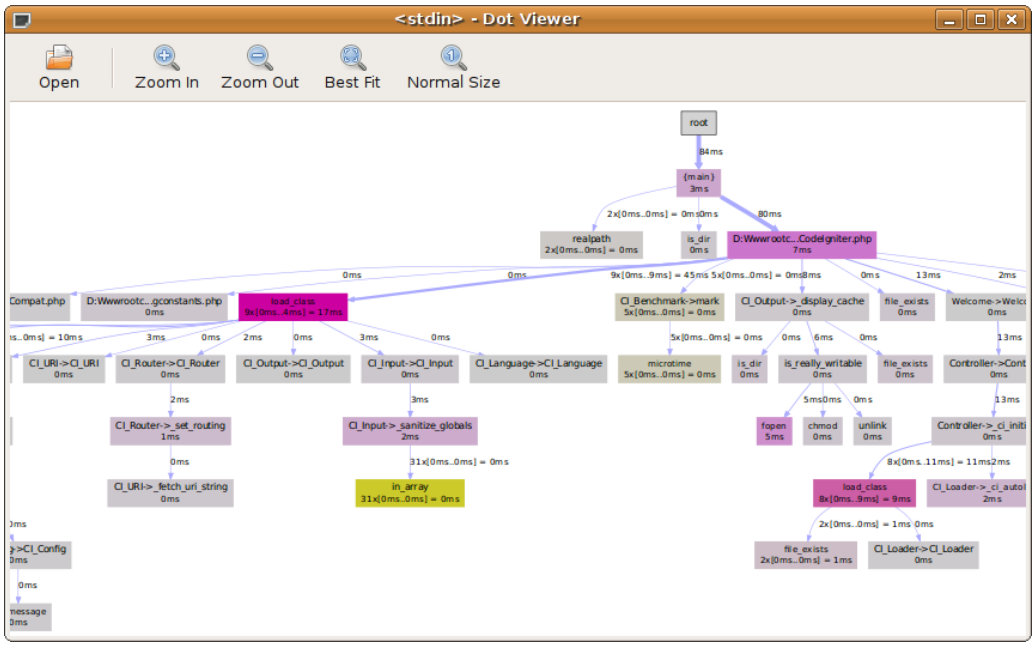


Рис. 1. Граф вызовов профайлера

– Использование оптимизирующего компилятора. Для создания эффективного ПО можно воспользоваться встроенными средствами оптимизации компилятора. Например, компилятору GCC возможно передать один из следующих флагов оптимизации:

– O0 (Отключение оптимизаций, основная цель: высокая скорость компиляции и предсказуемость результатов отладки. Эта опция задействована по умолчанию.);

– O1 (Мягкая оптимизация, некоторое увеличение времени компиляции, возможно ощутимое увеличение памяти для больших функций. Включены оптимизации, которые должны одновременно уменьшать общую занимаемую память и ускорять код.);

– 02 (Включает практически все доступные методы оптимизации, не ухудшающие один из двух показателей двоичного кода – память или скорость – за счѐт другого. Не включает развертку циклов и автоматическое встраивание функций.);

– O3 (Наиболее «агрессивная» оптимизация, включаются развертку циклов и автоматическое встраивание функций.);

– Os (Оптимизация размера программы. Включаются те опции из набора -O2, которые обычно не увеличивают объѐм кода. Также применяются некоторые другие оптимизации, направленные на снижение его объѐма) [5].

– Использование языка ассемблера. Высокоуровневые языки программирования не всегда позволяют эффективно выполнить некоторые операции, необходимые для выполнения криптографических преобразований, однако низкоуровневая реализация криптографического алгоритма снижает переносимость ПО, усложняет ее модификацию и сопровождение с учетом специфики конкретной микроархитектуры процессора. Использование программных вставок на языке ассемблера является компромиссом и позволяет наиболее оптимально реализовать требуемый криптографический алгоритм, особенно в критичных для производительности частях. Например, большинство пакетов ОС Linux с ассемблерными вставками работают на других микроархитектурах без дополнительного портирования [6], так как ассемблерный код используется только для оптимизации и предусматривает наличие замены на C/C++ или связан с реализацией дополнительных возможностей, что не блокирует сборку приложения на других микроархитектурах процес-

сора. Для повышения быстродействия отдельных процедур следует придерживаться следующих рекомендаций:

– избегать инструкции, содержащие 4 и более микроопераций;

– избегать инструкции, состоящие из 7 и более байт;

– избегать инструкции косвенной передачи управления;

– располагать команды с учетом возможности их «спаривания» (т.е. параллельного выполнения в обоих конвейерах) и минимизации простоев конвейера;

– выполнять регистровую оптимизацию (минимизировать количество обращений к памяти);

– вход в цикл следует выравнивать по 16-байтным границам;

– метки условных переходов выравнивать не стоит;

– метки безусловных переходов следует выравнивать по 16-байтным границам;

– наиболее вероятные ветви программы необходимо располагать непосредственно после команды перехода;

– массивы и структуры данных, размер которых кратен 32 байтам, следует выравнивать по 32-байтным границам;

– массивы и структуры данных, размер которых не кратен 32 байтам, следует выравнивать по 16-байтным границам;

– 2-байтные данные должны целиком содержаться внутри выравненного двойного слова;

– 4-х и 8-ми байтные данные должны быть выровнены, соответственно, по 4-х и 8-ми байтной границе [7].

Высокоуровневые методы оптимизации

– Оптимизация циклов. Большая часть времени исполнения криптографических операций ПО приходится на циклы, таким образом применение техник оптимизации циклов позволит увеличить быстродействие ПО [8]. Для оптимизации циклов используются следующие методы:

– вынесение инвариантных вычислений из циклов (вынесение тех операций, операнды которых не изменяются);

– замена операций с индуктивными переменными (изменение сложных операций с переменными, значения которых в процессе выполнения цикла образуют арифметическую прогрессию, на более простые операции);

– слияние и развертывание циклов (слияние двух вложенных циклов в один и замена цикла на линейную последовательность операций) [9].

– Сокращение количества обращений к внутренней памяти. Самая быстрая память ЭВМ – регистры процессора. Гораздо большей по объѐму, но заметно медленнее является внутренняя память (ОЗУ и ПЗУ). Соответственно, обращение к внутренней памяти для чтения или записи занимает больше времени, чем работа с регистрами процессора. Иерархия памяти ЭВМ представлена на рис. 2.

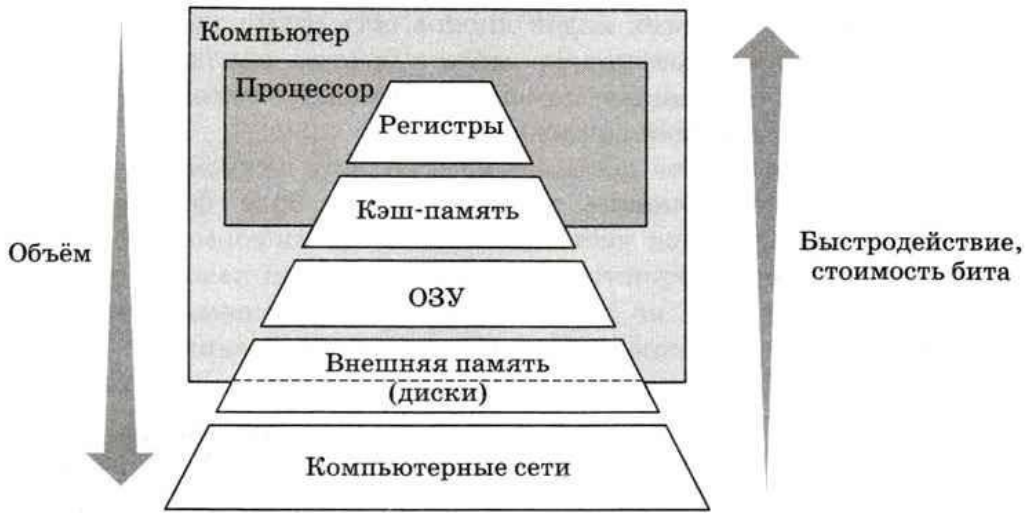


Рис. 2. Иерархия памяти ЭВМ

Быстродействие процессоров значительно выше, чем скорость работы ОЗУ, поэтому процессору приходится ожидать, пока из оперативной памяти произойдет отправление данных [10]. Для оптимизации необходимо ввести временную локальную переменную регистрового типа данных, в которую будет производиться запись промежуточных вычислений, а только через какое-то время произвести запись из этой переменной в оперативную память [11].

– Оптимизация хранения данных. В процессорах микроархитектуры ×86 и ARM данные не могут находиться в произвольной ячейке памяти. Каждый тип данных, кроме однобайтового, требует выравнивания, так как он может начинаться с любого адреса. Однако двухбайтовый тип данных должен начинаться только с четного адреса, четырехбайтный – с адреса, кратного 4, восьмибайтные – с адреса, кратного 8. Выравнивание ускоряет доступ к памяти за счет генерации кода, в котором на чтение и запись ячейки памяти требуется по одной машинной инструкции. Данные однобайтового типа данных всегда требуют одинакового количества машинных инструкций для доступа, поэтому для них нет предпочтительного выравнивания. Также, время доступа к иерархически организован-

ной памяти уменьшается благодаря сокращению количества обращений к оперативной памяти и совмещению обработки текущего фрагмента программы, и загрузке следующего из оперативной памяти в текущую буферную [12]. Пример выравнивания данных четырехбайтового типа данных приведен на рис. 3.

Адрес	0	1	2	3	4	5	6	7	8	9
Данные	Выровненные данные					Не выровненные данные				

Рис. 3. Пример выравнивания данных

– Свѐртка констант – часто используемый в современных компиляторах метод оптимизации, уменьшающий избыточные вычисления, путѐм замены константных выражений и переменных на их значения [13]. Прежде всего, упрощаются константные выражения, содержащие числовые значения. Также могут быть упрощены выражения, содержащие никогда не изменяемые переменные или переменные, объявленные как константы. Пример оптимизации кода на языке программирования C/C++ путем свѐртки констант приведен на рис. 4.

```c
//До оптимизации
#include <stdlib.h>

int main(int argc, char **argv)
{
    struct point
    {
        int x;
        int y;
    } p;
    int a = 32*32;
    int b = 32*32*4;
    long int c;
    //
    c = (a + b) * (4*4*sizeof(p) - 2 + 32);
    return 0;
}
```

```c
//После оптимизации
#include <stdlib.h>

int main(int argc, char **argv)
{
    struct point
    {
        int x;
        int y;
    } p;
    int a = 1024; // Свёрнуто из 32 * 32
    int b = 4096; // Свёрнуто из 32 * 32 * 4
    long int c;
    // 16 = 4*4, 30 = -2 + 32
    c = (a + b) * (16*sizeof(p) + 30);
    return 0;
}
```

Рис. 4. Свѐртка констант

– Распараллеливание вычислений. Параллельное выполнение криптографических операций позволяет повысить производительность криптографического программного модуля. Существует множество подходов к организации параллельных вычислений: многопоточное выполнение на одном физическом процессорном ядре, параллельное выполнение на нескольких процессорных ядрах, входящих в состав одного процессора, параллельное выполнение на множестве процессоров, которые могут находиться в физически разных компьютерах, объединенных в сеть. Но распарал-

леливание не всегда можно применить. Например, не все режимы работы блочных шифров допускают эффективное распараллеливание. Режимы (electronic codebook – электронной кодовой книги, режим простой замены) и CTR (counter mode – режим счетчика), благодаря возможности обрабатывать блоки открытого текста независимо друг от друга, могут быть легко распараллелены, а в режимах CBC (cipher block chaining – режим сцепления блоков шифротекста) и CFB (cipher feedback – режим обратной связи по шифротексту) хорошо распараллеливается только процедура расшифрования. Режим OFB (output feedback – режим обратной связи по выходу) не поддается распараллеливанию, но эффективность шифрования и расшифрования может быть повышена другими способами [14]. Также возможны вычисления хэш-функций, характеризующихся возможностью распараллеливания вычислений. В поиске такого решения будем опираться от вычислительной модели схемы Меркла – Дамгарда [15], которую можно описать следующим образом: на вход хэш-функции поступают данные, которые разбиваются на блоки фиксированной длины. Метод работает с блоками по очереди с помощью функции сжатия, каждый раз принимая входной блок с выходным от предыдущего прохода.

– Оптимизация с использованием алгоритмических методов. Примером являются такие методы, как алгоритм умножения Карацубы [16] для задач, связанных с арифметикой длинных чисел [17], Алгоритм Фюрера [18], оптимизация вычислений посредством заранее просчитываемых таблиц вспомогательных значений [19], применения методов факторизации чисел [20] и др. Пример вычисления алгоритмом умножения Карацубы представлен на рис. 5. Все вышеперечисленные методы позволяют добиться увеличения быстродействия криптографических операций. К минусам данного подхода оптимизации можно отнести повышенную сложность при программной реализации.

7641 * 8512

A = 76 * 85 = 6460;

B = 41 * 12 = 492;

C = (76 + 41) * (85 + 12) = 117 * 97 = 11349;

A * 10000 + (C – A – B) * 100 + B = 64600000 + 439700 + 492 = **65040192**;

Рис. 5. Пример вычисления алгоритмом умножения Карацубы

Вывод

В статье был произведен обзор методов оптимизации криптографического программного модуля, которые позволяют увеличить скорость выполнения криптографических преобразований.

Библиографический список

1. Рейтинг производительности процессоров 2020. Тесты Intel и AMD. – URL: https://benchmarkdb.ru/cpu/ (дата обращения: 31.05.2018).

2. Хэш-функции и хэш-адресация // В помощь Веб-Мастеру. – URL: https:// wm-help.net/lib/b/book/1493769350/105 (дата обращения: 31.05.2020).

3. Техника оптимизации программного кода. Оптимизация программного кода. Основные возможности оптимизации кода программистом и компилятором. – URL: https://printius.ru/tehnika-optimizacii-programmnogo-koda-optimizaciya-programmnogo-koda-osnovnye-vozmozhnosti-optimizac.html (дата обращения: 31.05.2020).

4. Оптимизация кода. – URL: https://www.viva64.com/ru/t/0084/ (дата обращения: 31.05.2020).

5. Использование GCC. – URL: staff.mmcs.sfedu.ru/~ulysses/IT/C++/using_gcc.html (дата обращения: 31.05.2020).

6. Анализ использования ассемблерных вставок в коде открытых проектов. – URL: https://www.opennet.ru/opennews/art.shtml?num=36551 (дата обращения: 31.05.2020).

7. Исследование методов оптимизации криптографических операций. – URL: https://docplayer.ru/28495281-Issledovanie-metodov-optimizacii-kriptograficheskih-operaciy.html (дата обращения: 31.05.2020).

8. Введение в технику оптимизации циклов. – URL: https://habr.com/ru/post/124910/ (дата обращения: 31.05.2020).

9. Системы программирования. Основные сведения о компиляции. – URL: http://mf.grsu.by/UchProc/livak/b_osnovy/opl_4.htm (дата обращения: 31.05.2020).

10. Процессор. Память. Устройства ввода и вывода. – URL: https://xn----7sbbfb7a7aej.xn--p1ai/informatika_10_136_pol/informatika_materialy_zanytii_10_136_pol_41_13.html (дата обращения: 31.05.2018).

11. Оптимизация кода: процессор. – URL: https://habr.com/ru/post/309796/ (дата обращения: 31.05.2020).

12. Искусство упаковки структур в C. – URL: https://tproger.ru/translations/art-of-structure-packing/ (дата обращения: 31.05.2020).

13. Steven, S. Muchnick. Advanced Compiler Design and Implementation / S. Steven. – 5th ed. – San Francisco : Morgan Kaufmann Publishers, 1997. – 856 p.

14. Симонова, О. Н. Особенности оценки качества и оптимизации алгоритмов симметричного шифрования / О. Н. Симонова // Молодой ученый. – 2016. – № 9 (113). – С. 79–81. – URL: https://moluch.ru/archive/113/29457/ (дата обращения: 31.05.2020).

15. Структура Меркла – Дамгора. – URL: https://ru.wikipedia.org/wiki/%D0%A1%D1%82%D1%80%D1%83%D0%BA%D1%82%D1%83%D1%80%D0%B0_%D0%9C%D0%B5%D1%80%D0%BA%D0%BB%D0%B0_%E2%80%94_%D0%94%D0%B0%D0%BC%D0%B3%D0%BE%D1%80%D0%B0 (дата обращения: 31.05.2020).

16. Умножение длинных чисел методом Карацубы. – URL: https://habr.com/ru/post/124258/ (дата обращения: 31.05.2020).

17. Длинная арифметика. – URL: https://ru.wikipedia.org/wiki/%D0%94%D0%BB%D0%B8%D0%BD%D0%BD%D0%B0%D1%8F_%D0%B0%D1%80%D0%B8%D1%84%D0%BC%D0%B5%D1%82%D0%B8%D0%BA%D0%B0 (дата обращения: 31.05.2020).

18. Алгоритм Фюрера. – URL: https://ru.wikipedia.org/wiki/%D0%90%D0%BB%D0%B3%D0%BE%D1%80%D0%B8%D1%82%D0%BC_%D0%A4%D1%8E%D1%80%D0%B5%D1%80%D0%B0 (дата обращения: 31.05.2020).

19. Электронная версия журнала «Информационные технологии в управлении и экономике». – URL: https://docplayer.ru/35442491-Elektronnaya-versiya-zhurnala-razmeshchena-na-sayte-i.html (дата обращения: 31.05.2020).

20. Факторизация целых чисел и криптография с открытым ключом. – URL: http://www.itlab.unn.ru/Uploads/coaChapter12.pdf (дата обращения: 31.05.2020).

Для цитирования:

Постников, Н. А. Обзор методов оптимизации криптографического программного модуля / Н. А. Постников // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 124–132.

УДК 004.934; 004.056.523

А. А. Афанасьев

ИДЕНТИФИКАЦИЯ ДИКТОРА ПРИ ИСПОЛЬЗОВАНИИ ПАРАМЕТРОВ ЭМПИРИЧЕСКОЙ ПЛОТНОСТИ РАСПРЕДЕЛЕНИЯ ХАРАКТЕРИСТИК РЕЧЕВОГО СИГНАЛА КОНТРОЛЬНОЙ ФРАЗЫ

Аннотация. Рассмотрена задача идентификации абонента при доступе к ресурсам инфокоммуникационной системы. В качестве канала биометрической информации о пользователе предлагается использовать параметры эмпирического распределения характеристик речевого сигнала контрольной фразы. В качестве таких параметров предлагается применение мел-кепстральных коэффициентов контрольной фразы речевого сигнала легитимного диктора.

A. A. Afanasev

IDENTIFICATION OF THE ANNOUNCER AT USING PARAMETERS OF EMPIRICAL FREQUENCY CURVE OF THE CONTROL PHRASE SPEECH SIGNAL CHARACTERISTICS

Abstract. The problem of subscriber identification at access to resources of an infocommunication system is considered. As a link of biometric information of the user it is offered to use parameters of empirical distribution of the control phrase speech signal characteristics. As such parameters application mel-frequency cepstral coefficients of a speech signal control phrase of legitimate announcer is offered.

Во многих прикладных приложениях инфокоммуникационных систем первоначально ставится задача идентификации абонента с целью предоставления ему доступа к ее ресурсам. В качестве одного из каналов биометрической информации о пользователе выступает его голос и параметры распределения характеристик речевого сигнала (РС) контрольной фразы.

Задача идентификации заключается в определении принадлежности представленного голоса диктора (объекта) тому или иному классу из множества известных классов, основываясь на векторе значений параметров, вычисленных по данному речевому сигналу.

Информацию о связи между значениями признаков объекта и его принадлежностью к определенному классу система извлекает из обученной совокупности объектов, для которой известны как значения параметров, так и классы. При этом задачей такой системы будет являться построение решающего правила $R^P : \{X\} \to \{A\}$, которое позволило бы распознать класс скрытого объекта $S \in \{\Omega\}$, опираясь на его образ $\{X\}$ в пространстве наблюдений. Доступная системе информация о функциях $R(S)$ и R^P, составляющих вместе с множествами $\{\Omega\}$, $\{A\}$ и $\{X\}$ первичную модель источника данных, ограничивается результатами наблюдений над конечным числом объектов S_j, $j = 1,\ldots,N$, составляющих так называемую обучающую совокупность. Каждый объект S_j в обучающей совокупности представлен номером своего класса и образом $\{X\}$.

Задача идентификации состоит в выделении одного диктора по наблюдаемому голосу из множества известных системе (наблюдателю) дикторов (множество $\{\Omega\}$). В зависимости от наличия во множестве $\{\Omega\}$ особого элемента, соответствующего решению «диктор неизвестен», задачу идентификации разделяют на открытую (решение «диктор неизвестен» во множестве $\{\Omega\}$ присутствует) и закрытую (решение «диктор неизвестен» во множестве $\{\Omega\}$ отсутствует).

Система идентификации характеризуется 2-мя вероятностями: ошибкой «пропуск цели" – $P_{\text{пц}}$ (ошибка первого рода) это принятие «чужого» диктора за легитимного, которого необходимо отобрать и ошибкой «ложное срабатывание» – $P_{\text{лс}}$, это отказ в доступе легитимному диктору. Система может перестраиваться таким образом, что ошибки одного рода могут быть уменьшены за счет увеличения ошибок другого рода (даже при сохранении всех других факторов, влияющих на вероятность ошибки: длительность и характер речевого сообщения, помехи и т.п.). Поиск путей снижения вероятности ошибки первого рода вызвал необходимость анализа существующих методов и подходов, используемых при автоматической идентификации диктора. Рассмотрим общую схему идентификации диктора, представленную на рис. 1 [1].

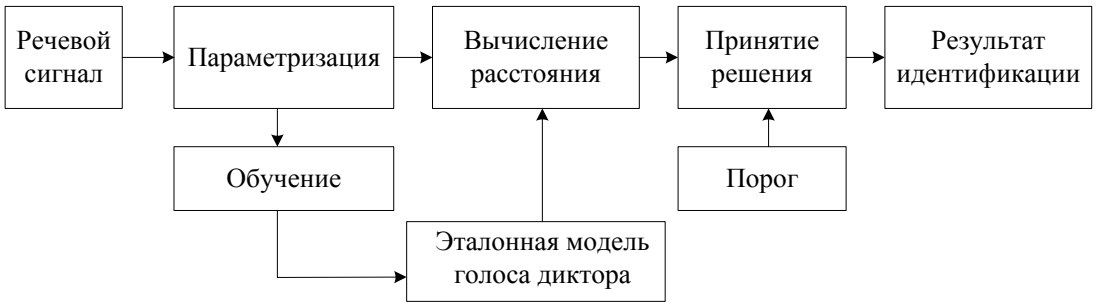


Рис. 1. Общая схема идентификации диктора

Изменение соотношения ошибок достигается за счет изменения порога принятия решения.

Часто для описания модели диктора используется математический аппарат на основе гауссовых смесей, при этом параметрами модели является набор параметров гауссовой смеси, аппроксимирующей плотность вероятности $f_i(x)$, $\Theta_i = \{ f_i(x)\} = \{\alpha_k \mu_k, \Sigma_k\}$, где α_k – вес *k*-го гауссиана, сумма гауссиан удовлетворяет условию $\sum_{k=1}^{N} \alpha_k = 1$; μ_k – вектор математического ожидания *k*-го гауссиана; Σ_k – ковариационная матрица *k*-го гауссиана. Для вычисления параметров гауссовой смеси используются *EM* – алгоритм [1]. Для модели, основанной на векторном квантовании, параметрами модели является набор векторов, характеризующих распределение векторов $\bar{X}$, т.е. матрица, содержащая *n* векторов $\{y_1 \ldots y_m\}$ размерностью *m*, и положительно определенную ковариационную матрицу W^i, получаемую из матрицы векторов $\Theta_i = \left(\{y_1 \ldots y_m\}^n, W^i \right)$ [1].

Для оценки функционирования текстонезависимой системы идентификации диктора также используют такой показатель как *вероятность ошибки идентификации* (эквивалентная ошибка) или обратный ему *вероятность правильной идентификации* [2].

Под вероятностью ошибки идентификации (эквивалентной ошибки) $P_{\text{ош}}$ понимается такое значение вероятности, при котором значения вероятностей «ложного срабатывания» и «пропуска цели» совпадают.

Базовая схема идентификации (см. рис. 1) состоит из блоков: параметризации, обучения, вычисления расстояния, принятия решения, они реализуют методы параметризации и распознавания образов. К используемым на практике методам параметризации РС

при текстонезависимой идентификации относятся методы, основанные на спектральном описании, представлении с использованием параметров передаточной функции голосового тракта и кепстральных параметров [3].

Проведенные исследования показали, что задачи цифровой обработки речевого сигнала такие как шумоподавление, низкоскоростное кодирование, распознавание, идентификация гораздо эффективнее и правильнее решать, учитывая особенности слуха человека и общее психоакустическое восприятие РС человеком [4].

Общепринятые границы воспринимаемого слухом частотного диапазона составляют (20–20000 Гц). При восприятии РС слуховой аппарат человека делит его на области частот, которые называются критическими полосами [5]. Это такая полоса частот, которая возбуждает одну и ту же часть базилярной мембраны, являющейся одной из основных составных частей слухового аппарата человека. Ширина критической полосы остается примерно постоянной вплоть до значения частоты 500 Гц, а при увеличении значений частоты увеличивается.

Для определения расстояния между центральными частотами соседних критических полос часто используют специальную единицу частоты – барк. Связь между частотами, выраженными в Гц и барк, используется для преобразования линейной шкалы частот в шкалу, соответствующую спектральному восприятию человека (1).

$$z(f) = 13arctg(0.76f) + 3.5arctg((f/7.5)^2)\,[\text{барк}]. \qquad (1)$$

Более простое для применения и использования выражение можно представить (3)

$$B(\text{z}) = 1125\ln(1 + f/700), \qquad (2)$$

где f и z – частоты, выраженные в кГц и барк соответственно.

Явление разделения при психоакустическом восприятии человека спектра РС на частотные группы относится к одному из фундаментальных свойств слуха и широко используется в системах идентификации диктора. Вследствие разделения спектра РС на частотные полосы слух реагирует не на общую мощность сигнала, а на мощность, сосредоточенную в отдельных частотных полосах. При этом более интенсивные частотные составляющие в полосах могут маскировать менее интенсивные. На этом принципе построены системы низкоскоростного кодирования, учитывающие психоакустические особенности восприятия РС (рис. 2).

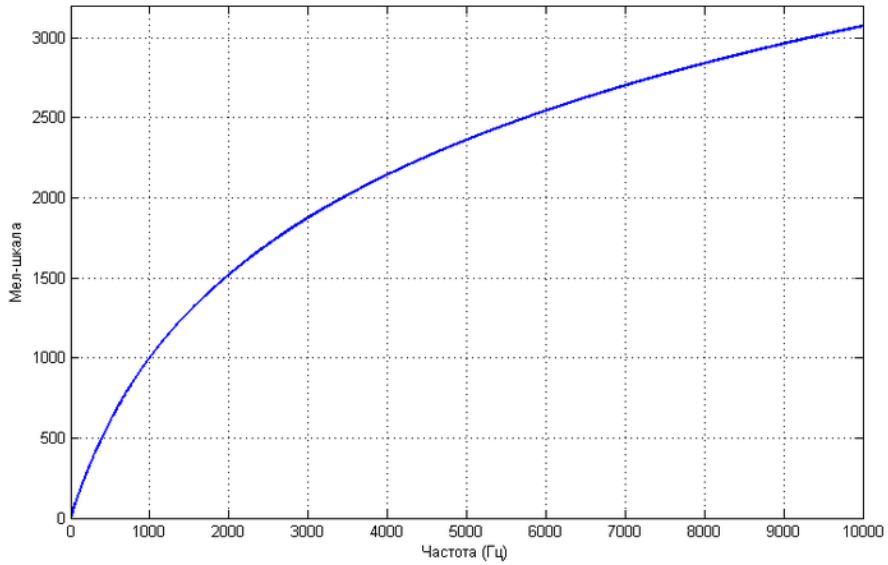


Рис. 2. Преобразование шкалы частот

Громкость звуков определяется как субъективное ощущение уровня речи. Для численной оценки громкости принято сравнивать уровень речи с чистым тоном частотой 1000 Гц. Значение звукового давления эталонного сигнала, равногромкого данному звуку, называется уровнем громкости этого звука. Вдоль каждой кривой уровня громкости, измеряемый в фонах (под одним фоном понимается уровень громкости звука, для которого уровень звукового давления равногромкого с ним звука частоты 1000 Гц равен 1 дБ), остается постоянным и полагается равным уровню звукового давления в дБ на частоте 1 кГц. Графически этот факт можно представить в виде кривых равной громкости, которые показаны на рис. 3.

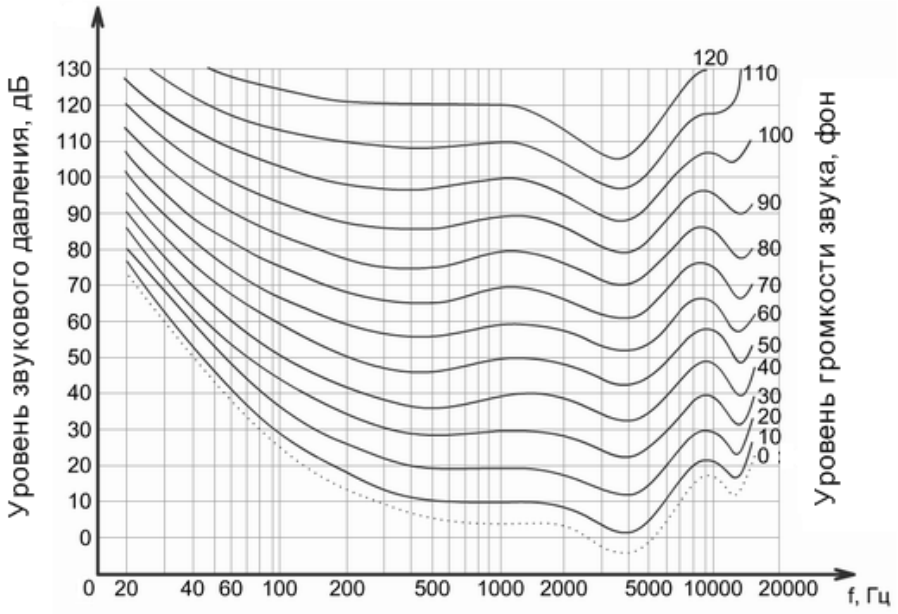


Рис. 3. Кривые равной громкости

Для представления интенсивности звука в определенной области частот используются единицы [мел]. Мел – единица высоты звука, основанная на восприятии этого звука нашими органами слуха. Мел удобно применять в целях анализа речи человека, так как его использование «приближает» алгоритмы обработки данных к человеческим параметрам восприятия, что благотворно сказывается на качестве распознавания.

Шкалы мел и барков являются эмпирически подобранными значениями, отражающими особенности слуха человека. На их основе рассчитываются мел-частотные кепстральные коэффициенты (MFCC), которые в последнее время получили весьма широкое использование при ЦОРС. Последовательность их расчета, следующая:

1. С использованием преобразования Фурье рассчитывается спектр речевого сигнала.
2. Полученный спектр пересчитывается для мел-шкалы.
3. Формируются треугольные перекрывающиеся окна, равномерно расположенные на мел-шкале.
4. Значения векторов спектра сигнала умножаются на оконную функцию после чего находится энергия сигнала, которая попадает в каждое из окон анализа.
5. Полученные коэффициенты (значения энергии) возводятся в квадрат и логарифмируются.
6. Используется еще раз преобразование Фурье, но, как правило, в связи с переходом к расчету значений квадрата комплексного спектра, применяется дискретное косинусное преобразование (ДКП). Особенностью ДКП является тот факт, что оно позволяет эффективно работать с четными функциями, при этом спектр остается вещественным, что очень удобно для его дальнейшего использования.

За счет нелинейной шкалы соотношения частот (см. рис. 2) наложение равномерных ОФ в области мел-шкалы приводит к высокому «оконному» разрешению (малой ширине окна) в области низких частот и низкому «оконному» разрешению (большой ширине окна) в области высоких частот для линейной частотной шкалы. По сути осуществляется приведение спектра РС к значениям воспринимаемым слуховым аппаратом человека, в частности базилярной мембраной, при этом осуществляется сокращение признакового пространства параметров, описывающих РС на сегменте анализа.

В представленном техническом решении для формирования эмпирической плотности распределения характеристик РС контрольной фразы использовалось моделирование в среде MATLAB. Использовалась функция [mfcc], которая предназначена для расчета мел-кепстральных коэффициентов [coeffs], а также их разностных значений [delta] и разностей от первоначально вычисленных разностей [deltaDelta], как аналогов первой и второй производной:

$$delta = \frac{\sum_{k=-M}^{M} k \cdot coeffs(k,:)}{\sum_{k=-M}^{M} k^2}; \quad deltaDelta = \frac{\sum_{k=-M}^{M} k \cdot delta(k,:)}{\sum_{k=-M}^{M} k^2}. \qquad (3)$$

Формат ее вызова следующий:

[coeffs,delta,deltaDelta,loc]=mfcc(audioIn,fs);

Вызов функции может быть осуществлен следующим образом:

[coeffs] = mfcc(audioIn,fs,'LogEnergy','Replace', 'Ignore');

По сути функция [mfcc] делит РС на сегменты и вычисляет для каждого из них 13 кепстральных коэффициентов. Формируется матрица, содержащая значения кепстральных коэффициентов. Ее строки определяют количество целых сегментов, а столбцы номера соответствующих коэффициентов на сегменте анализа.

Формирование эталонной плотности для обучения системы реализуется на контрольной фразе не менее 5 секунд, в которой используются не менее 70 % фонем русской речи. Проведенные исследования показали, что использование MFCC позволяет снизить ошибку первого рода при использовании классических решений по идентификации диктора и является наиболее предпочтительным перед другими решениями по параметризации РС (коэффициенты линейного предсказания, линейные спектральные частоты, спектральное описание РС). Перспективным направлением на наш взгляд представляется использование MFCC, представленных в качестве параметров эмпирической плотности распределения для обучения системы идентификации диктора, основанной на использовании нейросетевых технологиях обработки.

Библиографический список

1. Аграновский, А. В. Теоретические аспекты алгоритмов обработки и классификации речевых сигналов / А. В. Аграновский, А. В. Леднов. – Москва : Радио и связь, 2004. – 164 с.

2. Репалов, С. А. Разработка математических моделей и робастных алгоритмов идентификации дикторов по их речи : дис. … канд. техн. наук / Репалов С. А. – Москва : РГБ, 2003. – 140 с.

3. Максимов, А. В. Развитие систем автоматической текстонезависимой идентификации дикторов / А. В. Максимов, Н. М. Чавчавадзе, С. Ю. Мельников, М. В. Федюкин. – Москва : СТЭЛ – КС, 2009. – URL: http://www.stel.ru/news/pdf/melnikov%202009.pdf

4. Голунов, В. И. Верификация и идентификация говорящего / В. И. Голунов. – Санкт-Петербург, 2002. – URL: www.auditech.ru/article/ver_obz.doc

5. Рихтер, С. Г. Кодирование и передача речи в цифровых системах подвижной радиосвязи : учеб. пособие для вузов / С. Г. Рихтер. – Москва : Горячая линия : Телеком, 2010. – 304 .

Для цитирования:

Афанасьев, А. А. Идентификация диктора при использовании параметров эмпирической плотности распределения характеристик речевого сигнала контрольной фразы / А. А. Афанасьев // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 133–140.

УДК 004.056

С. П. Хворостухин

ПОСЛЕДОВАТЕЛЬНО-ПАРАЛЛЕЛЬНАЯ ОРГАНИЗАЦИЯ УЗЛА ШИФРОВАНИЯ НА БАЗЕ ПЛИС

Аннотация. Рассматриваются подходы к организации параллельных вычислений криптографических преобразований при реализации их на базе ПЛИС. Предлагаются подходы к последовательно-параллельной организации узла шифрования, приводятся результаты оценки производительности таких узлов. Предлагается вариант построения с динамическим перераспределением вычислительных ресурсов между функциями шифрования и расшифрования в зависимости от условий функционирования.

S. P. Khvorostukhin

SERIAL-PARALLEL ORGANIZATION OF THE ENCRYPTION BASED ON FPGA

Abstract. The prospect of parallel computing of cryptographic transformations when implementing them on FPGA. Approaches to the serial-parallel organization of the encryption are proposed, and the results of evaluating the performance are presented. A variant of construction with dynamic redistribution of computing resources between encryption and decryption functions is proposed, depending on the operating conditions.

С целью повышения производительности узлов шифрования могут быть использованы решения, предполагающие распараллеливание вычислений [1–3].

Однако, при реализации параллельной обработки существует необходимость сохранения исходной последовательности поступивших пакетов.

Предлагается организовать узел шифрования в виде набора ядер криптографических преобразований, которые последовательно загружаются информацией для преобразования. Загрузка осуществляется циклически, по готовности очередного ядра принять пакет информации. Распределение пакетов реализуется арбитром записи. Функций криптографического преобразования при этом выполняются каждым ядром независимо от остальных ядер. Выда-

ча преобразованной информации осуществляется пакетами, последовательно, в цикле аналогичном циклу загрузки. Последовательность выдачи задается арбитром чтения.

Для обеспечения хранения пакетов записываемой и выдаваемой информации целесообразно использовать буфер. Размер буфера должен быть достаточен для хранения пакета максимального размера.

Каждое ядро включает функции криптографического алгоритма, независящие от типа выполняемой операции и режима использования криптографического алгоритма, и набор функций, являющихся специфичными для типа операции (шифрование или расшифрование) и режима. Таким образом, каждое ядро можно представить в виде ядра криптоалгоритма и интерфейса функции. Схематичное представление узла шифрования/расшифрования представлено на рис. 1.

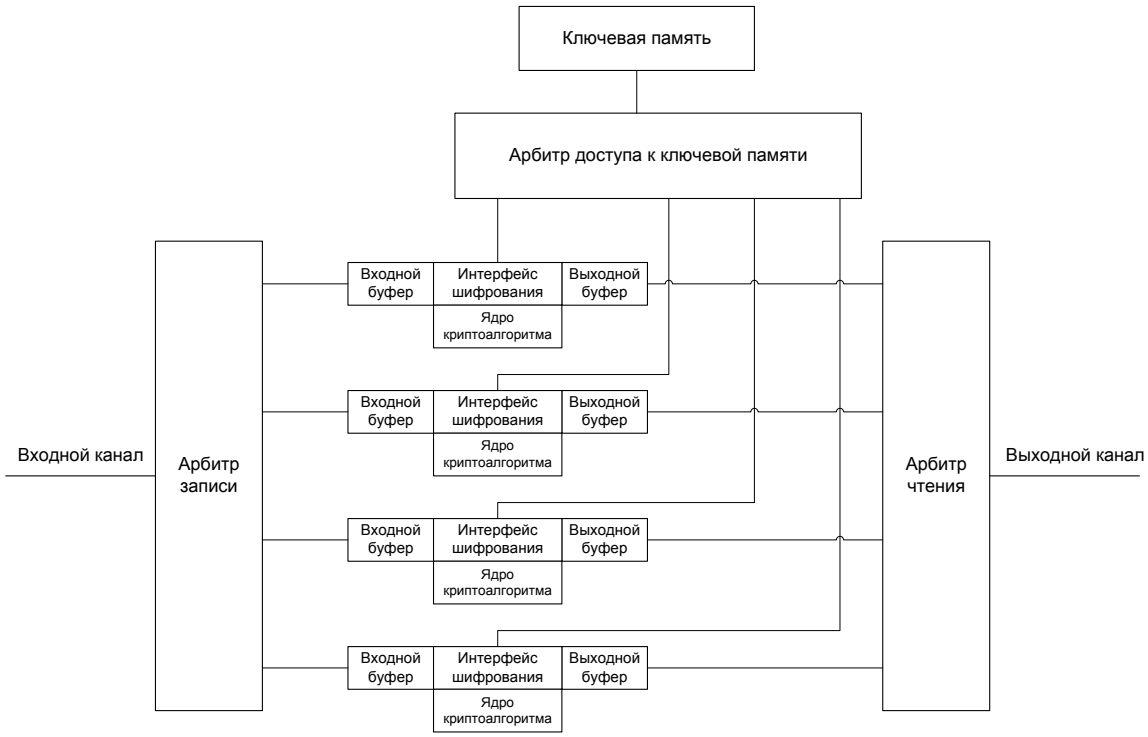


Рис. 1. Схематичное представление распараллеливания для функции шифрования

Был выполнен синтез узлов шифрования ГОСТ 34.12–2015 [4] на базе ПЛИС Spartan6 xc6slx100t. Варианты отличались количеством ядер криптоалгоритма, определенных для каждой функции. Для полученных вариантов была выполнена оценка производительности с использованием контрольных пакетов фиксированного размера, близкого к максимальному. Были получены результаты

зависимости производительности и зависимости затрачиваемых ресурсов ПЛИС от количества ядер криптоалгоритма на каждую функцию, представленные в табл. 1.

Таблица 1

Зависимость производительности и затрачиваемых ресурсов узла шифрования от количества ядер криптоалгоритма

Количество ядер криптоалгоритма, шт.	1	2	4	8	16
Максимальная скорость шифрования, Мбит/с	45	90	180	340	620
Объем затраченных ресурсов ПЛИС, %	58	63	70	83	96

При увеличении количества ядер повышение производительности ограничивается разделением общего ресурса. В предложенной схеме таким ресурсом является память ключей. При отсутствии параметров, требующих синхронизации между ядрами, каждому ядру может быть выделена собственная память, что позволяет убрать разделяемый ресурс. Однако, при необходимости поддержания синхронности некоторых параметров между ядрами, например, счѐтчики расхода ключа, предпочтительнее использовать общую память. Доступ к памяти должен выделяться последовательно. По завершении операций с ключом ядро должно освобождать доступ к памяти для следующего ядра.

При нестабильности межпакетного интервала входного потока информации, обеспечить равномерность нагрузки на ядра криптоалгоритма позволяет увеличение числа входных и выходных буферов каждого ядра. Группа буферов организуется по кольцевому принципу. Увеличение количества буферов позволяет хранить большее число пакетов в памяти и повышает устойчивость узла относительно переполнения.

Для некоторого класса задач, которые не требуют поддержки одинаковой производительности функций шифрования и расшифрования, например, режим предварительного шифрования, предлагается использовать вариант построения узла шифрования с динамическим распределением производительности между функциями шифрования и расшифрования. Такой вариант построения предполагает выделения всех ядер, реализующих криптографический алгоритм, в отдельный массив. Интерфейсы шифрования с входными и выходными буферами образуют отдельный массив. Такой же массив образуется для интерфейсов расшифрования. Доступ

интерфейсов к ядрам криптографического алгоритма реализуется через арбитра доступа. Схематичное представление предлагаемого узла представлено на рис. 2.

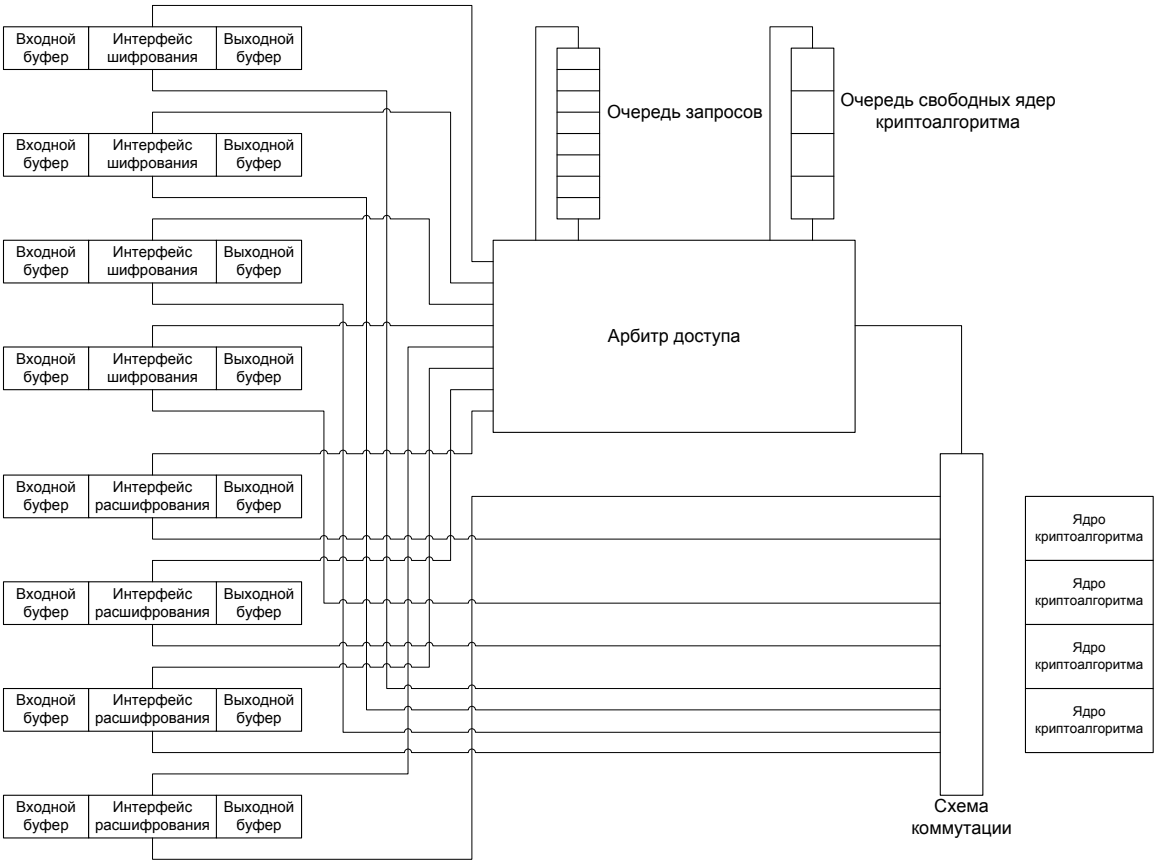


Рис. 2. Схематичное представление узла шифрования с динамическим распределением ресурсов между функциями шифрования и расшифрования

При получении пакета соответствующий интерфейс запрашивает у арбитра доступ к ядру криптоалгоритма. При наличии свободного ядра, арбитр выделяет доступ запросившему интерфейсу через схему коммутации. По завершении операции с ядром, интерфейс освобождает ядро и оно возвращается в очередь доступных для использования.

Такой вариант построения позволяет распределять имеющиеся вычислительные возможности между функциями шифрования и расшифрования в соответствии с условиями функционирования, позволяя, при необходимости, передать все ядра криптоалгоритма на выполнение только одной функции. При условии, что количество ресурсов ПЛИС, требуемых для реализации ядра криптоалгоритма, значительно больше ресурсов, требуемых на реализацию

интерфейса, повышается так же плотность заполнения ПЛИС и увеличивается количество доступных ядер криптоалгоритма для каждой функции.

Библиографический список

1. Родионов, А. Ю. Архитектура криптографического сопроцессора на ПЛИС / А. Ю. Родионов // Вопросы защиты информации. – 2016. – № 3. – С. 16–19.

2. Шелудько, В. М. Ускорение шифрования данных по блочным алгоритмам с помощью GPU и технологии CUDA / В. М. Шелудько // Техническая кибернетика, радиоэлектроника и системы управления : сб. материалов XI Всерос. науч. конф. молодых ученых, студентов и аспирантов. – Таганрог : Изд-во ЮФУ, 2012. – Т. 2. – С. 99.

3. Кожевников, А. А. Разработка и исследование высокоскоростных шифраторов на основе ПЛИС Spartan-6 / А. А. Кожевников // Техническая кибернетика, радиоэлектроника и системы управления : сб. материалов XI Всерос. науч. конф. молодых ученых, студентов и аспирантов. – Таганрог : Изд-во ЮФУ, 2012. – Т. 2. – С. 103.

4. ГОСТ 34.12–2015. Информационная технология. Криптографическая защита информации. Блочные шифры / Федеральное агентство по техническому регулированию и метрологии. – Москва : Стандартинформ, 2016.

Для цитирования:

Хворостухин, С. П. Последовательно-параллельная организация узла шифрования на базе ПЛИС / С. П. Хворостухин // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 141–145.

УДК 34.096; 519.2

Е. А. Малыгина

КОМПАКТНОСТЬ И СТЕРИЛЬНОСТЬ НЕЙРОСЕТЕВОГО ЭМБРИОНА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ДЛЯ ДОВЕРЕННОЙ НИЗКОРЕСУРСНОЙ ВЫЧИСЛИТЕЛЬНОЙ СРЕДЫ ВЫСОКОНАДЕЖНОЙ БИОМЕТРИКО-НЕЙРОСЕТЕВОЙ АУТЕНТИФИКАЦИИ ЛИЧНОСТИ

Аннотация. Рассматриваются особенности эмбриона человека и его естественного интеллекта как образец для разработки нейросетевого эмбриона искусственного интеллекта. Копируются отдельные свойства эмбриона и формулируются его функции, позволяющие безопасно развиваться в низкоресурсной доверенной вычислительной среде.

E. A. Malygina

COMPACTNESS AND STERILITY OF THE ARTIFICIAL INTELLIGENCE NEURAL NETWORK EMBRYO FOR A TRUSTED LOW-RESOURCE COMPUTING ENVIRONMENT OF HIGHLY RELIABLE BIOMETRIC-NEURAL NETWORK AUTHENTICATION OF THE INDIVIDUAL

Abstract. The features of the human embryo and its natural intelligence are considered as a model for the development of the neural network embryo of artificial intelligence. Individual properties of the embryo are copied and its functions are formulated, allowing it to develop safely in a low-resource trusted computing environment.

Широкое применение систем биометрической идентификации/аутентификации личности и систем разграничения доступа ставит задачу высоконадежного распознавания биометрических данных личности.

При этом аутентификация считается высоконадежной при выполнении следующих условий: приемлемой вероятности ошибок первого рода и гарантированно малой вероятностью ошибок второго рода, сопоставимой по своему значению с вероятностью случайного подбора кода неизвестного криптографического ключа при малом числе попыток подбора [1].

При обучении нейронной сети преобразованию тайного биометрического образа в ключ доступа полуавтоматический режим недопустим (не допускается участие оператора). Режим обучения может быть только автоматическим, появление сторонних наблюдателей – это угроза компрометации тайного биометрического образа человека и кода его ключа доступа [1].

Частичная или полная компрометация тайны используемого биометрического образа эквивалентна частичной или полной компрометации связанного с ним ключа доступа. Поэтому использование доверенной вычислительной среды для обучения и тестирования нейросетевого искусственного интеллекта распознавания биометрических данных личности является обязательным условием.

Построение различного рода систем распознавания образов с использованием нейронных сетей на всех этапах развития последних прямо или косвенно связано с изучением и робкой попыткой копирования отдельных моментов работы человеческого мозга. Известно, что мозг состоит из нейронов. У каждого нейрона есть один длинный отросток, аксон, и много коротких отростков, дендритов, Связи между ними называются синапсами и имеют достаточно сложную структуру, включающую порядка 10^{11} нейронов, у каждого из которых в среднем 7000 связей [2].

Каждый нейрон время периодически посылает по аксону нервный импульс, имеющий электрическую природу. При этом нейрон может находиться в двух разных состояниях: выключенном состоянии – частота подачи сигналов маленькая и, наоборот, в возбужденном состоянии – частота подачи импульсов значительно увеличивается. Отметим при этом, что нейрон выдает электрические сигналы через случайные промежутки времени. Интенсивность этого процесса меняется в зависимости от того, возбужден нейрон или нет.

При этом в работе [2] отмечено, что нейроны головного мозга могут быстро синхронизироваться и очень точно засекать весьма короткие промежутки времени. Частота импульсов в аксонах может составлять от 10 Гц в неактивном состоянии до приблизительно 200 Гц во время самой сильной активации, т.е. связи между нейронами активируются за десятки миллисекунд, и в полном цикле, например, при распознавании образов, находящегося перед ним, не может быть последовательной цепочки активаций длиннее десятка нейронов.

Если сравнивать процессы, происходящие в головном мозгу человека с процессами обычной вычислительной машины, в которой процессор исполняет длинные последовательные цепочки команд, обрабатывая их в синхронном режиме, то у мозга цепочки

короткие, зато работает он асинхронно и с очень высокой степенью параллелизации: нейроны активируются сразу во многих местах мозга, когда начинают распознавать необходимый объект [2].

При этом необходимо учитывать, что в головном мозге есть огромное число горизонтальных связей между нейронами одного уровня, множество замкнутых цепочек связанных нейронов, а также неожиданных и пока не совсем понятных связей нейронов из совершенно разных областей мозга, но целостной картины как все это работает в настоящее время еще не сложилось [2].

Анализируя данные, приведенные в работе [3], можно дополнить к вышесказанному отметить, что процессы, происходящие в нейронах нервной системы разнообразны, активность и количество нейронов во многом зависит от решаемой задачи и скорости ее решения, но при этом видно, что на выходе как отдельного нейрона или отдельных участков нейронной сети нет однозначного принятого решения «Да/Нет», а присутствует некий массив решений «Да/Нет», позволяющий более точно описать исследуемый образ.

Данная парадигма частично реализована при разработке систем высоконадежной биометрико-нейросетевой аутентификации по ГОСТ Р [1], когда на входе преобразователя биометрия-код используется 416 преобразованных параметров рукописного слова-пароля, а на выходе получается код ключа равный 256 бит (256 комбинаций «Да/Нет» (правая часть рис. 1), в отличии от предложенной исследователями США и Канады версии «нечеткого экстрактора» (левая часть рис. 1) [4–6].

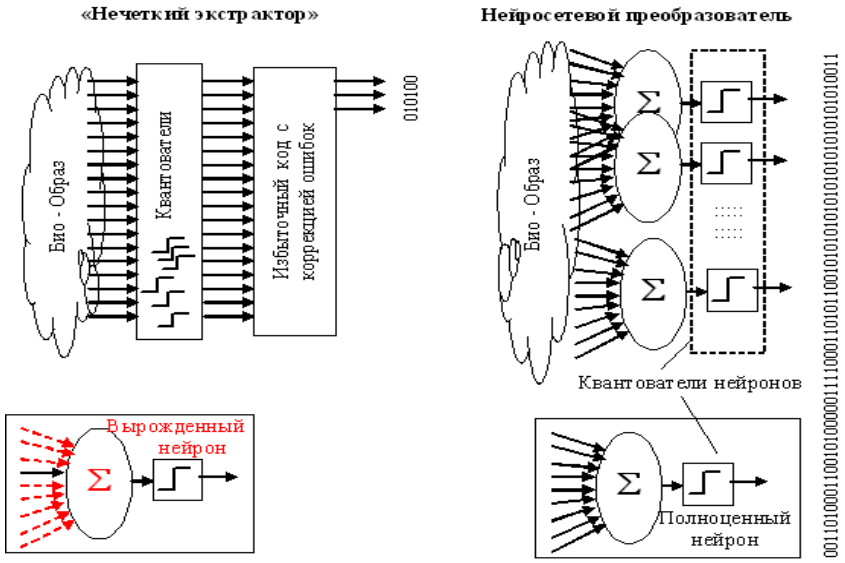


Рис. 1. Обобщенная схема обработки биометрических данных в «нечетком экстракторе» и нейросетевом преобразователе биометрия-код

Однако остается нерешенным еще один важный вопрос безопасности обработки и хранения данных. Обратимся к процессам, происходящим в эмбрионе будущего ребенка, находящегося в женском организме. Известно, что защита эмбриона от инфекции обеспечивается прежде всего факторами неспецифической резистентности и специфического иммунитета со стороны материнского организма [7]. Во-первых, механические барьеры: слизистая оболочка шейки матки, сомкнутый внутренний зев матки, устья маточных труб, слизистая пробка в шеечном канале; во-вторых, биохимические реакции в организме, направленные на подавление инфекционного агента и выведение его из организма. К ним можно отнести повышение температуры тела, увеличение содержания в крови кортикостероидов, увеличение синтеза лизоцима, разрушающего многие виды бактерий. В-третьих, важным звеном в элиминации возбудителя инфекции является фагоцитарная система (моноциты, макрофаги, гранулоциты). Эти клетки осуществляют свою функцию без участия иммунных механизмов, но могут инициировать и специфический иммунный ответ [7].

При вирусной инфекции активизируются киллерные клетки, которые лизируют бактериальные и вирусные инфекты. Специфические иммунные механизмы еще более сложные, так как вовлекают в защитный процесс не только систему иммунитета, но и нейроэндокринную и гемокоагуляционную системы [7].

Ученые открыли яркий случай иронии эволюции: фрагменты генетического материала вирусов, закрепившиеся в ДНК человека, со временем превратились в ключевые элементы иммунной системы, которая как раз и борется с опасными микроорганизмами, в том числе с вирусами [8].

Оказывается, что реликты древних вирусных инфекций неожиданным образом приносят пользу здоровью и сохраняются в геноме по принципам естественного отбора. Так, в 2015 г. Джоанна Высоцка (Joanna Wysocka) и ее коллеги из Стэндфордского университета выяснили, что эндогенные ретровирусы (ЭРВ) обеспечивают выживание людей: ЭРВ защищает трехдневные эмбрионы от других вирусов и регулирует генную активность клеток.

В составляющих плод восьми клетках ученые нашли не только ДНК родителей, но и генный материал HERVK – самого последнего ЭРВ, попавшего к человеку примерно 200 тысяч лет назад. HERVK вырабатывает белок, не позволяющий другим вирусам попасть в эмбрион, то есть вирус защищает человека от гриппа

и других опасных болезней. Более того, Rec (акцессорный белок HERVK) связывает некоторые клеточные РНК и регулирует деятельность рибосом. Таким образом, ЭРВ играет важную роль в раннем развитии человека, участвуя в генной активности эмбриональных клеток.

Человеческий зародыш-эмбрион был бы совсем беззащитен перед вирусной инфекцией, если бы его не защищали собственные вирусы [8].

Представим упрощенный аналог эмбриона искусственного нейросетевого интеллекта для систем биометрико-нейросетевой аутентификации (рис. 2).

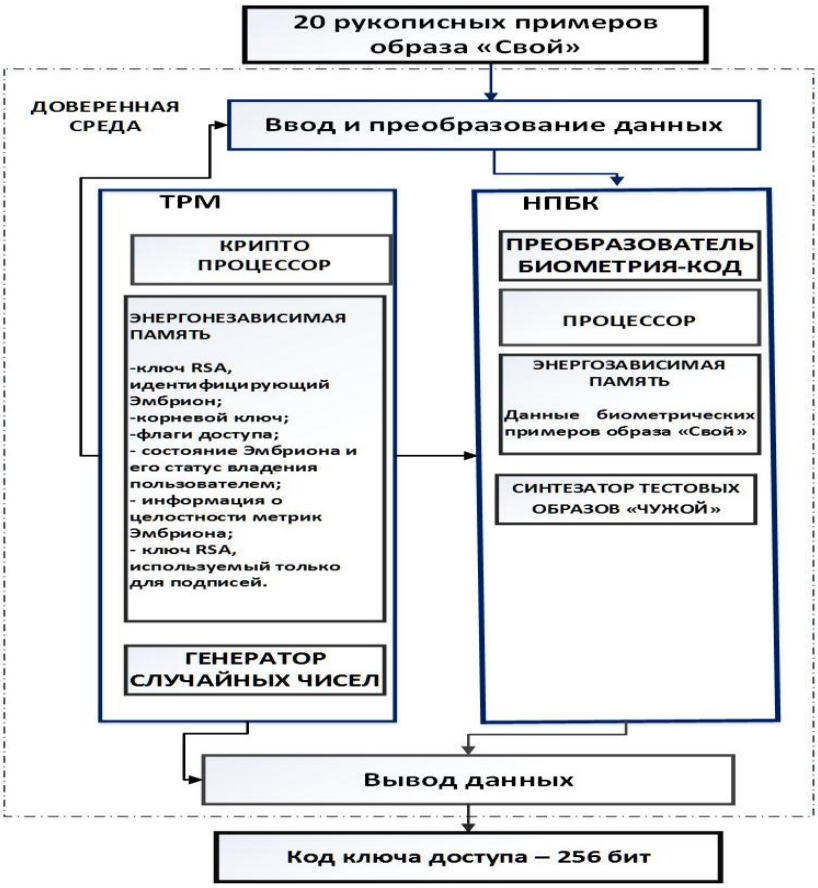


Рис. 2. Эмбрион искусственного нейросетевого интеллекта для высоконадежных систем биометрической аутентификации доверенной вычислительной среды

Он представляет собой симбиоз Trusted Platform Module (TPM) и нейросетевого преобразователя биометрия-код с модулем-синтезатором тестовых образов «Чужой».

Модуль TPM содержит в себе криптопроцессор, обеспечивающий средства безопасного создания ключей шифрования. Модуль

имеет следующие возможности: удалѐнную аттестацию, привязку и надѐжное защищѐнное хранение. Удаленный аудит создает связь аппаратных средств, загрузки системы и конфигурации компьютера, делая проверку достоверности и целостности программного обеспечения. Криптопроцессор шифрует данные таким способом, что они могут быть расшифрованы только на компьютере, где были зашифрованы, под управлением того же самого программного обеспечения. Привязка шифрует данные, используя ключ подтверждения TPM – уникальный ключ RSA, записанный в чип в процессе его производства. В архитектуре чипа реализованы следующие защитные механизмы: защищѐнное управление памятью; шифрование шины и данных; тестирование режимов блокирования; активное экранирование [9].

Нейросетевой преобразователь биометрия-код (НПБК) представляет в нашем случае обученную на 20 примерах рукописного образа «Свой» искусственную нейронную сеть с 416 входами и 256 выходами, преобразующий частично случайный вектор входных биометрических параметров «Свой» в однозначный код криптографического ключа (длинного пароля) и преобразующая любой иной случайный вектор входных данных в случайный выходной код.

Для проверки качества обучения системы высоконадежной биометрико-нейросетевой аутентификации необходимо проводить процедуру тестирования [10]. Для этого необходимо сформировать базы биометрических образов, размеры которых должны гарантировать подтверждение заданных характеристик.

Наиболее сложной частью задачи является тестирование стойкости обученных преобразователей биометрия-код к атакам подбора. Как правило, создать достаточно большой объем естественных биометрических образов по требованиям ГОСТ Р 52633.1–2009 технически невозможно [11]. По этой причине приходится многократно увеличивать размеры тестовой базы биометрических образов за счет синтетических биометрических образов, созданных по ГОСТ Р 52633.2–2010 [12].

Еще одной проблемой является то, что биометрические образы «Чужой» не являются изначально дискретными как пароли или ключи. Невозможно абсолютно точно указать дискретное кодовое состояние образа «Чужой».

В связи с этим возникают дополнительные проблемы по взаимному упорядочиванию образов «Чужой» в тестовых базах и га-

рантиям равномерности заполнения пространства возможных кодовых состояний.

При этом необходимо учитывать и тот факт, что базы «Чужой» должны быть сформированы с привлечением пользователей тестируемых систем в реальных условиях их эксплуатации.

Указанные выше проблемы не только усложняют процесс тестирования стойкости преобразователя биометрия-код к атакам подбора, но и требуют значительных временных, людских и финансовых затрат.

С целью устранения этих недостатков предлагается проводить тестирование систем высоконадежной биометрико-нейросетевой аутентификации личности с использованием полноценной обучающей базы образов «Свой», состоящей из 20 примеров. На ее основе циклически синтезируется единичные образы «Чужой», полученные методами мутаций и морфинга образов «Свой», созданные из нескольких примеров естественного образа «Свой».

После процесса тестирования и получения данных меры Хемминга с последующей записью их в таблицу, сгенерированный образ уничтожается и на вход подается следующий образ «Чужой». Затем цикл повторяется снова, эмпирически оцененное минимальное количество циклов – 128, при котором уже можно делать приемлемый прогноз возникновения вероятности ошибки появления кодового отклика «Свой» при предъявлении образа «Чужой».

Предложенный способ тестирования позволяет разработать малогабаритный защищенный модуль, использующий доверенную среду вычисления, в которой находятся: низко разрядный вычислитель, например, 8 битный, малый объем памяти, системный монитор, обеспечивающий функционирование эмбриона, нейросетевой преобразователь биометрия-код, примеры образа «Свой».

Использованные после обучения и тестирования примеры образа «Свой» удаляются из памяти эмбриона, в которой остаются только параметры обученного нейросетевого преобразователя биометрия-код, что обеспечивает безопасность самого образа «Свой» и экономит память эмбриона.

Таким образом, наличие механизмов синтезирования искусственных тестовых биометрических образов «Чужой» из естественных примеров образа «Свой», позволяет добиться компактности нейросетевого эмбриона искусственного интеллекта. При этом не компрометируется образ «Свой». Тестирование проводится самим пользователем в условиях реальной эксплуатации системы аутентификации.

Наличие криптографических механизмов контроля целостности программного обеспечения модуля TPM, защита и аудит операций обмена с внешней вычислительной средой, позволяет обеспечить стерильность эмбриона в доверенной вычислительной среде.

Заключение

Алгоритмов автоматического обучения больших искусственных нейронных сетей недостаточно для безопасного развития нейросетевого искусственного интеллекта. Обязательным условием является доверенная вычислительная среда, компактность и стерильность нейросетевого эмбриона искусственного интеллекта, криптографический контроль целостности исполняемых программ, постоянный аудит их функционирования.

Библиографический список

1. ГОСТ Р 52633.0–2006. Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации. – Москва : Стандартинформ, 2007. – 25 с.

2. Николенко, С. Глубокое обучение / С. Николенко, А. Кадурин, Е. Архангельская. – Санкт-Петербург, 2018. – 480 с. – (Сер.: Библиотека программиста).

3. Николлс, Дж. От нейрона к мозгу / Дж. Николлс, Р. Мартин, Б. Валлас, П. Фукс ; пер. с англ. П. М. Балабана, А. В. Галкина, Р. А. Гиниатуллина, Р. Н. Хазипова, Л. С. Хируга. – Москва : Едиториал УРСС, 2003. – 672 с.

4. Dodis, Y. Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy Data / Y. Dodis, L. Reyzin, A. Smith. – 2004. – April 13. – URL: www.cs.bu.edu/~reyzin/fuzzy.html

5. Иванов, А. И. Нечеткие экстракторы: проблема использования в биометрии и криптографии / А. И. Иванов // Первая миля. – 2015. – № 1. – С. 54–57.

6. Вятчанин, С. Е. Анализ эффективности систем аутентификации с нечеткими экстракторами и преобразователями биометрия-код / С. Е. Вятчанин, А. Ю. Малыгин, А. В. Сериков // Проблемы автоматизации и управления в технических системах : сб. докл. XXXII Междунар. науч.-техн. конф. (г. Пенза, 6–8 июня 2017 г.). – Пенза, 2017. – Т. 1. – С. 104–105.

7. Эмбрион человека находится под вирусной защитой. – URL: https://texnomaniya.ru/embrion-cheloveka-nahoditsya-pod-virusnoiy-zashitoiy

8. Круто ты попал в ДНК. – URL: https://texnomaniya.ru/kruto-ti-popal-v-dnk

9. Trusted Platform Module // Википедия. – URL: https://ru.wikipedia.org/wiki/Trusted_Platform_Module

10. ГОСТ Р 52633.3–2011. Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора. – Москва : Стандартинформ, 2018. – 12 с.

11. ГОСТ Р 52633.1–2009. Защита информации. Техника защиты информации. Требования к формированию баз естественных биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации. – Москва : Стандартинформ, 2010. – 24 с.

12. ГОСТ Р 52633.2–2010. Защита информации. Техника защиты информации. Требования к формированию синтетических биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации. – Москва : Стандартинформ, 2011. – 17 с.

Для цитирования:

Малыгина, Е. А. Компактность и стерильность нейросетевого эмбриона искусственного интеллекта для доверенной низкоресурсной вычислительной среды высоконадежной биометрико-нейросетевой аутентификации личности / Е. А. Малыгина // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 146–154.

УДК 519.2

Б. Б. Ахметов, В. А. Цимбал, С. А. Полковникова

МОДЕЛИРОВАНИЕ РАСПРЕДЕЛЕНИЯ ХИ-КВАДРАТ ДЛЯ СУЩЕСТВЕННО ЗАВИСИМЫХ БИОМЕТРИЧЕСКИХ ДАННЫХ

Аннотация. Рассматривается вопрос моделирования случайных многомерных процессов, происходящих в матрице 256×256, описывающих корреляционные связи между 256 разрядами выходного кода преобразователя биометрия-код с использованием симметричной связывающей матрицы. Симметризация корреляционных связей построена на совпадении выходной энтропии кодов.

B. B. Akhmetov, V. A. Tsymbal, S. A. Polkovnikova

MODELING THE DISTRIBUTION OF HI-SQUARE FOR ESSENTIALLY DEPENDENT BIOMETRIC DATA

Abstract. The question of modeling random multidimensional processes occurring in matrix 256x256, describing correlations between 256 discharges of output code of the converter biometrics-code using a symmetrical binding matrix.

Известно, что на малых выборках примеров образа «Свой» ошибка вычисления корреляционной матрицы получается достаточно большой [1]. Например, для обучающей выборки, состоящей из 20 примеров, ошибка вычисления коэффициентов корреляции может составлять $\Delta r = \pm 0.5$. При обращении корреляционных матриц эта ошибка усиливается пропорционально числу обусловленности $cond\left[R_N\right]$:

$$\Delta e \approx \Delta r \cdot cond\left[R_N\right]. \quad (1)$$

В свою очередь число обусловленности растет с ростом размерности обращаемой матрицы и может быть очень большим. Решением данной задачи можно осуществить при помощи нейросетевых эмуляторов квадратичных форм [1].

Нейроны нейросетевого эмулятора осуществляют следующее квадратичное преобразование.

$$y = \sum_{i=1}^{m} \frac{(E(v_i) - v_i)^2}{(\sigma(v_i))^2}, \qquad (2)$$

где m – число входов у нейрона или число степеней свободы; $\sigma(v_i)$ – стандартное отклонение i-го биометрического параметра, обрабатываемого нейроном.

Затем осуществляется квантование квадратичных данных, путем их сравнения с порогом k:

$$\begin{cases} z(y) = "0" \quad if \quad y \le k, \\ z(y) = "1" \quad if \quad y \rangle k \end{cases}. \qquad (3)$$

При этом вероятности возникновения ошибок первого и второго рода для решений, принимаемых отдельным нейроном, зависят от заданного значения порога k и распределения значений $p(y)$.

Данная задача имеет аналитическое решение для независимых данных. В этом случае выходные данные квадратичных нейронов описываются хи-квадрат распределением Пирсона:

$$p(y) = \frac{1}{2^{\frac{m}{2}} \cdot \Gamma\left(\frac{m}{2}\right)} \left\{ y^{\frac{m}{2}-1} \cdot \exp\left(\frac{-y}{2}\right) \right\}, \qquad (4)$$

где Γ(.) – гамма функция.

Нейроны, реализующие вычисления (2) и (3), можно называть хи-квадрат нейронами.

Однако простое аналитическое описание (4) не работает для сильно коррелированных биометрических данных. В связи с этим возникает необходимость моделирования влияния корреляционных связей биометрических данных на достоверности решений, принимаемых одним нейроном.

Из теории известно [2], что моделировать случайные многомерные процессы крайне сложно. Технически возможно вычислить симметричную матрицу корреляционных связей 256×256, описывающую корреляционные связи между 256 разрядами выходного кода, однако построить генератор, точно воспроизводящий корреляционные связи столь высокой размерности, невозможно.

Формально можно использовать 256 генераторов независимых случайных данных, умножив их некоторую связывающую данные матрицу A. Однако найти нужную связывающую матрицу

A, которая даст нужные корреляционные связи $r("x_k","x_j")$, трудно. Эта обратная задача относится к плохо обусловленным.

Так как задача не решается в прямом виде, предлагается использовать симметричную связывающую матрицу, которая имеет единичную диагональ и одинаковые элементы вне диагонали:

$$\begin{bmatrix} 1 & a & \cdots & a \\ a & 1 & \cdots & a \\ \cdots & \cdots & \cdots & \cdots \\ a & a & \cdots & 1 \end{bmatrix} \times \begin{bmatrix} \xi_{1,i} \\ \xi_{2,i} \\ \cdots \\ \xi_{m,i} \end{bmatrix} = \begin{bmatrix} y_{1,i} \\ y_{2,i} \\ \cdots \\ y_{m,i} \end{bmatrix} \Rightarrow R_m = \begin{bmatrix} 1 & r & \cdots & r \\ r & 1 & \cdots & r \\ \cdots & \cdots & \cdots & \cdots \\ r & r & \cdots & 1 \end{bmatrix}. \quad (5)$$

Только в этом случае, данные оказываются равно коррелированными. Если плавно изменять регулируемый параметр связывающей матрицы от 0 до 1, равная коррелированность также меняется от 0 до 1. Номограмма связи единственного регулируемого параметра матрицы с коэффициентом равной коррелированности выходных данных дана на рис. 1.

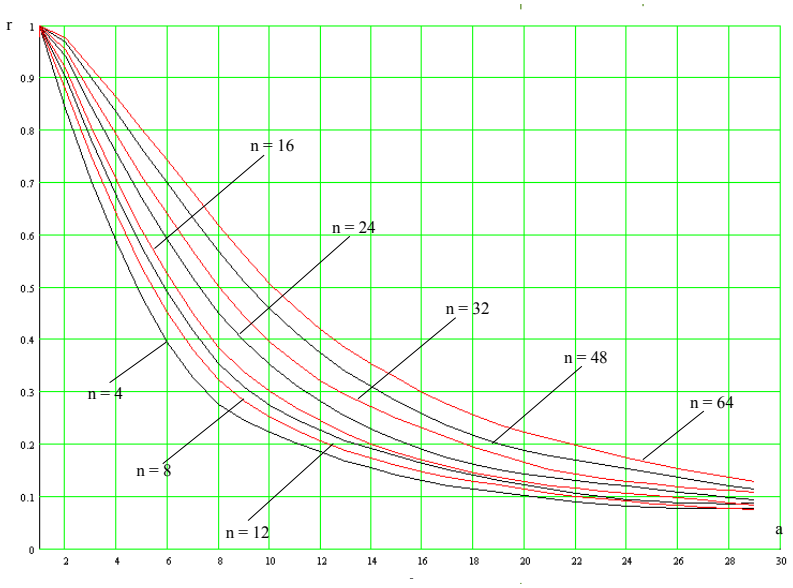


Рис. 1. Номограмма связи коэффициентов связывающей матрицы – a с коэффициентами равной коррелированности данных

Умножение непрерывных данных (континуумов) на связывающую матрицу порождает вектор непрерывных откликов $\bar{y}$. Для того, чтобы непрерывные данные преобразовать в дискретные данные, необходимо использовать 256 компараторов. Блок-схема моделирования симметричных равно коррелированных кодов приведена на рис. 2.

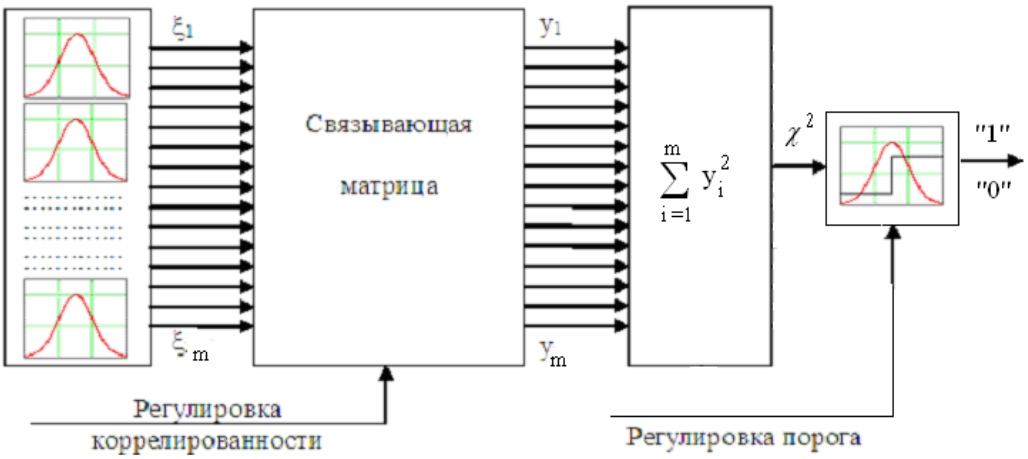


Рис. 2. Блок-схема моделирования симметричных равно коррелированных данных при получении зависимых хи-квадрат распределений с числом степеней свободы m

Отметим, что симметризация многомерных задач является типовым приемом, используемым при идентификации нелинейных динамических объектов, если они описываются рядом Вольтерра [4–6]. Этот прием имеет строгое математическое доказательство и позволяет сделать k-мерную задачу идентификации параметров ядра Вольтерра k-го порядка одномерной.

Примерно такого же эффекта удается добиться и симметризацией задачи, описываемой в данной статье. При этом единственное отличие симметризации ядер Вольтерра и симметризации корреляционных связей состоит в критерии эквивалентных замещений. Симметризация ядер Вольтерра построена на совпадении откликов нелинейного динамического объекта. Симметризация корреляционных связей строится на совпадении выходной энтропии кодов [7].

Библиографический список

1. Ахметов, Б. Б. Многомерный статистический анализ биометрических данных сетью частных критериев Пирсона / Б. Б. Ахметов, А. И. Иванов, А. В. Безяев, Ю. В. Фунтикова // Вестник Национальной академии наук Республики Казахстан. – 2015. – № 1. – С. 5–11.

2. Шалыгин, А. С. Прикладные методы статистического моделирования / А. С. Шалыгин, Ю. И. Палагин. – Ленинград : Машиностроение, 1986. – 320 с.

3. Эйкхофф, П. Основы идентификации систем управления / П. Эйкхофф. – Москва : Мир, 1975. – 517 с.

4. Ivanov, A. I. Simple Numerical Method of Separabel Volterra Kernels Symmetrization / A. I. Ivanov // Engineering Simulation. – 1999. – Vol. 16. – P. 411–416.

5. Иванов, А. И. Синтез нелинейных динамических моделей Винера – Гаммерштейна перераспределением памяти между входом и выходом / А. И. Иванов // Автоматика и телемеханика. – 1997. – № 11. – С. 21–32.

6. Иванов, А. И. Одномерный аналог многомерной идентификации Ли-Щецина / А. И. Иванов // Управляющие системы и машины. – 1999. – № 2. – С. 16–21.

7. Ахметов, Б. Б. Многомерные статистики существенно зависимых биометрических данных, порождаемые нейросетевыми эмуляторами квадратичных форм : монография / Б. Б. Ахметов, А. И. Иванов. – Алматы : Изд-во LEM, 2016. – 61 с.

Для цитирования:

Ахметов, Б. Б. Моделирование распределения хи-квадрат для существенно зависимых биометрических данных / Б. Б. Ахметов, В. А. Цимбал, С. А. Полковникова // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 155–159.

УДК 519.2; 519.6

Н. Н. Вершинин, В. А. Цимбал, Б. С. Ахметов, И. В. Урнев

ПОСТРОЕНИЕ ТАБЛИЦ ДОВЕРИТЕЛЬНОЙ ВЕРОЯТНОСТИ РЕШЕНИЙ, ПРИНИМАЕМЫХ ОТДЕЛЬНЫМИ НЕЙРОНАМИ КВАДРАТИЧНЫХ ЭМУЛЯТОРОВ

Аннотация. Рассматриваются вопросы получения данных для аналитического описания распределения хи-квадрат для существенно коррелированных биометрических данных при помощи численного моделирования. На основе данных построены таблицы для разного уровня коррелированности данных при разном числе степеней свободы.

N. N. Vershinin, V. A. Tsymbal, B. S. Akhmetov, I. V. Urnev

BUILDING TABLES OF CONFIDENCE PROBABILITY DECISIONS MADE BY INDIVIDUAL NEURONS OF SQUARE EMULATORS

Abstract. The issue of obtaining data for an analytical description of the distribution of hi-square for substantially correlated biometric data using numerical modeling is considered. Based on the data, tables are built for different levels of correlated data with different degrees of freedom.

К сожалению, простое аналитическое описание критерия хи-квадрат не работает для биометрических данных, по причине их сильной коррелированности [1–4]. В связи с этим необходимо смоделировать влияние корреляционных связей биометрических данных и получить таблицы достоверности принимаемых одним нейроном решений. На рис. 1 приведены полученные результаты.

Из данных рис. 1 видно, что при увеличении уровня коррелированности данных происходит смещение моды распределения в левую сторону. Это приводит к значительным изменениям доверительных вероятностей [2, 4].

Значения соответствующих доверительных вероятностей приведены в табл. 1. Первая строка табл. 1 для значения коррелированности 0.01 очень хорошо совпадает с аналитическим описанием для независимых данных.

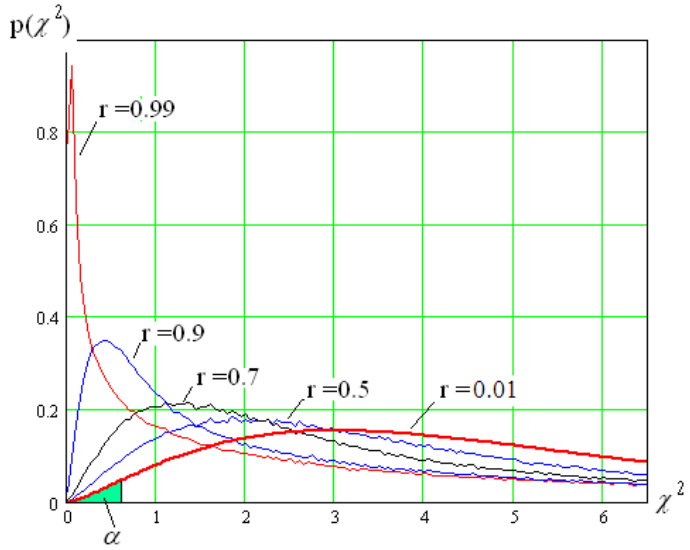


Рис. 1. Распределения хи-квадрат для разного уровня коррелированности данных при m = 5 степеней свободы

В частности, для доверительной вероятности P = 0.5 значение хи-квадрат составляет 4.35. Для некоррелированных данных эта величина составляет 5 (совпадает с числом степеней свободы m).

Из центрального столбца (P = 0.5) табл. 1 видно, что с увеличением уровня коррелированности значение хи-квадрат снижается с 4.35 до 2.29. Это эквивалентно снижению наблюдаемого числа степеней свободы по мере увеличения коррелированности данных.

Таблица 1

Значения хи-квадрат для разных уровней достоверности и разных значениях коррелированности данных при числе степеней свободы m = 5

m = 5		Квантили доверительной вероятности принятия верного решения – α										
		0.01	0.02	0.05	0.1	0.2	0.5	0.8	0.9	0.95	0.98	0.99
Коррелированность данных – r	0.01	0.56	0.76	1.15	1.61	2.35	4.35	7.29	9.29	11.10	13.40	15.10
	0.1	0.55	0.74	1.13	1.59	2.32	4.32	7.29	9.27	11.18	13.62	15.43
	0.2	0.52	0.71	1.09	1.53	2.24	4.23	7.28	9.43	11.55	14.35	16.49
	0.3	0.49	0.67	1.02	1.45	2.13	4.09	7.27	9.63	12.08	15.43	18.15
	0.4	0.45	0.62	0.94	1.34	1.98	3.90	7.26	9.97	12.86	16.89	20.08
	0.5	0.41	0.55	0.85	1.21	1.81	3.67	7.28	10.39	13.76	18.45	22.07
	0.6	0.35	0.48	0.74	1.06	1.60	3.40	7.37	10.94	14.8	20.12	24.32
	0.7	0.29	0.39	0.61	0.88	1.36	3.11	7.54	11.56	15.89	21.82	26.55
	0.8	0.22	0.29	0.47	0.68	1.08	2.80	7.73	12.19	16.97	23.5	28.63
	0.9	0.13	0.18	0.29	0.43	0.74	2.51	7.98	12.87	18.15	25.42	31.13
	0.99	0.02	0.03	0.06	0.12	0.36	2.29	8.18	13.47	19.12	26.94	32.99

Эта тенденция только усиливается с ростом числа входов (фиктивного числа степеней свободы) у нейронов Пирсона. Графики соответствующих кривых, полученные для m = 26, приведены на рис. 2.

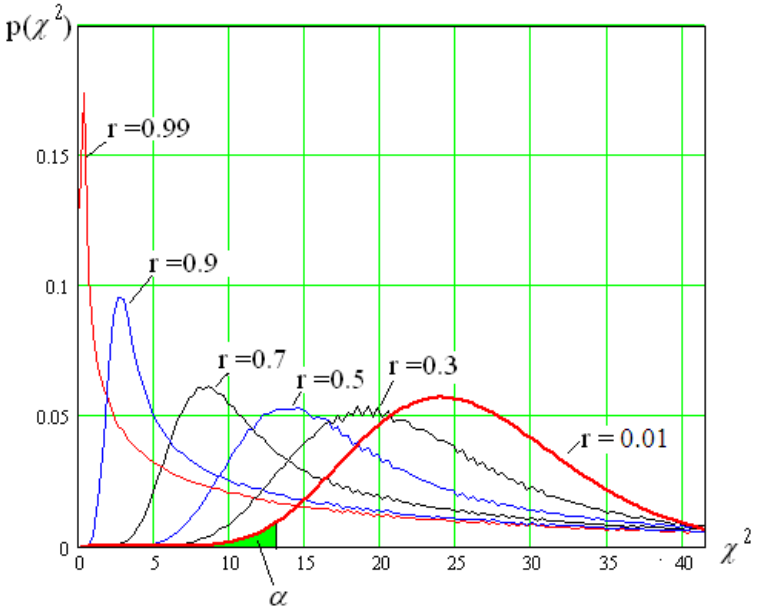


Рис. 2. Распределения хи-квадрат для разного уровня коррелированности данных при m = 26 степеней свободы

Сравнивая данные, приведенные на рис. 1 и 2, можно сделать вывод о нормализации распределений значений [3]. С ростом размерности решаемой задачи происходит приближение критерия хи-квадрат к нормальному распределению значений. Рост коррелированности данных приводит к обратному эффекту деформации нормального закона распределения.

Доверительные вероятности, полученные из кривых распределений рис. 3 приведены в табл. 2.

Таблица 2

Значения критерия хи-квадрат для разных уровней достоверности и разных значениях коррелированности данных при числе степеней свободы m = 26

m = 26		Квантили доверительной вероятности принятия решения – α										
		0.01	0.02	0.05	0.1	0.2	0.5	0.8	0.9	0.95	0.98	0.99
Коррелированность данных – r	0.01	12.20	13.42	15.37	17.29	19.83	25.33	31.80	35.56	38.87	42.82	45.61
	0.1	11.74	12.91	14.86	16.76	19.29	24.99	32.05	36.45	40.62	45.99	50.09
	0.2	10.85	11.98	13.84	15.67	18.17	24.06	32.32	38.42	44.92	54.01	61.14
	0.3	9.79	10.82	12.54	14.25	16.66	22.71	32.90	41.69	51.17	64.49	74.80
	0.4	8.60	9.53	11.09	12.68	14.96	21.14	33.8	45.31	57.75	74.91	88.45
	0.5	7.39	8.20	9.57	11.01	13.11	19.48	35.06	49.32	64.64	85.67	102.2
	0.6	6.05	6.75	7.92	9.17	11.08	17.75	36.48	53.47	71.64	96.7	116.3
	0.7	4.69	5.24	6.19	7.23	8.92	16.11	37.91	57.67	78.57	107.9	130.6
	0.8	3.27	3.67	4.38	5.19	6.64	14.66	39.58	62.01	85.68	118.8	144.5
	0.9	1.73	1.96	2.39	2.93	4.12	13.24	41.13	66.08	92.58	129.7	158.3
	0.99	0.21	0.25	0.36	0.65	1.90	12.01	42.54	69.84	99.19	139.8	171.0

На рис. 3 даны функции плотностей распределения критерия хи-квадрат с 32 степенями свободы для разных значений коррелированности данных.

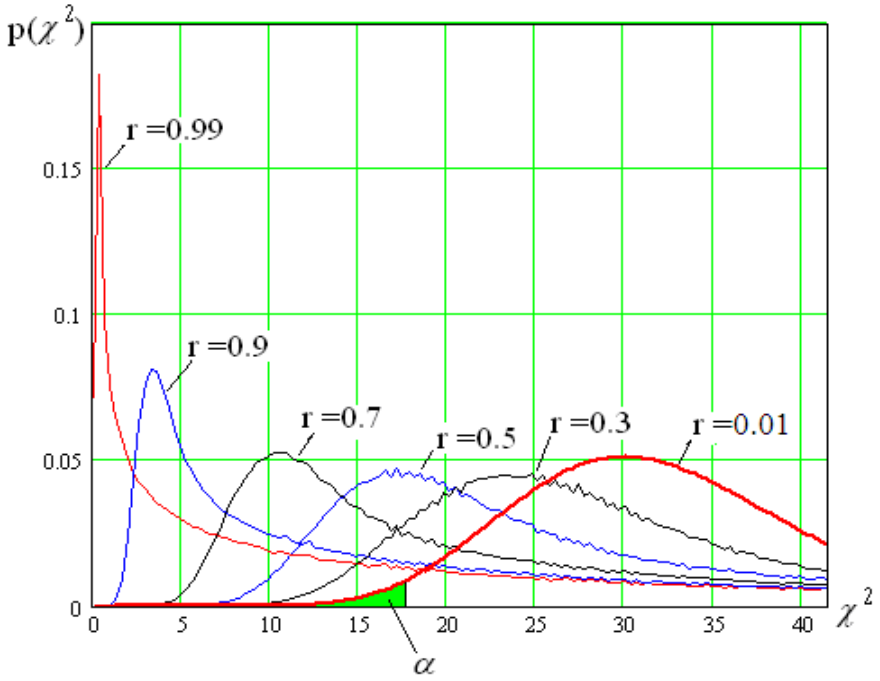


Рис. 3. Распределения хи-квадрат для разного уровня коррелированности данных при m = 32 степеней свободы

Сравнивая данные, приведенные на рис. 1, 2 и 3, следует обратить внимание на рост деформации исходного распределения хи-квадрат распределения с ростом размерности решаемых задач. Так, для высокого уровня коррелированности данных r = 0.99 и исходных данных r = 0.00 повышение размерности распределения приводят к монотонному росту расстояний между их модами (5-2-0.05) = 2.99, (26-2-0.5) = 23.5, (32-2-0.7) = 29.3.

Значения доверительных вероятностей решений одиночных нейронов Пирсона при m = 32 приведены в табл. 3.

Очевидно, что для каждого показателя степени свободы m может быть построена своя система кривых, описывающих плотности хи-квадрат распределений и для них построены иные таблицы квантилей уровня достоверности проверяемых гипотез. Эти данные могут быть использованы для получения промежуточных таблиц, соответствующих данным с промежуточными числами степеней свободы, которые можно получить обычной интерполяцией.

Таблица 3

Значения хи-квадрат для разных уровней достоверности и разных значениях коррелированности данных при числе степеней свободы m = 32

m = 32		Квантили доверительной вероятности принятия решения –α										
		0.01	0.02	0.05	0.1	0.2	0.5	0.8	0.9	0.95	0.98	0.99
Коррелированность данных -r	0.01	16.36	17.78	20.08	22.28	25.15	31.32	38.46	42.58	46.17	50.46	53.47
	0.1	15.68	17.07	19.33	21.53	24.43	30.85	38.76	43.76	48.48	54.75	59.58
	0.2	14.45	15.78	17.95	20.07	22.92	29.65	39.25	46.55	54.37	65.41	73.9
	0.3	13	14.21	16.2	18.18	20.97	27.93	40.07	50.75	62.37	78.59	91.15
	0.4	11.41	12.48	14.29	16.13	18.76	25.94	41.33	55.45	70.64	91.55	108
	0.5	9.74	10.69	12.31	13.96	16.39	23.9	42.95	60.46	79.19	105	125
	0.6	7.974	8.776	10.14	11.6	13.81	21.79	44.81	65.65	87.9	118.6	142.8
	0.7	6.156	6.799	7.902	9.11	11.06	19.81	46.61	70.83	96.5	132.4	160.6
	0.8	4.282	4.736	5.56	6.499	8.183	18.05	48.69	76.18	105.4	145.9	177.5
	0.9	2.256	2.52	3.015	3.643	5.073	16.29	50.61	81.28	113.9	159.6	194.7
	0.99	0.270	0.315	0.438	0.808	2.336	14.78	52.36	85.95	122.1	172.1	210.4

Библиографический список

1. Технология использования больших нейронных сетей для преобразования нечетких биометрических данных в код ключа доступа : монография / Б. С. Ахметов, А. И. Иванов, В. А. Фунтиков, А. В. Безяев, Е. А. Малыгина. – Алматы : Изд-во LEM, 2014. – 144 с.

2. Ахметов, Б. Б. Многомерный статистический анализ биометрических данных сетью частных критериев Пирсона / Б. Б. Ахметов, А. И. Иванов, А. В. Безяев, Ю. В. Фунтикова // Вестник Национальной академии наук Республики Казахстан. – 2015. – № 1. – С. 5–11.

3. Шалыгин, А. С. Прикладные методы статистического моделирования / А. С. Шалыгин, Ю. И. Палагин. – Ленинград : Машиностроение, 1986. – 320 с.

4. Ахметов, Б. Б. Многомерные статистики существенно зависимых биометрических данных, порождаемые нейросетевыми эмуляторами квадратичных форм : монография / Б. Б. Ахметов, А. И. Иванов. – Алматы : Изд-во LEM, 2016. – 61 с.

Для цитирования:

Вершинин, Н. Н. Построение таблиц доверительной вероятности решений, принимаемых отдельными нейронами квадратичных эмуляторов / Н. Н. Вершинин, В. А. Цимбал, Б. С. Ахметов, И. В. Урнев // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 160–164.

УДК 519.2

А. В. Потапов

ПРОВЕРКА ЭКВИВАЛЕНТНОСТИ ЭНТРОПИИ КОДОВ РЕАЛЬНОЙ КОРРЕЛЯЦИОННОЙ МАТРИЦЫ И ЕЕ СИММЕТРИЧНОГО АНАЛОГА

Аннотация. Показано, что энтропия кодов для реальной корреляционной матрицы и ее аналога с равными коэффициентами корреляции расходится. При этом с ростом размеров используемых данных ошибка уменьшается.

A. V. Potapov

CHECKING THE EQUIPI OF ENTROPY CODES OF THE REAL CORRELATION MATRIX AND ITS SYMMETRICAL ANALOGUE

Abstract. It is shown that the entropy of codes for the real correlation matrix and its analogue with equal correlation ratios diverges. However, as the size of the data used increases, the error decreases.

Процедура симметризации корреляционных связей не имеет строгого аналитического доказательства, поэтому необходимо оценить возможную погрешность от замещения реальных корреляционных матриц на их симметричный аналог [1].

Если рассматривать двухмерную задачу, то погрешность оказывается нулевой [2, 3]. Это происходит из-за симметричности двухмерных корреляционных матриц:

$$R_2 = \begin{bmatrix} 1 & r_{1,2} \\ r_{2,1} & 1 \end{bmatrix}. \qquad (1)$$

Все корреляционные матрицы симметричны по определению, то есть:

$$r_{1,2} = r_{2,1} = \frac{|r_{1,2}| + |r_{2,1}|}{2} = \tilde{r}\,. \qquad (2)$$

Более сложной оказывается ситуация для корреляционных матриц третьего порядка:

$$R_3 = \begin{bmatrix} 1 & r_{1,2} & r_{1,3} \\ r_{2,1} & 1 & r_{2,3} \\ r_{3,1} & r_{3,2} & 1 \end{bmatrix}. \qquad (3)$$

Для корреляционных матриц общего вида коэффициенты корреляции, находящиеся на диагонали, не совпадают [4]. Учитывая симметрию коэффициентов корреляции относительно диагонали можно представить, как:

$$\tilde{r} = \frac{|r_{1,2}| + |r_{1,3}| + |r_{2,3}|}{3}. \quad (4)$$

Затем необходимо проверить, насколько энтропия кодов с реальной корреляционной матрицей будет отличаться от энтропии кодов с ее симметричным аналогом. Проверить это можно с помощью имитационного моделирования, выполняемого по блок-схеме рис. 1 [1].

Если реализовать численный эксперимент программно, используя порядка 10 000 случайных числа, полученных от трех программных генераторов нормальных данных, то получим поток трехбитных случайных кодов. Далее следует вычислить вероятности появления каждого из возможных кодовых состояний и оценить их совместную энтропию для реальной (не полностью симметричной) корреляционной матрицы:

$$H("x_1, x_2, x_3") = -\sum_{i=1}^{8} P_i("x_1, x_2, x_3") \cdot \log_2(P_i("x_1, x_2, x_3")). \quad (5)$$

Далее тот же самый численный эксперимент следует провести для равно коррелированных данных с полностью симметричной корреляционной матрицей третьего порядка:

$$\tilde{H}("x_1, x_2, x_3") = -\sum_{i=1}^{8} \tilde{P}_i("x_1, x_2, x_3") \cdot \log_2(\tilde{P}_i("x_1, x_2, x_3")). \quad (6)$$

Очевидно, что результаты вычислений энтропии для реальной корреляционной матрицы и ее аналога с равными коэффициентами корреляции должны расходиться. При этом расхождение будет является случайной величиной, амплитуда которой зависит от размеров массива использованных случайных данных [5]. Если предлагаемые преобразования эквивалентны, то с ростом размеров используемых данных ошибка должна уменьшаться. Результаты численных экспериментов приведены в табл. 1.

Из данных табл. 1 видно, что при симметризации корреляционных связей трехмерных данных относительные ошибки лежат в интервале от 0.7 % до 1.8 %. Наблюдается также падение относительной ошибки с ростом размеров выборки кодов, на которых производились вычисления. Если увеличивать размерность задачи, то относительная ошибка также снижается.

Таблица 1

Значения относительной ошибки вычисления энтропии, наблюдаемой при замене реальных корреляционных связей на симметричные

№ опыта	$r_{1,2}$	$r_{1,3}$	$r_{2,3}$	$\tilde{r}$	Δ% при N = 20 000	Δ% при N = 40 000	Δ% при N = 80 000
1	-0.461	0.244	0.260	0.321	0.012	0.009	0.007
2	-0.461	0.244	0.260	0.321	0.014	0.011	0.008
3	0.178	0.161	-0.505	0.281	0.018	0.013	0.012
4	0.178	0.161	-0.505	0.281	0.017	0.013	0.010
5	0.418	-0.134	0.405	0.319	0.015	0.014	0.011
6	0.418	-0.134	0.405	0.319	0.017	0.013	0.012

Полученные результаты являются подтверждением возможности упрощения задачи оценки энтропии путем замещения реальной асимметричной матрицы корреляционных связей на ее симметричный аналог, при этом, замечено, что чем больше размерность задачи, тем меньше на ее энтропию влияет каждый отдельный биометрический параметр.

Библиографический список

1. Ахметов, Б. Б. Многомерные статистики существенно зависимых биометрических данных, порождаемые нейросетевыми эмуляторами квадратичных форм : монография / Б. Б. Ахметов, А. И. Иванов. – Алматы : Изд-во LEM, 2016. – 61с.

2. Шалыгин, А. С. Прикладные методы статистического моделирования / А. С. Шалыгин, Ю. И. Палагин. – Ленинград : Машиностроение, 1986. – 320 с.

3. Эйкхофф, П. Основы идентификации систем управления / П. Эйкхофф. – Москва : Мир, 1975. – 517 с.

4. Ахметов, Б. Б. Многомерный статистический анализ биометрических данных сетью частных критериев Пирсона / Б. Б. Ахметов, А. И. Иванов, А. В. Безяев, Ю. В. Фунтикова // Вестник Национальной академии наук Республики Казахстан. – 2015. – № 1. – С. 5–11.

5. Иванов, А. И. Синтез нелинейных динамических моделей Винера – Гаммерштейна перераспределением памяти между входом и выходом / А. И. Иванов // Автоматика и телемеханика. – 1997. – № 11. – С. 21–32.

Для цитирования:

Потапов, А. В. Проверка эквивалентности энтропии кодов реальной корреляционной матрицы и ее симметричного аналога / А. В. Потапов // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 165–167.

УДК 519.2

С. В. Туреев

ИСПОЛЬЗОВАНИЕ НОРМИРОВАННОГО КРИТЕРИЯ ХИ-КВАДРАТ ДЛЯ АНАЛИТИЧЕСКОГО ОПИСАНИЯ СТАТИСТИК АСИММЕТРИЧНЫХ РАСПРЕДЕЛЕНИЙ БИОМЕТРИЧЕСКИХ ПАРАМЕТРОВ В ТЕСТОВЫХ БАЗАХ

Аннотация. Показано, что нормированное хи-квадрат распределение можно достаточно эффективно использовать при аппроксимации асимметричных распределений биометрических параметров тестовых баз.

S. V. Tureev

USE OF THE RATIONED HEE-SQUARE CRITERION TO ANALYTICALLY DESCRIBE THE STATISTICS OF ASYMMETRICAL DISTRIBUTIONS OF BIOMETRIC PARAMETERS IN TEST BASES

Abstract. It is shown that the normalized hi-square distribution can be used effectively enough in approximation of asymmetrical distributions of biometric parameters of test bases.

При формировании больших баз тестовых образов необходимо классифицировать вводимые образы по группам стабильности, уникальности и качества их параметров [1, 2]. Таким образом, для каждого образа из базы тестовых образов должны быть указаны средняя стабильность, средняя уникальность и среднее качество всех биометрических параметров данного образа.

Возможность достаточно эффективного использования нормированного распределения хи-квадрат для описания распределения функционалов биометрических образов проиллюстрируем для тестовых баз рукописных образов.

Для этого возьмѐм обучающие выборки 100 пользователей «Свой», в каждой выборке содержится 20 примеров рукописного слова длиной 5 букв, и 2000 образов «Чужой» по одному примеру рукописного слова такой же длины, как и образ «Свой».

Затем вычислим положение и степень разбросанности биометрических параметров всех используемых распределений. Полученные значения математических ожиданий и дисперсии используем

для расчѐта стабильности, уникальности и качества параметров тестовых образов.

При вычислении показателя стабильности *i*-го контролируемого биометрического параметра необходимо воспользоваться выражением (1):

$$c(v_i) = \frac{\sigma_{Чужой}(v_i)}{\sigma_{Свой}(v_i)}, \qquad (1)$$

где $\sigma_{Чужой}(v_i)$ – стандартное отклонение i-го биометрического параметра множества образов «Чужой»; $\sigma_{Свой}(v_i)$ – стандартное отклонение *i*-го биометрического параметра множества образов «Свой».

Чтобы распределить биометрические образы по классам средней стабильности воспроизведения их параметров необходимо учесть показатель средней стабильности всех параметров классифицируемого биометрического образа.

Средняя стабильность $E(c(v))$ рассчитывается как среднее арифметическое всех контролируемых параметров $c(v)$ биометрического образа.

Полученная гистограмма плотности распределения средней стабильности биометрических параметров тестовых образов представлена на рис. 1.

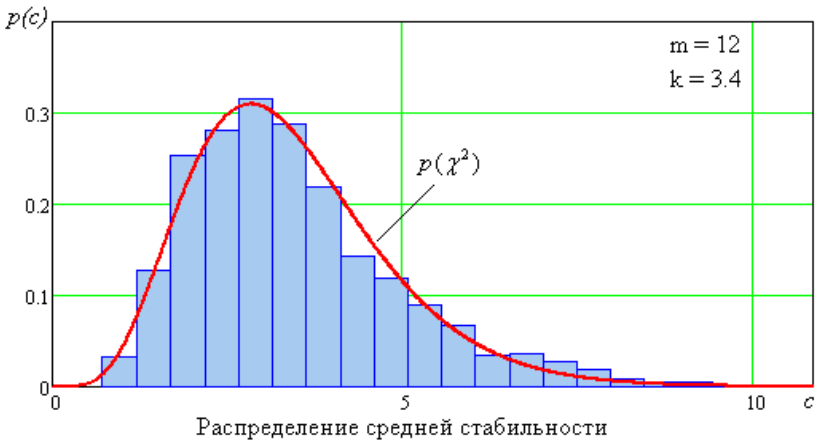


Рис. 1. Плотность нормированного хи-квадрат распределения и распределение средней стабильности параметров биометрических образов

Визуально видно, что полученное распределение имеет положительную асимметрию, что подтверждает невозможность использования нормального закона распределения значений для опи-

сания данного распределения. Распределения средней стабильности имеет следующие характеристики:

1) математическое ожидание – 3,452;
2) стандартное отклонение – 1,517;
3) коэффициент асимметрии – 1,045.

Полученные данные были использованы при нахождении параметров нормированного хи-квадрат распределения. Для нахождения оптимальной функции аппроксимации использовался метод наименьших квадратов. Подбиралось количество степеней свободы и корректирующий коэффициент (смещение математического ожидания) таким образом, чтобы основные статистические моменты двух распределений стали максимально близки, т.е. ошибка расхождения двух плотностей распределения была минимальной.

В итоге оптимальной найденной функцией является нормированное распределение хи-квадрат с числом степеней свободы равным 12 и корректирующим коэффициентом равным 3,4:

1) математическое ожидание – 3,401;
2) стандартное отклонение – 1,389;
3) коэффициент асимметрии – 0,82;
4) ошибка расхождения – 0,091.

Показатель уникальности *i*-го биометрического параметра, отражающий отличие контролируемого параметра от среднестатистического значения этого параметра, характерного для всех пользователей, вычисляется по формуле (2):

$$u(v_i) = \frac{\left| E_{Чужой}(v_i) - E_{Свой}(v_i) \right|}{\sigma_{Чужой}(v_i)}, \qquad (2)$$

где $E_{Чужой}(v_i)$ – математическое ожидание *i*-го биометрического параметра множества биометрических образов «Чужой»;

$E_{Свой}(v_i)$ – математическое ожидание *i*-го биометрического параметра множества биометрических образов «Свой».

Показатель средней уникальности всех параметров биометрического образа, вычисляется по формуле (3):

$$E(u(v_i)) = \frac{1}{n}\sum_{i=1}^{n} u(v_i). \qquad (3)$$

Построенная гистограмма плотности распределения средней уникальности и найденная плотность нормированного хи-квадрат распределения представлена на рис. 2.

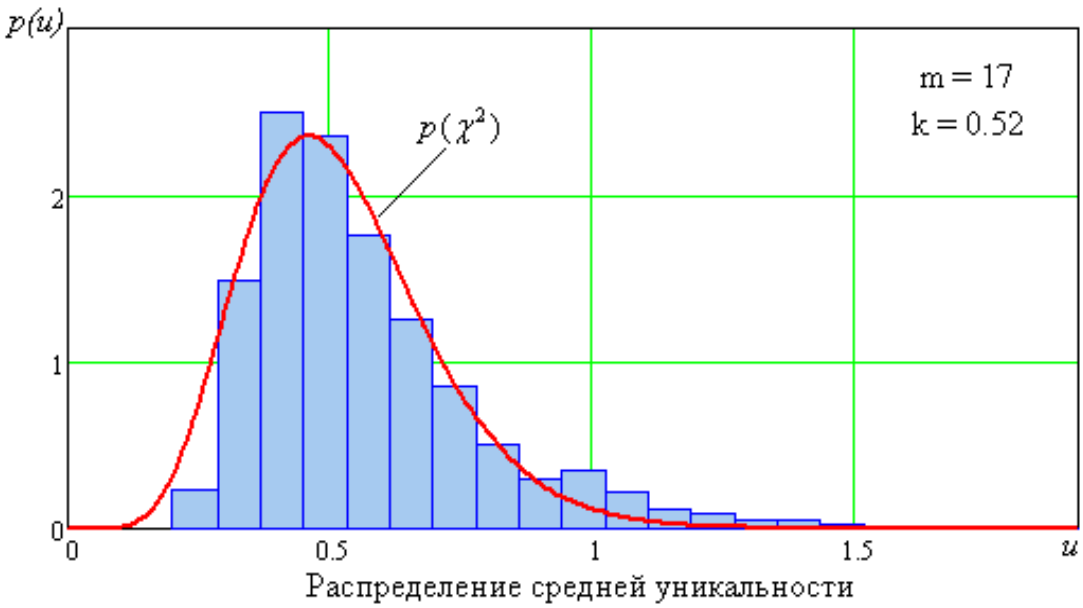


Рис. 2. Плотность нормированного хи-квадрат распределения и распределение средней уникальности параметров биометрических образов

Основные моменты экспериментально полученного распределения:

1) математическое ожидание – 0,568;
2) стандартное отклонение – 0,223;
3) коэффициент асимметрии – 1,551.

Оптимальной функцией, описывающей распределение средней уникальности функционалов тестовых образов, является нормированное хи-квадрат распределение с 17 степенями свободы и корректирующим коэффициентом равным 0,52:

1) математическое ожидание – 0,52;
2) стандартное отклонение – 0,179;
3) коэффициент асимметрии – 0,69;
4) ошибка расхождения – 0,157.

Затем был вычислен показатель качества *i*-го биометрического параметра (4):

$$q(v_i) = \frac{\left| E_{Чужой}(v_i) - E_{Свой}(v_i) \right|}{\sigma_{Чужой}(v_i) + \sigma_{Свой}(v_i)}. \qquad (4)$$

На рис. 3 представлены гистограмма плотности распределения среднего качества и описывающая данное распределение плотность нормированного распределения хи-квадрат.

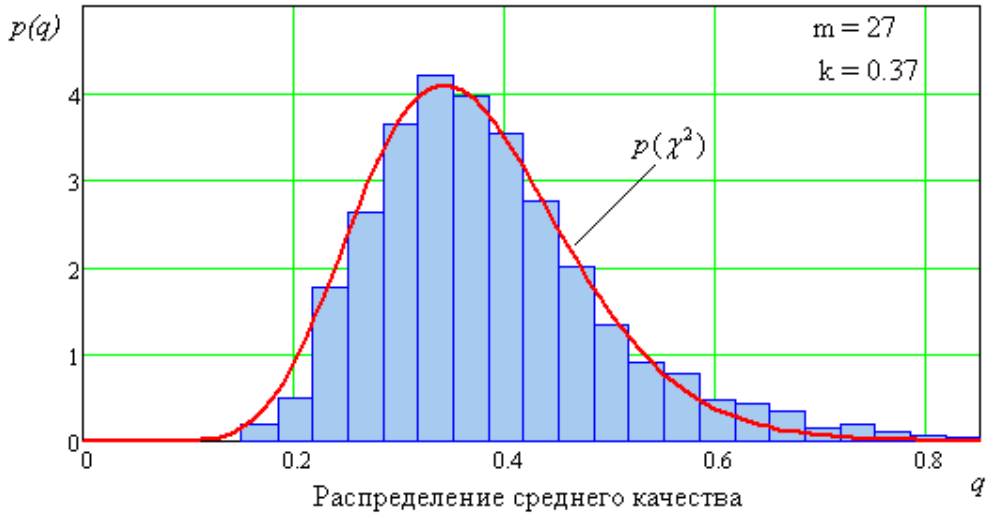


Рис. 3. Плотность нормированного хи-квадрат распределения и распределение среднего качества параметров биометрических образов

Для распределения среднего качества функционалов тестовых образов получили следующие моменты:

1) математическое ожидание – 0,385;
2) стандартное отклонение – 0,111;
3) коэффициент асимметрии – 0,95.

Минимальное расхождение плотностей распределения получается при аппроксимации функцией нормированного хи-квадрат с 27 степенями свободы и корректирующим коэффициентом равным 0,37:

1) математическое ожидание – 0,37;
2) стандартное отклонение – 0,101;
3) коэффициент асимметрии – 0,544;
4) ошибка расхождения – 0,096.

Нормированное хи-квадрат распределение может быть использовано при формировании баз естественных биометрических образов «Чужой».

Базы естественных биометрических образов «Чужой», предназначенные для тестирования средств биометрической аутентификации, должны отражать статистику распределения активного населения (страны, региона) по возрасту, половому признаку, роду занятия, квалификации и иным характеристикам, присущим людям, для которых создано тестируемое средство высоконадежной биометрической аутентификации [2, 3].

Формирование баз естественных биометрических образов «Чужой» должно осуществляться, исходя из десятикратного пре-

вышения, находящихся в базе случайных образов «Чужой», по отношению к ожидаемой стойкости к атакам подбора тестируемого средства аутентификации [1].

Прогноз стойкости к атакам подбора (обратной величины вероятности ошибки второго рода) должен осуществляться внутренними средствами тестирования средства аутентификации [1, 4]. Число биометрических образов полной базы может быть вычислено, используя следующее выражение (5)

$$N_{Полн} = \frac{10}{P_2}. \tag{5}$$

При отсутствии доверия к внутренним средствам тестирования (либо при отсутствии таких средств) прогноз вероятности ошибок второго рода – P_2 вычисляется приближенно, исходя из гипотезы нормального закона распределения значений вероятности ошибок, по формуле (6):

$$P_2 \approx \frac{1}{\sqrt{2\pi}} \int_{\sqrt{n}E(q(v))}^{\infty} \exp(\frac{x^2}{2}) \cdot dx, \tag{6}$$

где n – число учитываемых средством аутентификации биометрических параметров; $E(q(v))$ – среднее качество всех учитываемых средством биометрической аутентификации биометрических параметров.

В связи с тем, что формирование полной базы естественных биометрических образов «Чужой» для тестирования средств высоконадежной биометрической аутентификации сопряжено с огромными затратами времени и иных материальных ресурсов допускается использовать неполные естественные базы [4]. Неполнота базы приводит к снижению достоверности тестирования, но во многих случаях вполне допустима.

При формировании неполной базы из N биометрических образов «Чужой» показатель неполноты базы можно определить по формуле (7):

$$\varphi = \frac{\log(N)}{\log(N_{Полн})}. \tag{7}$$

Неполные базы необходимо в дальнейшем при тестировании многократно увеличивать за счет дополнения их синтетическими биометрическими образами [3].

Требования к показателю неполноты формируемой базы могут определяться методом тестирования и методом синтеза дополняющих базу синтетических биометрических образов.

Нормированное распределение хи-квадрат достаточно качественно описывает распределения средней стабильности, уникальности и качества параметров биометрических образов [5]. Площади плотностей распределения для средней стабильности расходятся на 9,1 %, для средней уникальности – 15,7 % и на 9,6 % для среднего качества.

Снизить ошибку расхождения можно только перейдя от целого значения количества степеней свободы к дробному (фрактальному).

На рис. 4–6 приведены графики зависимости ошибки расхождения (площадь расхождения исходного распределения и распределения описывающей функции) от количества степеней свободы.

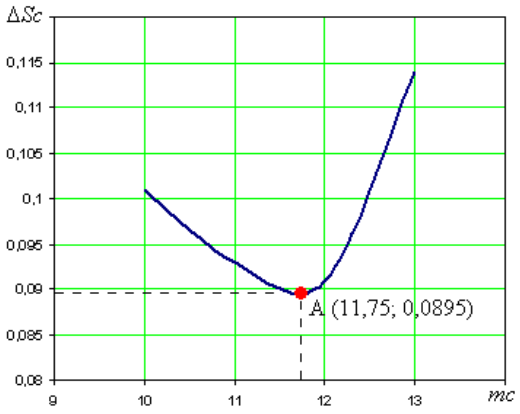


Рис. 4. График зависимости ошибки расхождения распределения от количества степеней свободы при оптимальном количестве степеней свободы в точке A(11,75; 0,089)

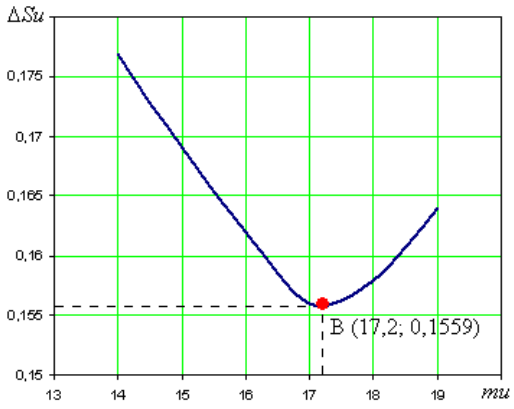


Рис. 5. График зависимости ошибки расхождения распределения от количества степеней свободы при оптимальном количестве степеней свободы в точке B(17,2; 0,156)

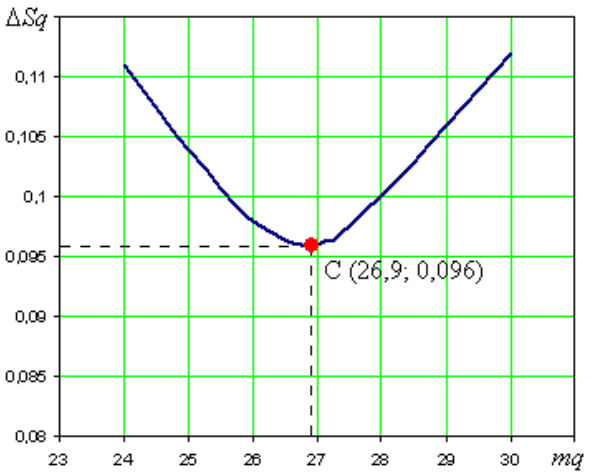


Рис. 6. График зависимости ошибки расхождения распределения от количества степеней свободы при оптимальном количестве степеней свободы в точке C(26,9; 0,096)

Точки A(11,75; 0,089), B(17,2; 0,156) и C(26,9; 0,096) показывают оптимальное количество степеней свободы при котором ошибка минимальна. Нормированное хи-квадрат распределение биометрических параметров позволяет правильно размножать обучающие или тестовые биометрические образы.

Также возможно определить условия, обеспечивающие равномерность заполнения синтетическими биометрическими образами поля выходных биометрических кодов.

Если образы представлены только одним примером, то для оценки их родства достаточно вычислить расстояние Хемминга между их кодами [6]. Сложнее обстоит дело при сравнении двух образов, представленных несколькими примерами. В этом случае необходимо найти центры сравниваемых образов по входам нейросетевого преобразователя биометрия-код или по его выходам [4, 7].

Центр по входам определяется простым усреднением входных биометрических данных по каждому из контролируемых биометрических параметров. Центр по входам можно рассматривать как среднестатистический пример исследуемого родительского образа. Если подать его на вход преобразователя биометрия-код, то на выходе должен появиться код, соответствующий центру кодов данного образа. К сожалению, эта простая конструкция не всегда работает правильно. Из-за нелинейной связи входного и выходного многомерных пространств, и плохой обусловленности задачи могут возникать существенные ошибки в определении центра выходных кодов исследуемого биометрического образа.

Действительно надежным и точным является только способ проявки центра кодов биометрического образа. В соответствии с этим способом необходимо контролировать число «0» и «1» в нестабильных разрядах, увеличивая число проверок через предъявление нейронной сети очередного примера образа. Результаты по каждому разряду следует накапливать до того момента пока число «1» или число «0» в серии надежно не превысит половину.

Очевидно, что число потомков должно быть привязано к среднему значению родственности множества всех размножаемых биометрических образов. То есть, в размножаемой группе биометрических образов необходимо вычислить среднее расстояние Хемминга между размножаемыми биометрическими образами – E(h) и функцию вероятности появления значений меры Хемминга – P(h) для размножаемых биометрических образов.

Располагая математическим ожиданием значений меры Хемминга – E(h), можно вычислить для него среднее число примеров, соответствующее, гипотетической ситуации, когда все размножаемые образы будут иметь одинаковое родство. Например, имеется 21 образ, а требуется получить 2100 образов.

Экспериментально подтверждено, что 21 образ дает 21(21-1)/2 неповторяющихся пар, что составляет число 210. Ограничимся использованием этих 210 пар, тогда каждая пара образов-родителей должна давать по 10 образов потомков. Расчет E(h) и P(h) для 210 пар биометрических образов является достаточно простой задачей невысокой вычислительной сложности.

В нашем примере расстоянию между образами-родителями, равному E(h) соответствует 10 образов-потомков. При меньшем расстоянии Хемминга число потомков должно уменьшаться, так как между родительскими образами растет родство. При большем, чем среднее E(h) расстояние между образами родителями число их потомков должно увеличиваться.

Пользуясь этим, удается рассчитать связь числа потомков с расстоянием Хемминга между центрами образов-родителей. Так как известно число примеров у образов-родителей со средним значением меры Хемминга, то для других пар образов (расстояния между которыми неравно среднему) число потомков вычисляется следующим образом (8):

$$N_{AB} = N_{средн} \cdot P(h_{AB}), \qquad (8)$$

где $P(h_{AB})$ – вероятность появлении в размножаемой выборке расстояния Хемминга – h_{AB}.

Выражением (8) следует пользоваться в случаях, когда функция распределения значений меры Хемминга существенно несимметрична (смещена относительно центра – половины максимального значения меры Хемминга выходных кодов преобразователя). В случае E(h) ≈ 0.5 *max*(*h*) плотность распределения значений меры Хемминга близка к нормальной [5].

Следовательно, в этом случае нет необходимости полностью восстанавливать функцию вероятности появления значений меры Хемминга – P(h), выражение (8) упрощается:

$$N_{AB} = N_{средн} \cdot \frac{1}{\sqrt{2\pi}\cdot\sigma(h)} \int_0^{h_{AB}} exp\{\frac{-(E(h)-h)^2}{2\sigma^2(h)}\}dh, \qquad (9)$$

где *б*(*h*) – среднеквадратическое отклонение меры Хемминга для размножаемых пар биометрических образов.

Таким образом, расчет числа потомков для каждой из пар размножаемых биометрических образов особых трудностей не вызывает.

При малых значениях меры Хемминга число потомков для близких образов-родителей оказывается нулевым. Для пар родителей со средним родством число потомков точно совпадает со средним значением числа потомков.

Для пар образов-родителей с очень малым родством дает двукратное увеличение числа потомков по сравнению со средним показателем.

Результаты вычислений числа ожидаемых потомков по выражению (9) дают нецелые, дробные оценки, которые должны округляться до ближайшего минимального целого значения.

Заключение

Нормированное хи-квадрат распределение можно достаточно эффективно использовать при аппроксимации асимметричных распределений биометрических параметров тестовых баз. При этом возможно выполнить условия, обеспечивающие равномерность заполнения синтетическими биометрическими образами поля выходных биометрических кодов.

Библиографический список

1. ГОСТ Р 52633.0–2006. Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации. – Москва : Стандартинформ, 2007. – 25 с.

2. ГОСТ Р 52633.1–2009. Защита информации. Техника защиты информации. Требования к формированию баз естественных биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации. – Москва : Стандартинформ, 2010. – 24 с.

3. ГОСТ Р 52633.2–2010. Защита информации. Техника защиты информации. Требования к формированию синтетических биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации. – Москва : Стандартинформ, 2011. – 17 с.

4. ГОСТ Р 52633.3–2011. Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора. – Москва : Стандартинформ, 2018. – 12 с.

5. Нейросетевая защита персональных биометрических данных / Ю. К. Язов, В. И. Волчихин, А. И. Иванов, В. А. Фунтиков, И. Г. Назаров. – Москва : Радиотехника, 2012. – 157 с.

6. Расстояние Хэмминга // Википедия. – URL: https://ru.wikipedia.org/wiki/Расстояние_Хэмминга

7. ГОСТ Р 52633.5–2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия – код доступа. – Москва : Стандартинформ, 2012. – 20 с.

Для цитирования:

Туреев, С. В. Использование нормированного критерия хи-квадрат для аналитического описания статистик асимметричных распределений биометрических параметров в тестовых базах / С. В. Туреев // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 168–178.

УДК 519.2

А. И. Солопов

АНАЛИТИЧЕСКОЕ ОПИСАНИЕ ИСКУССТВЕННЫХ НЕЙРОННЫХ СЕТЕЙ С ИСПОЛЬЗОВАНИЕМ ЛЕНТОЧНЫХ МАТРИЦ

Аннотация. Рассмотрена попытка аналитического описания искусственных нейронных сетей, построенных на некоторых общепринятых упрощениях, в частности с использованием ленточных матриц. В результате исследований выявлено, что синтезировать матрицы, делающие зависимыми данные, достаточно просто. Создать программную реализацию корреляторов также, оказывается, не сложно, однако намного труднее создать и сбалансировать качественный генератор случайных чисел.

A. I. Solopov

ANALYTICAL DESCRIPTION OF ARTIFICIAL NEURAL NETWORKS USING TAPE MATRIX

Abstract. An attempt to analytically describe artificial neural networks built on some common simplifications, in particular using tape matrix, is considered. Studies have shown that it is quite easy to synthesize matrixes that make data dependent. Creating a program implementation of correrators is also not difficult, however, it is much more difficult to create and balance a quality random number generator.

Наряду с численным описанием искусственных нейронных сетей [1, 2] достаточно большой интерес представляют попытки их аналитического описания, построенные на некоторых общепринятых упрощениях. Одним из таких упрощений является использование ленточных матриц [3, 4].

Переход к ленточным корреляционным матрицам осуществим достаточно просто. Для этого достаточно заменить полные коррелирующие матрицы на ленточные матрицы. Например, для простейшей коррелирующей матрицы подобная замена будет выглядеть следующим образом:

$$\begin{bmatrix} a & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & a & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & a & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & a & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & a & 1 & 1 \\ 1 & 1 & 1 & 1 & 1 & a & 1 \\ 1 & 1 & 1 & 1 & 1 & 1 & a \end{bmatrix} \Rightarrow \begin{bmatrix} ka & 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & a & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & a & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & a & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & a & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & a & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & ka \end{bmatrix}. \quad (1)$$

В правой матрице k – это коэффициент пересчета, учитывающий то, что первая и последняя строки этой матрицы имеют по два ненулевых элемента, все другие матрицы имеют по 3 ненулевых элемента. Коэффициент пересчета k выбирается по номограмме, исходя из заданного значения a, ширины ленты и требуемого значения r в конечной корреляционной матрице.

В частности, для a = 4, при ленте шириной из трех элементов и r = 0,52, коэффициент k = 0,92. В конечном итоге замена левой коррелирующей матрицы на правую коррелирующую матрицу этого же выражения приводит к изменениям конечных корреляционных матриц следующего вида:

$$\begin{bmatrix} 1 & r & r & r & r & r & r \\ r & 1 & r & r & r & r & r \\ r & r & 1 & r & r & r & r \\ r & r & r & 1 & r & r & r \\ r & r & r & r & 1 & r & r \\ r & r & r & r & r & 1 & r \\ r & r & r & r & r & r & 1 \end{bmatrix} \Rightarrow \begin{bmatrix} 1 & r & 0 & 0 & 0 & 0 & 0 \\ r & 1 & r & 0 & 0 & 0 & 0 \\ 0 & r & 1 & r & 0 & 0 & 0 \\ 0 & 0 & r & 1 & r & 0 & 0 \\ 0 & 0 & 0 & r & 1 & r & 0 \\ 0 & 0 & 0 & 0 & r & 1 & r \\ 0 & 0 & 0 & 0 & 0 & r & 1 \end{bmatrix}. \quad (2)$$

Подчеркнем, что при таком подходе формирования связанных данных ширина ленты конечной корреляционной матрицы может быть любой от 1 до n, где n – порядок корреляционной матрицы. Кроме того, все преобразования с полными коррелирующими матрицами, могут быть распространены и на ленточные матрицы.

Соответственно, ленточные корреляционные матрицы могут иметь одинаковые и случайные элементы лент, элементы с одинаковыми и разными знаками.

Одним из наиболее часто используемых приемов аналитического упрощения является сведение реальных связей и процессов к зависимым процессам. В связи с этим рассмотрим синтез векто-

ров, зависимых данных с Марковской корреляционной матрицей [5–7]. Подобные связи получаются, если формировать данные следующим образом:

$$\begin{cases} v_1 = x_1, \\ v_2 = (x_2 + a \cdot v_1) / \sqrt{1 + a^2}, \\ v_3 = (x_3 + a \cdot v_2) / \sqrt{1 + a^2}, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ v_k = (x_k + a \cdot v_{k-1}) / \sqrt{1 + a^2}, \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \end{cases} \tag{3}$$

где x_k – это случайные числа, полученные от генератора независимых случайных чисел с нормальным законом распределения значений, нулевым математическим ожиданием и единичной дисперсией.

Преобразования дают корреляционную матрицу с убыванием значений коэффициентов корреляции пропорционально степени этих коэффициентов, причем степень определяется расстоянием коэффициента корреляции от диагонали:

$$\left[r_{v_i v_j} \right] = \begin{bmatrix} 1 & r & r^2 & \vdots\vdots\vdots & r^n \\ r & 1 & r & \vdots\vdots\vdots & r^{n-1} \\ r^2 & r & 1 & \vdots\vdots\vdots & r^{n-2} \\ \vdots\vdots\vdots & \vdots\vdots\vdots & \vdots\vdots\vdots & \vdots\vdots\vdots & \vdots\vdots\vdots \\ r^n & r^{n-1} & r^{n-2} & \vdots\vdots\vdots & 1 \end{bmatrix}. \tag{4}$$

Связь коэффициента корреляции – r и задаваемого параметра – a приведена на графике рис. 1.

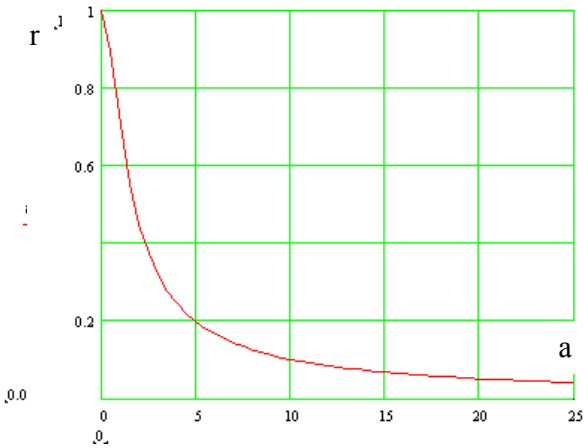


Рис. 1. График связи задаваемого параметра «a» и значений коэффициентов Марковской корреляционной матрицы

Следует отметить, что связь задаваемого коэффициента «а» и коэффициентов корреляции достаточно проста. Порядок корреляционной Марковской матрицы или длина генерируемого вектора не влияют на функцию связи r(a).

В том случае, если регулируемые параметры в Марковском корреляторе сделать разными,

$$\begin{cases} v_1 = x_1, \\ v_2 = (x_2 + a_1 \cdot v_1) / \sqrt{1 + {a_1}^2}, \\ v_3 = (x_3 + a_2 \cdot v_2) / \sqrt{1 + {a_2}^2}, \\ \vdots\vdots\vdots\vdots\vdots\vdots\vdots\vdots\vdots\vdots \\ v_k = (x_k + a_{k-1} \cdot v_{k-1}) / \sqrt{1 + {a_{k-1}}^2}, \\ \vdots\vdots\vdots\vdots\vdots\vdots\vdots\vdots\vdots\vdots \end{cases} \qquad (5)$$

то получим гораздо более сложную корреляционную матрицу связей между генерируемыми биометрическими параметрами

$$\left[r_{v_i v_j} \right] = \begin{bmatrix} 1 & r_1 & {r_{1,2}}^2 & \vdots\vdots\vdots & (r_{1,n-1})^n \\ r_1 & 1 & r_2 & \vdots\vdots\vdots & (r_{2,n-2})^{n-1} \\ {r_{1,2}}^2 & r_2 & 1 & \vdots\vdots\vdots & (r_{3,n-3})^{n-2} \\ \vdots\vdots\vdots & \vdots\vdots\vdots & \vdots\vdots\vdots & \vdots\vdots\vdots & \vdots\vdots\vdots \\ (r_{1,n-1})^n & (r_{2,n-2})^{n-1} & (r_{3,n-3})^{n-2} & \vdots\vdots\vdots & 1 \end{bmatrix}. \qquad (6)$$

Вывод: синтезировать матрицы, делающие зависимыми данные достаточно просто. Создать программную реализацию корреляторов, оказывается не сложно, намного труднее создать и сбалансировать качественный генератор случайных чисел.

Библиографический список

1. Галушкин, А. И. Синтез многослойных систем распознавания образов / А. И. Галушкин. – Москва : Энергия, 1974.

2. Нейроинформатика / Е. М. Миркес, А. Н. Горбань, В. Л. Дунин-Барковский, А. Н. Кирдин, А. Ю. Новоходько, Д. А. Россиев, С. А. Терехов, М. Ю. Сенашова, В. Г. Царегородцев. – Новосибирск : Наука : Сибирское предприятие РАН, 1998. – 296 с.

3. Ланкастер, П. Теория матриц / П. Ланкастер. – Москва : Наука, 1973.

4. Гантмахер, Ф. Р. Теория матриц / Ф. Р. Гантмахер. – 5-е изд. – Москва : Физматлит, 2004. – 560 с.

5. Рухин, А. Л. Корреляционные матрицы цепочек для марковских последовательностей и тестирование случайности / А. Л. Рухин // Теория вероятности и ее применение. – 2006. – № 4 (51). – С. 712–731.

6. Марков, А. А. Распространение закона больших чисел на величины, зависящие друг от друга / А. А. Марков // Известия физико-математического общества при Казанском университете. – 2-я серия. – 1906. – Т. 15. – С. 135–156.

7. Яглом, А. М. Вероятность и информация / А. М. Яглом, И. М. Яглом. – Москва : Наука, 1973. – 512 с.

Для цитирования:

Солопов, А. И. Аналитическое описание искусственных нейронных сетей с использованием ленточных матриц / А. И. Солопов // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 179–183.

УДК 004.934; 004.056.523

А. А. Афанасьев

ФОРМИРОВАНИЕ ВЫБОРОК ПАРАМЕТРОВ РЕЧЕВОГО СИГНАЛА ПРИ ОБУЧЕНИИ СИСТЕМЫ БИОМЕТРИЧЕСКОЙ ИДЕНТИФИКАЦИИ ДИКТОРА

Аннотация. Рассмотрена задача разработки системы идентификации диктора. Предложен метод декомпозиции речевого сигнала при формировании выборок параметров речевого сигнала. Использование данного подхода при предварительном обучении системы позволит повысить качество идентификации пользователя при доступе к ресурсу инфокоммуникационной системы.

A. A. Afanasev

FORMING OF SPEECH SIGNAL PARAMETERS SELECTIONS AT TRAINING THE SYSTEM OF BIOMETRIC IDENTIFICATION OF THE ANNOUNCER

Abstract. The problem of the identification system development of the announcer is considered. The decomposition technique of a voice signal at forming selections of parameters of a voice signal is offered. Use of this approach at preliminary training of a system will allow to increase quality of user identification at access to a resource of an infocommunication system.

При доступе к ресурсу инфокоммуникационной системы часто необходима предварительная идентификация пользователя с целью дальнейшего доступа к ресурсу пропускной способности и сервисов инфокоммуникационной сети связи. Для выполнения данной процедуры в существующих системах в настоящее время может использоваться однократное или непрерывное подтверждение легитимности пользователя на этапе вхождения и работы в инфокоммуникационной системе. Одним из частных вариантов реализации биометрической идентификации пользователя являются его индивидуальные голосовые особенности.

При обучении системы биометрической идентификации диктора одной из ключевых задач является правильное формирование выборок параметров речевого сигнала (РС). Для решения данной задачи предлагается осуществлять правильную сегментацию РС,

что приведет к уменьшению межсегментной корреляции между параметрами выделяемыми на отдельном сегменте.

В основной массе рекомендаций по обработке РС используется подход, основанный на определении значений параметров модели речеобразования путем кратковременного анализа сегментов речи фиксированной длительности от 10 до 30 мс [1]. Фиксированный сегмент анализа речевых данных используется в большинстве систем обработки РС, что является существенным недостатком данных устройств. При этом использование фиксированных сегментов анализа при реализации систем обработки РС на которых предполагается наличие квазистационарных участков не соответствует особенностям природы возникновения и существования РС. В качестве ограничений при обработке РС полагается, что параметры РС изменяются с течением времени достаточно медленно, что позволяет рассматривать РС как стационарный на временных интервалах порядка 2,5–60 мс, называемых "окнами" или сегментами. Введение динамически изменяемой длительности сигнала может быть осуществлено при использовании выражений (1)–(3).

$$\Delta T \in (t_{\text{н}} ... t_{\text{k}}) = \text{var}; \Delta T > 20\,\text{мс}; \frac{\Delta \text{T}}{\text{T}_{\text{от}}} \in \{N\}; \quad (1)$$

$$t_{\text{н}} = \Delta tn, \text{при sign}(S(t_{\text{н}})) + sign(S(\Delta \text{t(n-1)})) = 0; \quad (2)$$

$$t_{\text{k}} = \Delta tk, \text{при sign}\,(S(t_k)) = \text{sign}(S(t_{\text{н}})) \,\&\, signS(\Delta \text{t(k-1)}) = \\ = sign(S(\Delta \text{t(n-1)})) \,\&\, (t_k - t_{\text{н}}) > 20\,\text{мс}, \quad (3)$$

где ΔT – длительность сегмента анализа; $t_{\text{н}}$, $t_{\text{к}}$ – время начала и окончания сегмента анализа; $T_{\text{от}}$ – период ОТ; n – номер отсчета в начале сегмента; k – номер отсчета в конце сегмента; Δt – интервал дискретизации, N – область натуральных чисел.

Использование данных выражений позволяет получать сегменты, отсчеты которых в начале и окончании будут иметь одинаковые знаки конечной разности первого порядка, такой подход является новым направлением при выделении сегментов анализа. При этом с высокой вероятностью можно утверждать, что начальный и конечный отсчеты во вновь сформированном сегменте будут иметь значения, близкие к моменту изменения полярности знака отсчетов РС, что значительно уменьшит возможные искажения на стыках сегментов. Техническое решение описано в [2], показано на рис. 1 и является новым направлением развития систем обработки РС.

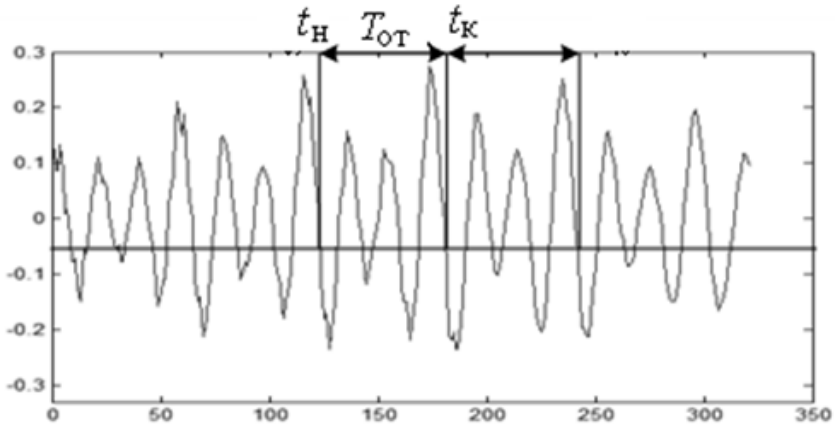


Рис. 1. Динамическое выделение сегмента анализа речевого сигнала

При таком техническом решении вычисляемые для обучения системы идентификации параметры передаточной функции голосового тракта считаются неизменными, что дает возможность ввести в использование понятие «однородного сегмента РС». Таким образом, однородность сегмента РС связана с природой речеобразования и заключается в малом отличии параметров передаточной функции голосового тракта при моделировании РС на основе данного подхода.

На шумоподобных участках РС формирование границы окончания сегмента анализа реализуется в соответствии с вышеизложенным подходом на интервалах соответствующих длительности около 20 мс. Таким образом, возможно использование более длительных интервалов анализа, особенно на сегментах имеющих квазипериодическую вокализованную природу образования, однако может возникать ошибка в определении участков однородности при переходе между звуками одного фонетического ряда. В задачах прикладного характера распределение значений РС априорно является случайным, более того сам вид распределения остаѐтся неизвестным. Увеличение границ анализируемого сегмента может быть основано на использовании более полной информации о статистических взаимосвязях между мгновенными значениями РС, получаемыми с помощью непараметрических методов, то есть при отсутствии данных о характере распределения. При адаптивном изменении границ сегмента анализа РС, количество наблюдаемых отсчѐтов, составляющих данный участок, априорно является неизвестной величиной. Оно может принимать различные значения в зависимости от особенностей произносимой речи. Соответственно количество наблюдений (отсчѐтов), составляющих анализируемый участок, зависит в данном случае от исхода самих наблюдений. Следовательно, для определения длительности сегмента анализа необходимо

применять методы последовательной проверки статистических гипотез, суть которых сводится к определению «эффективных» выборок при которых будет иметь место расширение анализируемого сегмента, и, как следствие, увеличение набора отсчетов. В данном случае предлагается метод, основанный на последовательном критерии отношения вероятностей (критерий Вальда), которая заключается в вычислении отношение вероятностей получения выборок (функций правдоподобия) на каждом этапе [3].

В основе данного технического решения лежит тот факт, что однородные участки речевого сигнала имеют неизменную функцию плотности вероятности. При каждом увеличении сегмента на длительность периода основного тона на основе вновь полученного набора отсчѐтов РС производятся следующие операции [4]:

1. Вычисляется значение правдоподобия $L_0(S,m)$ при справедливости гипотезы H_0 – основной гипотезы.

2. Производится ядерная оценка плотности распределения $f_1(S,m)$ участка речевого сигналаf1si,m для основной гипотезы.

3. Производится вычисление значения правдоподобия $L_1(S,m)$ – альтернативной гипотезыL1S,m по $f_1(S,m)$ f1si,m. Далее вычисляется статистика критерия Вальда:

$$\ln\frac{L_1(S,m)}{L_0(S,m)} = \ln\left(L_1(S,m) - \ln L_0(S,m)\right) = Z\left[S\right], \qquad (4)$$

При этом анализируемые выборки (сегменты) принадлежат к «эффективной» области Сэф пространства выборок при выполнении $\ln B\big|_{n<n'} < Z\left[S\right] < \ln A\big|_{n<n'}$, lnBn<n'<Z[S]<lnAn<n'где n – общее количество отсчѐтов в анализируемом сегменте, n'- максимально возможное число отсчѐтов в сегменте, с учѐтом ограничений G.114 МСЭ. В случае $Z\left[S\right] < \ln B\big|_{n<n'}$ ZS<lnBn<n', набор отсчѐтов принадлежит к «абсолютно эффективной» области $C_{эф}^{абс}$ Сэфабс (сегменту анализа). Оставшийся вариант при ZS > lnA $Z\left[S\right] > \ln A\big|_{n<n'}$ определяет набор отсчѐтов принадлежащих к «неэффективной» области $C_{н.эф}$ Сн.эф (сегменту анализа). Ограничительные константы *A и* B, определяются на основании ошибок первого α и второго β рода:

$$\begin{cases} \ln A = \ln\dfrac{1-\beta}{\alpha}, \\ \ln B = \ln\dfrac{\beta}{1-\alpha}, \end{cases} \qquad (5)$$

Анализ функции отношения вероятностей позволяет выделить важный показатель, характеризующий РС. Это интервал, являющийся промежутком времени, в течение которого отсчеты статистически взаимосвязаны, следовательно, являются участками однородности при формировании РС. Данный факт соотносится с результатами анализа образования вокализованных и шумоподобных сигналов [5].

На выделенных сегментах параметры передаточной функции голосового тракта не претерпевают значительного изменения, что дает возможность использовать их фиксированные значения при моделировании более длительных участков речевого сигнала,чем это возможно при фиксированных длительностях сегментов анализа [5]. Такой подход показывает пути уменьшения объема данных при подготовке данных о РС, что приводит к получению необходимого и достаточного количества параметров для его качественного представления при обработке.

Формирование переменной длины сегмента анализа РС дает возможность по-новому рассмотреть большинство подходов к обработке РС и выделению его характеристик для обучения системы идентификации. При их оценке в качестве параметров для обучения предлагается использовать линейные спектральные частоты (ЛСЧ), которые описывают поведение передаточной функции голосового тракта на сегменте анализа [6].

Рассмотрим полином знаменателя передаточной функции синтезирующего фильтра $A_M(z)$ (6):

$$A_M(\mathrm{Z}) = 1 - \sum_{m=1}^{M} a_m z^{-m} = \prod_{m=1}^{M} (1 - z_{0m} z^{-1}); \qquad (6)$$

где z_{0m} – корни полинома $A_M(z)$.

Общий порядок расчета ЛСЧ заключается в следующем. Из полинома передаточной функции образуются новые полиномы $P(Z)$ и $Q(Z)$ (7).

$$\begin{cases} P(Z) = A_M(Z) + Z^{-(M+1)} A_M(Z^{-1}); \\ Q(Z) = A_M(Z) - Z^{-(M+1)} A_M(Z^{-1}); \end{cases} \qquad (7)$$

При таком подходе полином P(z) является симметричным, а полином Q(z) антисимметричным. Для того чтобы вернуться к исходному полиному, определяющему передаточную функцию голосового тракта, необходимо выполнить (8).

$$A_M(\mathrm{Z}) = \frac{P(Z) + Q(Z)}{2}. \qquad (8)$$

Полиномы в выражении (2) имеют тривиальные вещественные корни: $z_{oP} = -1$ и $z_{oQ} = 1$, для исключения которых нужно разделить $P(z)$ на $1 + z^{-1}$, а $Q(z)$ на $1 - z^{-1}$. Тривиальным корням полиномов (2) будут соответствовать значения Z лежащие на единичной окружности на соответствующих угловых частотах 0 и π.

Пример расположения ЛСК на единичной окружности и на частотной оси для $M = 4$, показан на рис. 2, где точками отображены корни $P(z)$, а кружочками – корни полинома $Q(z)$.

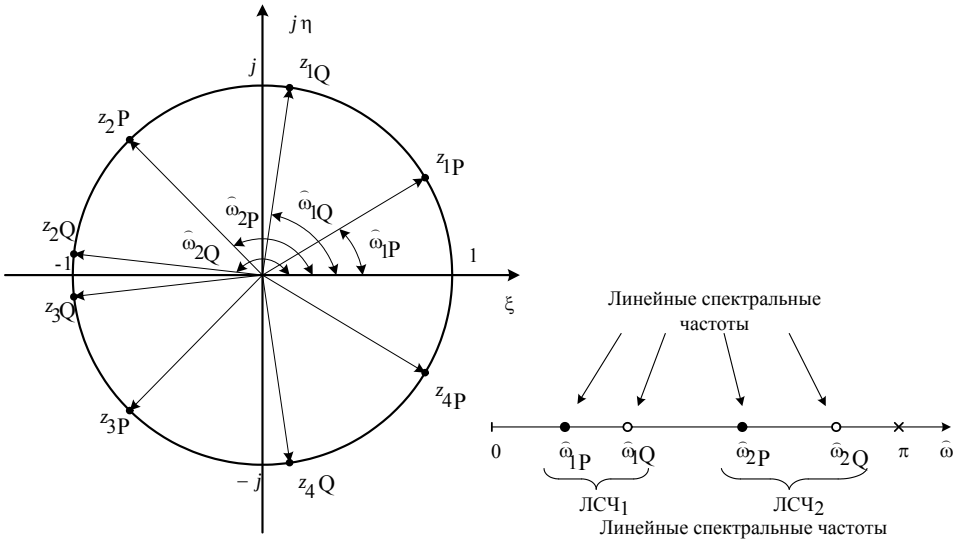


Рис. 2. Пример расположения нулей полиномов при $M = 4$

Корни полученных полиномов (7) будут являться (ЛСЧ). Исследования показали, что ЛСЧ на участках РС, расширенных с помощью предложенной модели, не претерпевают значительных изменений по отношению к начальному в *N* отсчетов (табл. 1).

Таблица 1

Значения ЛСЧ (в радианах) на однородных сегментах

N	1:196	1:246	1:296	1:346
	0,252	0,252	0,251	0,251
	0,37	0,371	0,371	0,372
	0,474	0,474	0,474	0,476
	0,618	0,62	0,619	0,618
	1	0,998	0,989	0,983
	1,664	1,623	1,629	1,607
	1,98	1,978	1,977	1,974
	2,108	2,108	2,102	2,1
	2,544	2,538	2,544	2,537
	2,82	2,827	2,822	2,823

При этом значительно ослабевают корреляционные связи между соответствующими наборами ЛСЧ при переходе от сегмента к сегменту, что позволяет значительно снизить вычислительную сложность алгоритма обучения системы идентификации пользователя, сформировав при этом более наглядный аудио портрет пользователя инфокоммуникационной системы. Иллюстрация результатов исследований приведена на рис. 3.

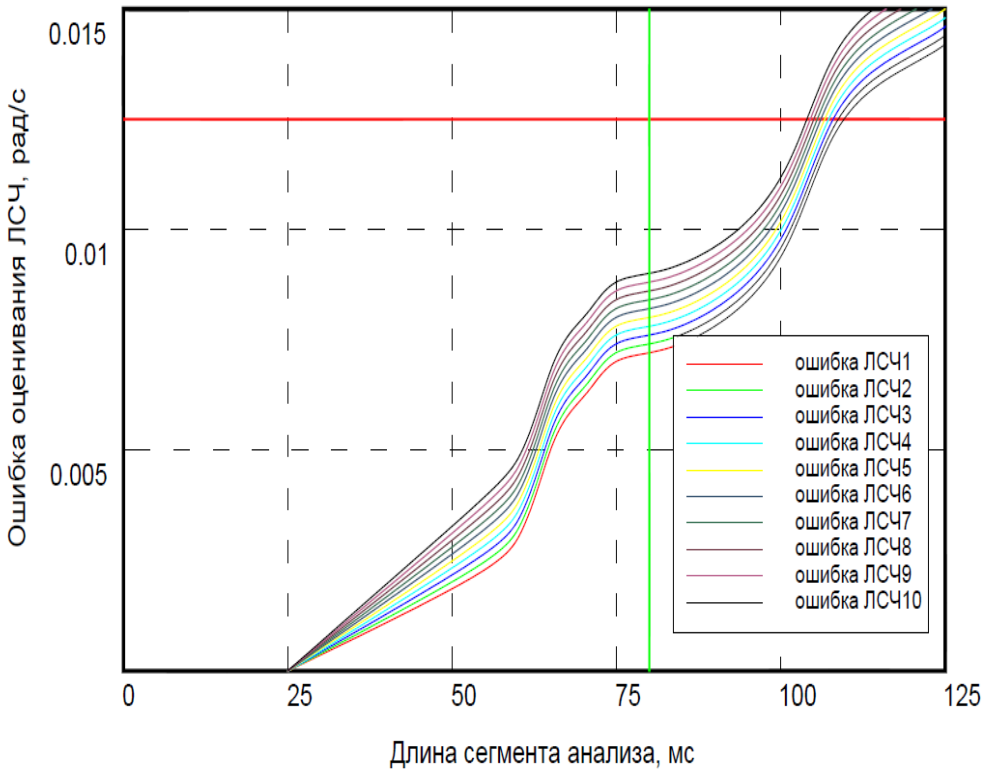


Рис. 3. Ошибка оценки линейных спектральных частот на однородных сегментах по отношению к начальному

Предлагаемое техническое решение позволило использовать значения ЛСЧ для обучения системы идентификации диктора, при этом они вычисляются для всего анализируемого участка РС, границы которого адаптивно расширены непараметрическим способом, основанным на методах последовательной проверки статистических гипотез.

Перспективным направлением развития данного направления является использование мел-кепстральных коэффициентов. Задачи обработки РС шумоподавление, низкоскоростное кодирование, распознавание, идентификация гораздо эффективнее и правильнее решать, учитывая особенности слуха человека и общее психоакустическое восприятие РС человеком.

За счет нелинейной шкалы соотношения частот используемой при оценке мел-кепстральных коэффициентов наложение равномерных ОФ в области мел-шкалы приводит к высокому «оконному» разрешению (малой ширине окна) в области низких частот и низкому «оконному» разрешению (большой ширине окна) в области высоких частот для линейной частотной шкалы. По сути осуществляется приведение спектра РС к значениям воспринимаемым слуховым аппаратом человека, в частности базилярной мембраной, при этом осуществляется сокращение признакового пространства параметров описывающих РС на сегменте анализа.

Таким образом, представленные технические решения по обработке РС делают адекватным их применение для подготовки выборок параметров с целью обучения системы идентификации диктора. До недавнего времени существующие подходы к сегментации РС и выделению его характеристик не соответствовали теории речеобразования и восприятия РС органами слуха, что не только уменьшало преимущества используемых методов обучения систем идентификации диктора, но и создавало предпосылки повышения вероятности ошибки первого рода систем разграничения доступа по голосу.

Современное состояние инфокоммуникационных систем и тенденции их развития свидетельствуют о перспективности представленных технических решений по модернизации и совершенствованию алгоритмов формирование выборок параметров речевого сигнала при обучении системы биометрической идентификации диктора.

Библиографический список

1. Рихтер, С. Г. Кодирование и передача речи в цифровых системах подвижной радиосвязи : учеб. пособие для вузов / С. Г. Рихтер. – Москва : Горячая линия – Телеком, 2010. – 304 с.

2. Пат. 2445718 Российская Федерация, МПК G10L 19/00. Способ выделения сегментов обработки речи на основе анализа корреляционных зависимостей в речевом сигнале / Афанасьев А. А., Новиков Е. И., Трубицын В. Г., Титов О. Н. ; заявитель и патентообладатель Академия ФСО России. – № 2010136618/08 ; заявл. 31.08.2010 ; опубл. от 20.03.2012, Бюл. № 8.

3. Вальд, А. Последовательный статистический анализ / А. Вальд ; под ред. А. Ф. Лапко. – Москва : Физматлит, 1960. – 328 с.

4. Косарев, Е. Л. Методы обработки экспериментальных данных / Е. Л. Косарев. – Москва : Физматлит, 2008. – 208 с.

5. Афанасьев, А. А. Вычисление линейных спектральных частот на однородных участках речевого сигнала / А. А. Афанасьев, Р. С. Власов // Современные технологии в науке и образовании : сб. тр. II Междунар. науч.-техн. форума. – Рязань : РГРТУ, 2019. – Т. 1. – С. 109–112.

6. Афанасьев, А. А. Алгоритмы обработки речевого сигнала при переменной длительности сегмента анализа / А. А. Афанасьев, Р. С. Власов, В. Г. Лисичкин, А. В. Питолин // Вестник Воронежского государственного технического университета. – 2019. – Т. 15, № 4. – С. 41–48.

Для цитирования:

Афанасьев, А. А. Формирование выборок параметров речевого сигнала при обучении системы биометрической идентификации диктора / А. А. Афанасьев // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 184–192.

УДК 004.056; 004.89; 57.087

В. Д. Платонов, В. Ю. Киселёв, А. П. Иванов

РАЗРАБОТКА УЧЕБНОГО АППАРАТНО-ПРОГРАММНОГО КОМПЛЕКСА БЕСПРОВОДНОЙ ПЕРЕДАЧИ ДАННЫХ

Аннотация. Разработана структура учебного аппаратно-программного комплекса для исследования помехозащищенности беспроводной передачи данных. Проведено тестирование работоспособности разработанного стенда. Экспериментально получена зависимость коэффициента правильного приема пакета от соотношения сигнал/помеха.

V. D. Platonov, V. Yu. Kiselev, A. P. Ivanov

DEVELOPMENT OF EDUCATIONAL HARDWARE AND SOFTWARE COMPLEX FOR WIRELESS DATA TRANSMISSION

Abstract. A structure of the educational hardware-software complex for studying the noise immunity of wireless data transmission has been developed. Testing of the developed bench operation has been conducted. The dependence of the correct packet reception ratio on the signal-to-interference ratio was experimentally obtained.

В настоящее время технологии беспроводной передачи данных стали гораздо чаще встречается в нашей жизни. В современном мире почти не осталось электронных устройств, которые не связываются между собой по одному из беспроводных протоколов или не имеют выхода в глобальную сеть интернет. И этому не нужно удивляется, беспроводные технологии имеют ряд преимуществ перед проводными аналогами.

Главными плюсами являются простота, скорость и стоимость развертывания сетей передачи данных. Там, где используются километры дорогостоящих проводов, при беспроводной связи обходятся всего парой приѐмо-передающих станций, так же где нет возможность организовать проводную связь из-за рельефа местности, водоемов или других факторов беспроводная сеть просто организуется.

В тоже время беспроводные технологии имеют и недостатки. Одним из главных является низкая помехозащищенность, а как следствие и более низкая надежность связи по сравнению с проводными технологиями. Помехи или атаки могут быть вызваны как другими устройствами, работающими в данных диапазонах частот,

так и злоумышленниками, которые хотят нарушить целостность и доступность передаваемой информации.

Исходя из этого была определена цель проекта - разработка учебного аппаратно-программного комплекса для исследования помехозащищенности беспроводной передачи данных, который позволит студентам получать практические навыки и сформировать представление о работе беспроводной передаче данных.

Была рассмотрена следующая классификация видов атак на компоненты телекоммуникационных систем (ТКС) [1]:

– атаки, направленные на перехват передаваемой информации;

– атаки, направленные на блокирование приема информации;

– атаки, направленные на модификацию передаваемой информации;

– атаки, направленные на фальсификацию передаваемой информации.

В аппаратно-программном комплексе было решено реализовать наиболее опасный вид атак, от которых сложнее всего защитится, так как он реализуется на физическом уровне – атак, направленных на блокирование приема информации.

Готовые решения для исследования помехозащищенности в беспроводных каналах передачи данных уже существуют, но у них много минусов. Они очень дорогостоящие, а так же эти системы сложно приспособить для решения разных задач. Их аппаратная база исключает возможность использования своих протоколов передачи данных, различных алгоритмов шифрования и т.д. В разработанном комплексе отсутствуют указанные недостатки. Он позволяет решать задачи необходимые для приобретения студентами необходимых компетенций [2].

Разработанный аппаратно-программный комплекс состоит из приемной части, передающей части и генератора помех, который имитирует действия злоумышленника. Разработанная структура комплекса представлена на рис. 1.

Аппаратно-программный комплекс состоит из следующих компонентов:

– рабочее место студента, передающего информацию (АРМ 1);

– рабочее место студента, принимающего информацию (АРМ 2);

– передающая часть (передатчик);

– приемная часть (приемник);

– генератор атак (помех).

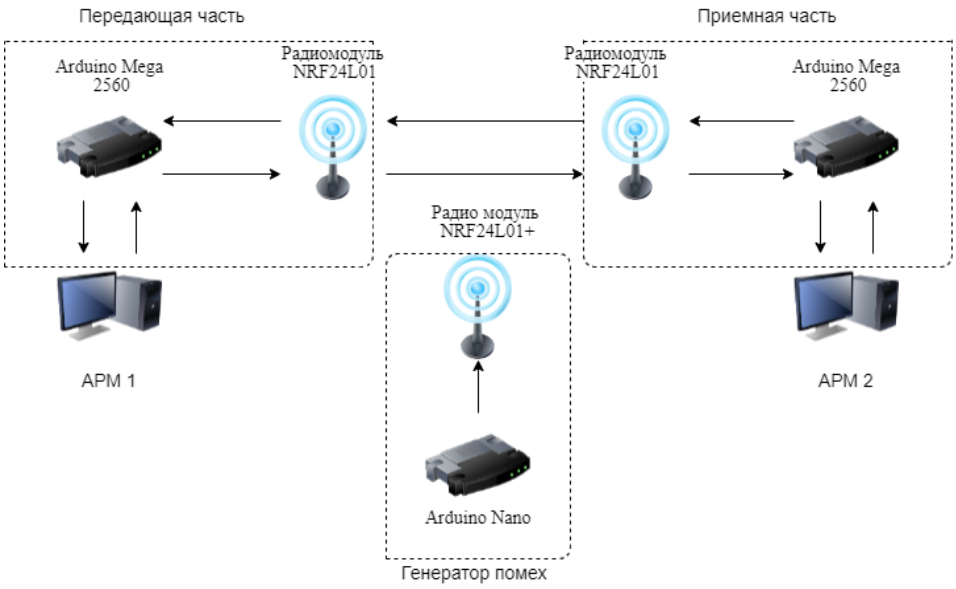


Рис. 1. Структура аппаратно-программного комплекса

АРМ 1 и АРМ 2 представляют собой персональные компьютеры (ПК) на базе ОС Windows с установленным на них разработанным программным обеспечением (ПО) с помощью которого осуществляется сопряжение всех устройств системы и управление комплексом.

Основные технические характеристики разработанного аппаратно-программного комплекса:

– диапазон частот от 2,4 до 2,5 ГГц;
– скорость передачи данных 256, 1024 и 2048 кбит/с;
– размер пакета от 1 до 32 байт;
– мощность передатчика –18, –12, –6 и 0 дБм;
– возможность повторной передачи пакета;
– число повторов пакета от 1 до 10000.

Графический интерфейс ПО комплекса представлен на рис. 2.

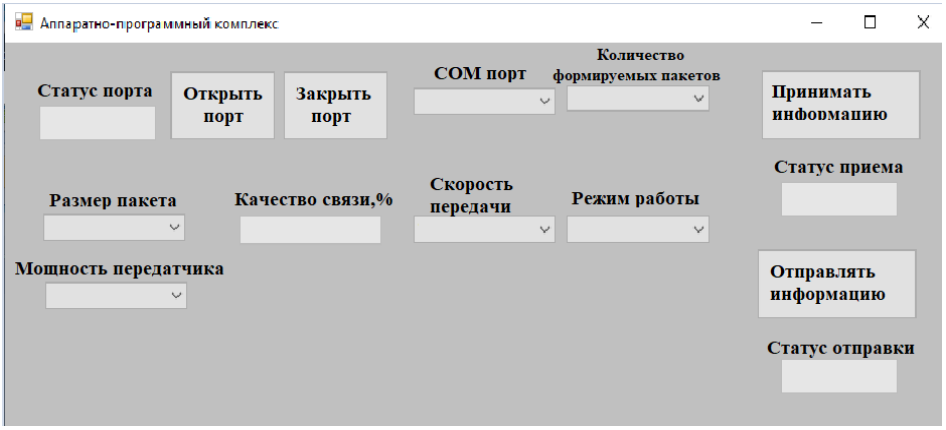


Рис. 2. Графический интерфейс ПО комплекса

Как видно на рис. 1, передатчик и приемник комплекса построен на одинаковой аппаратной базе – программируемой плате Arduino MEGA 2560 и радио модуля NRF24L01.

Основу программируемой платы Arduino MEGA 2560 составляет микроконтроллер ATmega2560 работающий с тактовой частотой 16 МГц. Вычислительной мощности данного микроконтроллера вполне достаточно для обеспечения согласованной работы всех элементов системы.

Радиомодуль NRF24L01 являются легко программируемым, надежным и дешевыми средством беспроводной передачи данных. Он может работать в полудуплексном режиме (одновременно отправлять и принимать сообщения), в сети связи возможна одновременная передача информации на 8 таких же устройств по каждому из 127 каналов связи, в него уже встроены приемный и передающий модули. Дальность связи NRF24L01 без потери качества составляет 100 м. Работает модуль в диапазоне частот от 2,4 до 2,5 ГГц с шагом в 0,1 ГГц.

Генератор атак реализован программируемой плате Arduino NANO на базе микроконтроллера ATmega3280, так как вычислительная мощность Arduino MEGA 2560 была бы избыточна, а ее стоимость значительно выше. В качестве радиомодуля выбран NRF24L01+. Он отличается от радиомодуля NRF24L01 наличием дополнительного усилителя и внешней антенной, что позволяет изменять мощность выходного сигнала атаки для установки необходимого соотношения сигнал/шум в ходе проведения исследований помехозащищенности беспроводной передачи данных.

Авторы считают, что:

– создание учебного аппаратно-программного комплекса для исследования помехозащищенности беспроводной передачи данных и практика его использования в образовательном процессе позволяют сделать вывод об оправданности и необходимости создания аппаратно-программных средств обучения, имитирующих действие реально используемых в предметных областях телекоммуникационных систем в случае, когда прямое обращение к ним невозможно;

– в результате использования учебного аппаратно-программного комплекса для исследования помехозащищенности беспроводной передачи данных студенты получают умения и отрабатывают навыки в части обязательных для специальности 10.05.02 «Информационная безопасность телекоммуникационных систем» компетенций [3]:

– способность формулировать задачи, планировать и проводить исследования, в том числе эксперименты и математическое моделирование объектов, явлений и процессов телекоммуникационных систем, включая обработку и оценку достоверности их результатов (ПК-2);

– способность оценивать технические возможности и вырабатывать рекомендации по построению телекоммуникационных систем и сетей, их элементов и устройств (ПК-3);

– способность проектировать защищенные телекоммуникационные системы и их элементы, проводить анализ проектных решений по обеспечению заданного уровня безопасности и требуемого качества обслуживания, разрабатывать необходимую техническую документацию с учетом действующих нормативных и методических документов (ПК-5);

– способность проводить анализ эффективности технических и программно-аппаратных средств защиты телекоммуникационных систем (ПК-8).

Библиографический список

1. Иванов, А. П. Анализ моделей атак нарушителя на компоненты телекоммуникационных систем / А. П. Иванов, Е. Д. Кашаев // Информация и безопасность. – 2013. – № 3. – С. 439, 440.

2. Власов, М. В. Разработка стенда для исследования атак нарушителя на компоненты телекоммуникационных систем / М. В. Власов, Д. В. Малашкин, А. П. Иванов // Информационные технологии в науке и образовании. Проблемы и перспективы : сб. науч. ст. Всерос. межвуз. науч.-практ. конф. (г. Пенза, 14 марта 2018 г.) / под ред. Л. Р. Фионовой. – Пенза : Изд-во ПГУ, 2018. – С. 202–204.

3. Об утверждении федерального государственного образовательного стандарта высшего образования по специальности 10.05.02 Информационная безопасность телекоммуникационных систем (уровень специалитета) : приказ Министерства образования и науки РФ № 1426 от 16 ноября 2016 г. – URL: http://fgosvo.ru/uploadfiles/fgosvospec/100502.pdf (дата обращения: 24.05.2020).

Для цитирования:

Платонов, В. Д. Разработка учебного аппаратно-программного комплекса беспроводной передачи данных / В. Д. Платонов, В. Ю. Киселѐв, А. П. Иванов // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 193–197.

УДК 519.2

Н. Н. Вершинин, А. С. Боровский,
И. В. Урнев, Г. К. Чистова

ЧИСЛЕННАЯ МОДЕЛЬ ИДЕАЛЬНОГО НЕЙРОСЕТЕВОГО ПРЕОБРАЗОВАТЕЛЯ БИОМЕТРИЯ-КОД

Аннотация. Идеальный нейросетевой преобразователь биометрия-код является основой для разработки методики повышения достоверности автоматизированного прогнозирования вероятности ложной аутентификации «Чужого». Получены распределения значений меры Хэмминга кодов-откликов «Чужой» идеального преобразователя, достоверно описываемые классическим биномиальным законом распределения независимых данных.

N. N. Vershinin, A. S. Borovsky, I. V. Urnev, G. K. Chistova

NUMERICAL MODEL OF THE IDEAL NEURAL CONVERTER BIOMETRICS-CODE

Abstract. The ideal biometric code converter is the basis for developing a technique to improve the reliability of automated prediction of the probability of false authentication of "Alien". The distributions of the values of the "Alien" response codes of the ideal converter are reliably described by the classical binomial law of distribution of independent data.

Схема моделирования закона распределения меры Хемминга кодов-откликов «Чужой» идеального преобразователя биометрия-код приведена на рис. 1. Для «чистоты» эксперимента не будем учитывать погрешности генераторов случайных чисел, представив их идеальными [1].

На схеме рис. 1 блок-1 генерирует N векторов независимых нормально распределенных данных, эмулируя множество биометрических образов «Чужие».

Данные с идеального генератора поступают на вход идеального преобразователя, заранее обученного распознавать множество биометрических образов «Свой».

При обучении используются биометрические образы пользователя «Свой», биометрические образы пользователя «Чужой» и заданный код аутентификации.

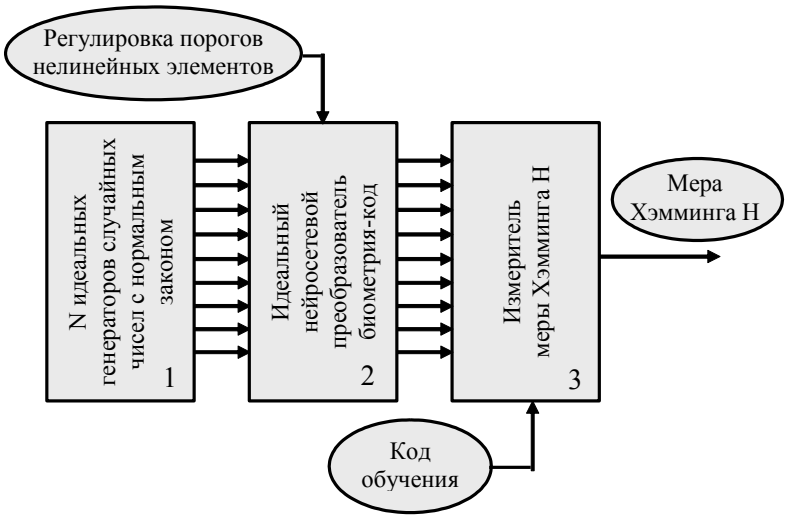


Рис. 1. Схема численного моделирования закона распределения меры Хемминга кодов-откликов «Чужой» идеального преобразователя биометрия-код

При подаче на входы правильно обученного преобразователя биометрических образов пользователя «Свой» на выходе формируется заданный при обучении код аутентификации.

Мера Хэмминга (выражающая число несовпавших бит) кодов-откликов «Свой» и заданного при обучении кода аутентификации равно нулю. При подаче на входы правильно обученного преобразователя случайных биометрических образов множества «Чужие» на выходе формируются случайные коды-отклики. При этом, регулируя параметры нелинейных элементов, можно смещать вероятности состояний на выходе преобразователя биометрия-код.

Идеальный преобразователь биометрия-код построен таким образом, что входные данные каждого нейрона не повторялись во входных данных других нейронов. На выходе преобразователя формируются N случайных кодов-откликов, вычисляется мера Хэмминга кодов-откликов «Чужой» и заданного при обучении преобразователя кода аутентификации пользователя «Свой». Формируется распределение меры Хэмминга кодов-откликов «Чужой» и заданного при обучении преобразователя кода аутентификации пользователя «Свой».

Зарубежные ученые в качестве модели статистического описания выходных состояний идеального преобразователя используют гипотезу классического биномиального закона распределения независимых данных. В отличие от используемой в зарубежных исследованиях гипотезы, численные эксперименты по схеме рис. 1, и гипотезы, предложенные в работах Надеева Д. Н. [5, 6], показали,

что выходные случайные коды идеального преобразователя биометрия-код достоверно описываются классическим биномиальным законом распределения независимых данных значений с параметрами числа степеней свободы – n и математического ожидания нормированной меры Хэмминга $-m\frac{H}{n}$ [4–7]. Гистограммы распределений меры Хемминга кодов-откликов «Чужой» идеального преобразователя при различных значениях параметра $m\frac{H}{n}$ для n = 256 приведены на рис. 2. Здесь и далее в целях наглядности при изображении дискретных распределений вместо гистограмм будут использоваться огибающие.

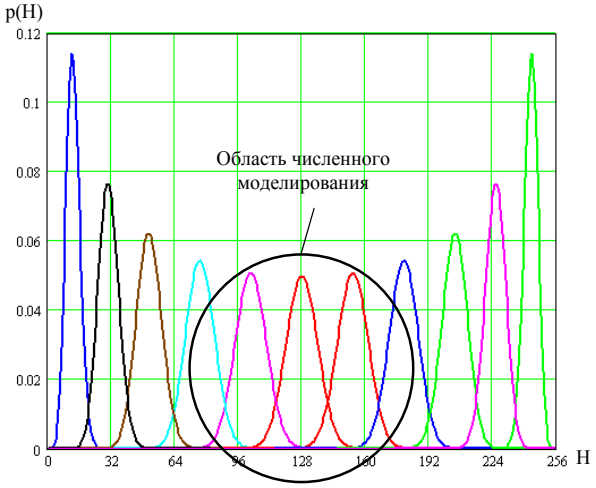


Рис. 2. Биномиальный закон распределения меры Хемминга кодов-откликов «Чужой» идеального нейросетевого преобразователя биометрия-код

Выбор обозначенной на рис. 2 области численного моделирования обусловлен предположением о правильном обучении идеального преобразователя биометрия-код, когда $m\frac{H}{n} = 0{,}5$. При дефектах обучения параметр $m\frac{H}{n}$ принимает другие значения. Экспериментальные и теоретические значения математических ожиданий $m\frac{H}{n}$ и среднеквадратических отклонений меры Хемминга $\sigma\frac{H}{n}$ для идеального преобразователя с 256 выходами приведены на рис. 3 и 4.

Установлено, что результаты численного эксперимента согласуются с теоретическими расчетами. Среднеквадратическое отклонение численных данных падает пропорционально квадратному корню от числа экспериментов.

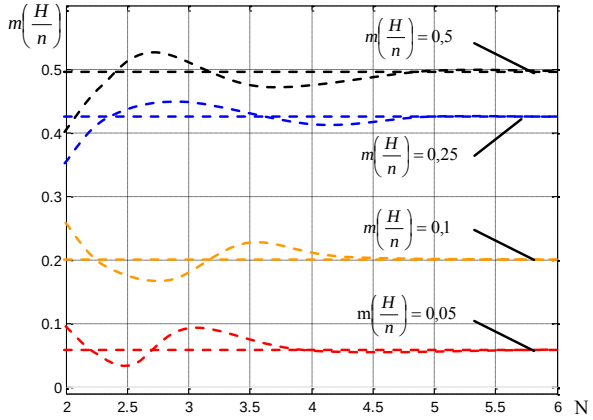


Рис. 3. Экспериментальные и теоретические значения математических ожиданий нормированной меры Хемминга $m\dfrac{H}{n}$ для идеального преобразователя с 256 выходами

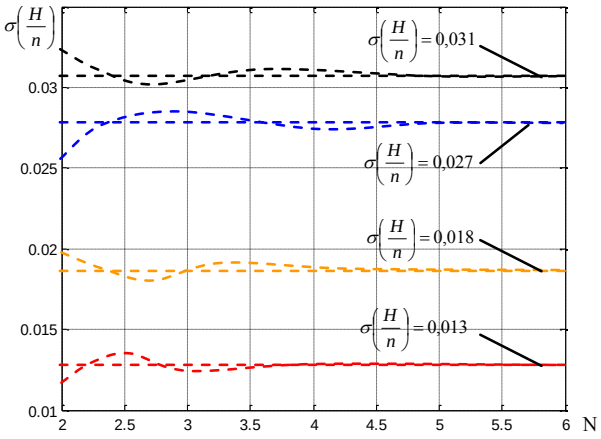


Рис. 4. Экспериментальные и теоретические значения среднеквадратических отклонений нормированной меры Хемминга $\sigma\dfrac{H}{n}$ для идеального преобразователя с 256 выходами

Наряду с этим в экспериментах явно прослеживается методическая погрешность, обусловленная не идеальностью использованных программных генераторов случайных чисел. При использовании в численном эксперименте более качественных генераторов случайных чисел методическая погрешность должна уменьшиться.

Одинаковые значения погрешностей при любых значениях выбираемого параметров $m\dfrac{H}{n}$ и n дает право утверждать, что распределения значений меры Хэмминга кодов-откликов «Чужой» идеального преобразователя достоверно описываются классическим биномиальным законом распределения независимых данных.

Для биномиального закона распределения вероятность угадывания k разрядов из общего числа n разрядов выходного кода идеального преобразователя биометрия-код равна:

$$P\left(n, m\left(\frac{H}{n}\right)\right) = \frac{n!}{k!(n-k)!}\left(m\left(\frac{H}{n}\right)\right)^{k}\left(1-m\left(\frac{H}{n}\right)\right)^{n-k}. \qquad (1)$$

Предложено вычислять вероятность ложной аутентификации «Чужого» идеальным преобразователем биометрия-код следующим образом

$$P_2 = \left(m\left(\frac{H}{n}\right)\right)^{n}. \qquad (2)$$

Проведение вычислений по выражению (2) позволяет получать оценку вероятности ложной аутентификации «Чужого» обученным преобразователем с некоррелированными выходными данными.

Таким образом можно сделать вывод о том, что распределения значений меры Хэмминга кодов-откликов «Чужой» идеального преобразователя достоверно описываются классическим биномиальным законом распределения независимых данных.

Библиографический список

1. Надеев, Д. Н. Статистическое описание идеального нейросетевого преобразователя биометрия-код: «информационный» вечный двигатель / Д. Н. Надеев, А. Ю. Малыгин, А. И. Иванов // Нейрокомпьютеры: разработка, применение. – 2007. – № 12. – С. 15–17.

2. Надеев, Д. Н. Моделирование биномиального зависимого закона распределения значений: синтез таблиц вероятностей ошибок первого и второго рода для высоконадежной биометрической защиты / Д. Н. Надеев // Современные технологии безопасности. – 2006. – № 4. – С. 33–38.

3. Нейросетевые преобразователи биометрических образов человека в код его личного криптографического ключа : монография / В. И. Волчихин, А. И. Иванов, В. А. Фунтиков, О. В. Ефимов и др. ; под ред. А. Ю. Малыгина. – Москва : Радиотехника, 2006. – Кн. 29. – 48 с. – (Сер.: Нейрокомпьютеры и их применение).

4. Надеев, Д. Н. Аналитико-фрактальное описание закона распределения значений выходных кодов биометрико-нейросетевых преобразователей / Д. Н. Надеев, А. И. Иванов // Современные технологии безопасности. – Пенза, 2007. – С. 44–49.

5. Надеев, Д. Н. Моделирование вероятностных состояний выходных бит преобразователей биометрия-код: описание закона распределения / Д. Н. Надеев // Антитеррористическая безопасность. – Пенза, 2007. – С. 54–56.

6. Надеев, Д. Н. Численно-аналитическое описание закона распределения значений выходных кодов биометрико-нейросетевых преобразователей / Д. Н. Надеев // Надежность и качество – 2007 : сб. материалов Междунар. симп. – Пенза : Изд-во ПГУ, 2007. – С. 33–38.

7. Иванов, А. И. Оценка фрактальности нейросетевых преобразователей биометрия-код при высоких входных размерностях / А. И. Иванов, Д. Н. Надеев, М. Е. Агеев, О. С. Захаров // Надежность и качество – 2007 : сб. материалов Междунар. симп. – Пенза : Изд-во ПГУ, 2007. – С. 28–33.

Для цитирования:

Вершинин, Н. Н. Численная модель идеального нейросетевого преобразователя биометрия-код / Н. Н. Вершинин, А. С. Боровский, И. В. Урнев, Г. К. Чистова // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 198–203.

УДК 519.2

А. С. Боровский, Н. Н. Вершинин, Л. А. Авдонина,
С. А. Полковникова

ИЗМЕРЕНИЕ ПОКАЗАТЕЛЯ ФРАКТАЛЬНОСТИ РЕАЛЬНОГО НЕЙРОСЕТЕВОГО ПРЕОБРАЗОВАТЕЛЯ БИОМЕТРИЯ-КОД

Аннотация. Реальный нейросетевой преобразователь биометрия-код имеет предельное значение длины выходного кода. Для кодов большей длины роста выходного качества принимаемых решений нет, фрактальные свойства преобразователя полностью исчезают, уступая место детерминированным, с точки зрения корреляции, выходным состояниям.

A. S. Borovsky, N. N. Vershinin, L. A. Avdonina,
S. A. Polkovnikova

MEASURING THE FRACTALITY OF A REAL NEURONET BIOMETRICS CODE CONVERTER

Abstract. A real biometric code converter has the ultimate value of output code length. For codes of greater growth of output quality of decisions are not, the fractal properties of the converter completely disappear, giving way to deterministic, in terms of correlation, output states.

Вероятность ложной аутентификации «Чужого» соответствует событию, когда злоумышленник угадывает все биты выходного кода преобразователя биометрия-код. Вероятность ложной аутентификации «Чужого» P_2 при вероятности появления «0» в разрядах кода $p = 0.5$ и нулевой корреляции между разрядами рассчитывается как

$$P_2 = p^n. \quad (1)$$

При больших длинах ключа n оценки вероятности ложной аутентификации «Чужого» становятся неоправданно высокими. Классический биномиальный закон дает завышенные результаты, так как не учитывает корреляционные связи между реальными биометрическими данными.

При появлении корреляции между разрядами выходного кода длина эквивалентного по стойкости выходного кода становится меньше реальной длины выходного кода преобразователя. Вероятность ложной аутентификации «Чужого» становиться равной

$$P_2 = p^{\mathrm{g}}. \qquad (2)$$

Получается, что теперь не имеет значения, для какой длины ключа оценивается вероятность ложной аутентификации «Чужого», а определяющим параметром становиться среднеквадратическое отклонения $\sigma\dfrac{H}{n}$ нормированной меры Хэмминга.

Тогда вероятность появления состояния «0» в разрядах выходного кода преобразователя эквивалентна математическому ожиданию нормированной меры Хэмминга $m\dfrac{H}{n}$. После этого формула (2) примет вид

$$P_2(g) = \left(m\left(\frac{H}{n} \right) \right)^{\mathrm{g}}. \qquad (3)$$

Так как длина эквивалентного по стойкости выходного кода преобразователя становится меньше реальной длины выходного кода преобразователя биометрия-код и принимает дробные (нецелые) значения, то оценка длины эквивалентного по стойкости выходного кода преобразователя может быть получена с использованием теории фрактальной размерности [1, 2].

В работе [3] показано, что для объектов фрактальной размерности нарушается условие, согласно которому среднеквадратическое отклонение распределения случайной величины должно уменьшаться в $\sqrt{2}$ раз при удвоении числа наблюдений.

$$б(64) \approx \frac{б(32)}{\sqrt{2}}\ б(128) \approx \frac{б(64)}{\sqrt{2}}\ \ldots\ б(2n) \approx \frac{б(n)}{\sqrt{2}}. \qquad (4)$$

Полученные результаты численных экспериментов показывают, что при увеличении в два раза длины ключа n преобразователя биометрия-код происходит уменьшение среднеквадратического отклонения нормированной меры Хэмминга $\sigma\dfrac{H}{n}$ в $\sqrt{2}$ раз. Однако условие (4) выполняется только при малых значениях длины ключа n. Например, для преобразователя, имеющего выходной

ключ 256, при увеличении n среднеквадратическое отклонение $\sigma\frac{H}{n}$ начинает снижаться медленнее. При длинах ключа, больших 500, среднеквадратическое отклонение $\sigma\frac{H}{n}$ практически перестает меняться, постепенно приближаясь к своему предельному значению $\sigma\frac{H}{n}=0.17$. Графики изменения среднеквадратического отклонения $\sigma\frac{H}{n}$ при увеличении длины ключа n приведен на рис. 1.

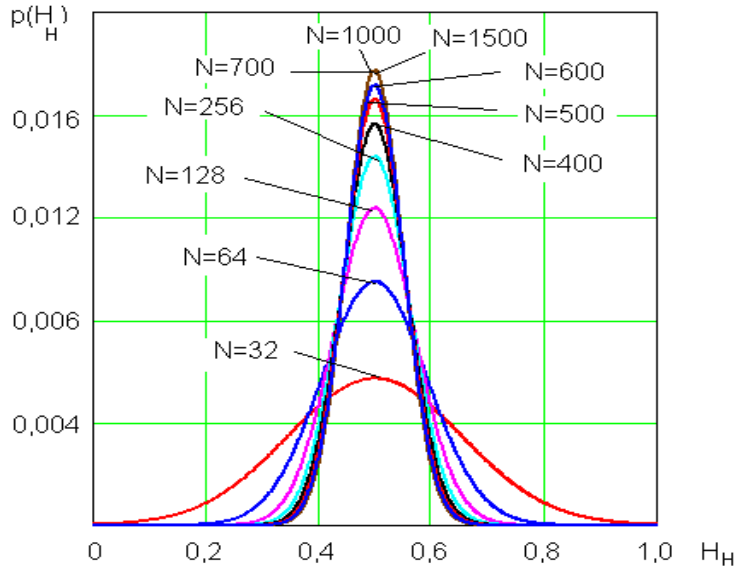


Рис. 1. Графики изменения среднеквадратического отклонения $\sigma\frac{H}{n}$ при увеличении длины ключа *n*

Отклонение от условия (4) может быть оценено по показателю самоподобия (фрактальности) Хэрста [1, 2]: Показатель Хэрста преобразователя биометрия-код может быть вычислен по формуле:

$$H(n)=\frac{\log_2(б_н(n))}{\log_2(б_н(2n))}, \qquad (5)$$

где $\sigma(n)$ – среднеквадратическое отклонение нормированной меры Хэмминга при размерности ключа длиной n.

График изменения показателя Хэрста при увеличении длины ключа обученного преобразователя биометрия-код, аппроксимированный экспонентой:

$$H(\mathrm{n})=0.5\cdot(1-e^{-0.28\cdot n})+0.5 \qquad (6)$$

приведен на рис. 2.

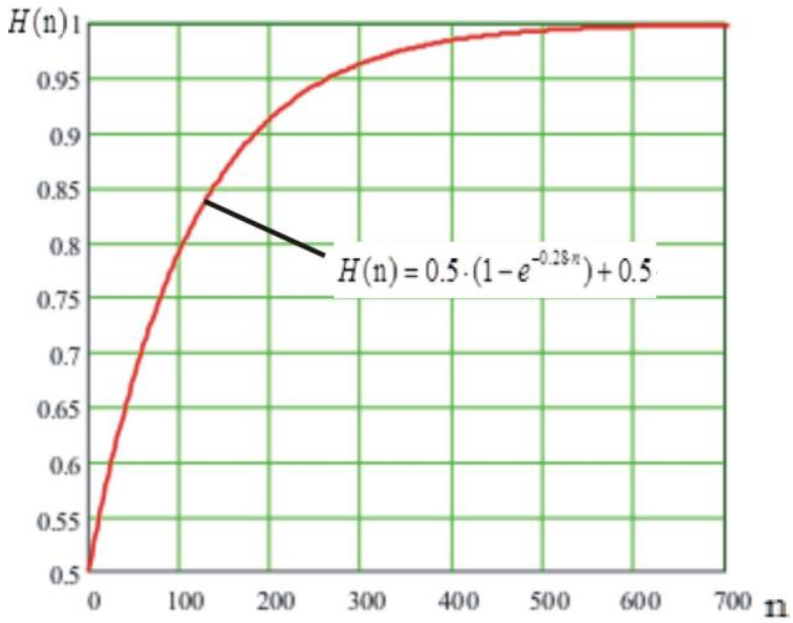


Рис. 2. Связь показателя Херста с длиной ключа *n*

Из данных рис. 2 видно, что показатель Хэрста для малых длин ключа преобразователя близок к $H(n) = 0.5$ (события случайны и слабо коррелированы). При увеличении длины ключа показатель $H(n)$ растет и при некотором значении n достигает единицы $H(n) = 1$ (наблюдения перестают быть независимыми). В таких условиях перестаем получать приращение качества принимаемых преобразователем решений.

Результаты численных экспериментов по проверке отклонения от условия независимости кодов-откликов «Чужой» преобразователя биометрия-код с ростом длины ключа n преобразователя показывают, что каждый из реализованных на практике нейросетевых преобразователей биометрия-код имеет свое предельное значение длины выходного кода. Например, предельное значение длины ключа однослойного 256-выходного преобразователя равно N = 512. Для кодов большей длины роста выходного качества принимаемых решений нет. Фрактальные свойства преобразователя полностью исчезают, уступая место детерминированным с точки зрения корреляции выходным состояниям.

Библиографический список

1. Иванов, А. И. Оценка фрактальности нейросетевых преобразователей биометрия-код при высоких входных размерностях / А. И. Иванов, Д. Н. Надеев, М. Е. Агеев, О. С. Захаров // Надежность и качество – 2007 : сб. материалов Междунар. симп. – Пенза : Изд-во ПГУ, 2007. – С. 28–33.

2. Калуш, Ю. А. Показатель Хѐрста и его скрытые свойства / Ю. А. Калуш, В. М. Логинов // Сибирский журнал индустриальной математики. – 2002. – Т. 5, вып. 4. – С. 29–37.

3. Кликушин, Ю. Н. Метод фрактальной классификации сложных сигналов / Ю. Н. Кликушин // Журнал радиоэлектроники. – 2000. – Т. 4. – URL: https://jre.cplire.ru/iso/apr00/1/text.html

Для цитирования:

Боровский, А. С. Измерение показателя фрактальности реального нейросетевого преобразователя биометрия-код / А. С. Боровский, Н. Н. Вершинин, Л. А. Авдонина, С. А. Полковникова // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 204–208.

УДК 519.24

А. Г. Банных

НЕЙРОСЕТЕВЫЕ ЭКВИВАЛЕНТЫ НОВЫХ СТАТИСТИЧЕСКИХ КРИТЕРИЕВ ДЛЯ ПРОВЕРКИ ГИПОТЕЗЫ СИММЕТРИЧНОСТИ РАСПРЕДЕЛЕНИЯ ДАННЫХ МАЛОЙ ВЫБОРКИ

Аннотация. Целью работы является оценка вероятности ошибок первого и второго рода для двух новых статистических критериев проверки гипотезы симметричности малых выборок биометрических данных. Рассмотрен критерий, построенный как сумма правого и левого «хвостов» исследуемой малой выборки. Показан дискретный характер распределений для этого типа новых статистических критериев. Кроме того, рассматривается критерий отношения максимума к минимуму данных малой выборки. Распределение данных такого статистического критерия является непрерывным. Подтверждена возможность перевода новых статистических критериев в форму эквивалентных им искусственных нейронов.

A. G. Banny

NEURAL NETWORK EQUIVALENTS OF NEW STATISTICAL CRITERIA TO TEST THE SYMMETRY HYPOTHESIS OF SMALL SAMPLE DATA DISTRIBUTION

Abstract. The aim of the work is to assess the probability of first- and second-class errors for two new statistical criteria for testing the symmetry hypothesis of small samples of biometric data. The criterion, built as the sum of the right and left tails of the small sample studied, is considered. The discrete nature of distributions for this type of new statistical criteria is shown. In addition, the criterion for the ratio of the maximum to the minimum of small sample data is considered. The distribution of data of such a statistical criterion is continuous. The possibility of translating new statistical criteria into the form of the equivalent of artificial neurons has been confirmed.

Одной из проблем биометрии является необходимость получения оценок малых выборок. По своему справочнику [1] А. И. Кобзарь выделил группу из 21 критерия, ориентированных на проверку гипотезы нормальности. Каждому из этих критериев можно построить эквивалентный ему нейрон [2, 3, 5]. Объединив классические критерии удается получить сеть из 21 искусственных нейронов. Чем больше нейронов в сети, тем выше оказывается

доверие к ее решениям. Группа из 21 статистических критериев может быть увеличена до 32, добавив к ним еще 11 критериев симметрии [1]:

1) критерий третьего статистического момента 1930 г.;
2) критерий Смирнова 1947 г.;
3) критерий Вилкоксона 1945 г.;
4) критерий Финчи 1977 г.;
5) критерий Ходжеса-Лемана 1982 г.;
6) критерий Кенуя 1979 г.г.;
7) критерий Аттила-Керсинга-Цуккини 1982 г.;
8) критерий Бхаточая-Гаствирта-Райта 1982 г.;
9) критерий Бооса 1982 г.;
10) критерий Гупты 1967 г.;
11) критерий Фрезера 1957–1963 гг.

В дополнение к уже созданным в прошлом веке статистическим критериям анализа симметрии распределений могут быть созданы новые статистические критерии. Например, хорошим дополнением является новый статистический критерий суммы правого и левого «хвостов» анализируемых данных.

Для выборки в 21 опыт новый критерий и его характеристики даны в табл. 1

Таблица 1

$\begin{cases} x \leftarrow \dfrac{sort(x) - E(x)}{\sigma(x)} \\ \tilde{x}_i \leftarrow -3 + \dfrac{2 \cdot i}{3},\ i = 0, 1, 2, 3 \\ sx \leftarrow n_0 + n_2 \\ \quad z \leftarrow "1"\ if\ \ sx < 3.5 \\ \quad z \leftarrow "0"\ if\ \ sx \geq 3.5 \\ P_1 \approx 0.25 \approx P_2 \approx 0.14 \approx P_{EE} \approx 1.95 \end{cases}$	где – n_i – число опытов выборки, попавших в i-й интервал гистограммы между его границами $\tilde{x}_i$ и $\tilde{x}_{i+1}$, $corr(sx, \mu_3) = -0.29$, $corr(sx, sS) = 0.24$, $corr(sx, \mu_4) = 0.219$, $corr(sx, og) = -0.039$; μ_3 – третий статистический момент; μ_4 – четвертый статистический момент; sS – статистика критерия Смирнова

На рис. 1 представлены спектральные линии нового критерия «суммы хвостов» для нормального распределения и его асимметричной альтернативы в виде хи-квадрат критерия с 9 степенями свободы.

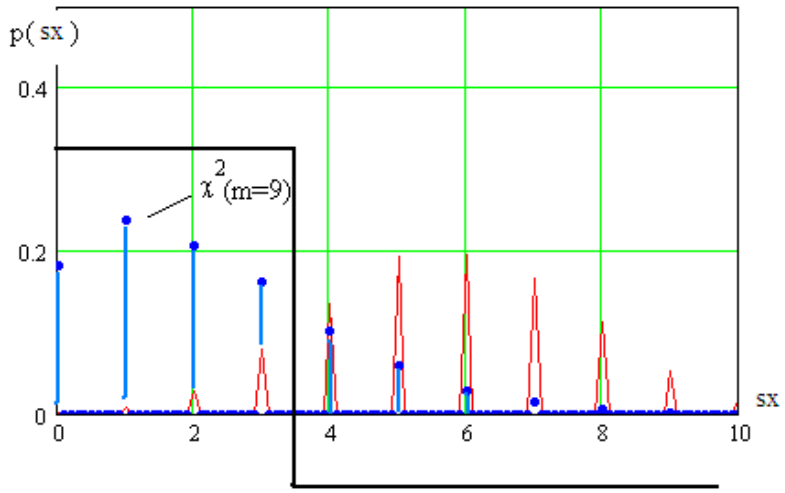


Рис. 1. Дискретный выходной спектр состояния критерия «суммы хвостов»

Принципиально важным является то, что новый статистический критерий дает решения слабо коррелированные с решениями классических статистических критериев (μ_3, μ_4, sS), как это отображено в правой части табл. 1. Классические статистические критерий имеют модули показателей коэффициентов корреляции, находящиеся в интервале от 0.7 до 0.97.

Аналогичный результат дает еще один новый статистический критерий «отношения границ» выборки, данные по этому критерию приведены в табл. 2.

Таблица 2

$x \leftarrow sort(x)$ $og \leftarrow \frac{E(x) - x_0}{x_{20} - E(x)}$ $z \leftarrow "1"\ if\ \ og < 0.75$ $z \leftarrow "0"\ if\ \ og \geq 0.75$ $P_1 = P_2 = P_{EE} \approx 0.251$	где $corr(og, \mu_3) =$ -0.343, $corr(og, \mu_4) = 0.377$, $corr(og, sS) = 0.54$, $corr(og, sx) = 0.219$

Этот новый критерий дает непрерывные спектры своих выходных состояний, приведенные на рис. 2.

Как видно из правой части табл. 2 новый статистический критерий и эквивалентные ему нейрон имеют слабо коррелированные выходные состояния как с классическими статистическими критериями, так и с новым критерием «суммы хвостов».

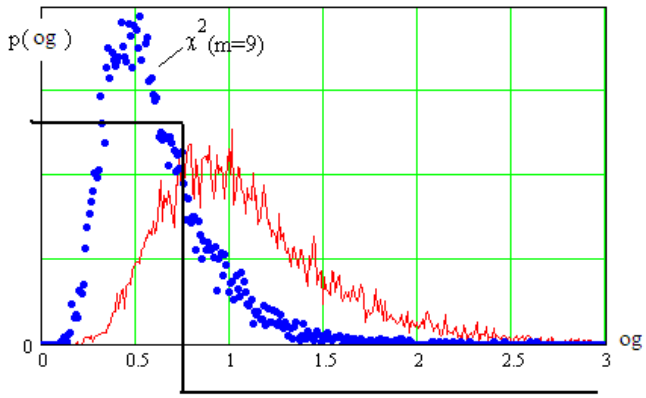


Рис. 2 Выходные состояния нового критерия «отношения границ» выборки – og

Таким образом, рассмотренные выше два статистических критерия и эквивалентные им искусственные нейроны позволяют увеличить число искусственных нейронов с 32 до 34. При этом задача синтеза статистических критериев прошлого XX в. и нынешнего XXI в. существенно меняются. В прошлом веке исследователи создавали статистические критерии, ориентируясь только на рост их мощности при тех или иных ограничениях. Синтез новых статистических критериев должен в этом веке учитывать возможность совместного их применения, то есть от них требуется низкий уровень корреляционной сцепленности с уже известными статистическими критериями.

Библиографический список

1. Кобзарь, А. И. Прикладная математическая статистика. Для инженеров и научных работников / А. И. Кобзарь. – Москва : ФИЗМАТЛИТ, 2006. – 816 с.

2. Коллекция искусственных нейронов, эквивалентных статистическим критериям для их совместного применения при проверке гипотезы нормальности малых выборок биометрических данных / А. И. Иванов, А. Г. Банных, Е. Н. Куприянов, В. С. Лукин, К. А. Перфилов, К. Н. Савинов // Безопасность информационных технологий : сб. науч. ст. по материалам I Всерос. науч.-техн. конф. (г. Пенза, 24 апреля 2019 г.). – Пенза, 2019. – С. 156–164.

3. Иванов, А. И. Искусственные молекулы, собранные из искусственных нейронов, воспроизводящих работу классических статистических критериев / А. И. Иванов, А. Г. Банных, А. В. Безяев // Вестник пермского университета. Сер.: Математика. Механика. Информатика. – 2020. – № 1 (48). – С. 26–32.

4. Иванов, А. И. Искусственный нейрон для контроля по критерию вариаций коэффициентов эксцесса малых выборок биометрических данных

с нормальным распределением / А. И. Иванов, А. Г. Банных, А. В. Безяев // Информационно-управляющие телекоммуникационные системы, средства поражения и их техническое обеспечение : сб. науч. ст. Всерос. науч.-техн. конф. / под общ. ред. В. С. Безяева. – Пенза : АО «НПП "Рубин"», 2019. – С. 84–94.

5. Иванов, А. И. Учет влияния корреляционных связей через их усреднение по модулю при нейросетевом обобщении статистических критериев для малых выборок / А. И. Иванов, А. Г. Банных, Ю. И. Серикова // Надежность. – 2020. – № 20 (2). – С. 28–34. – URL: https://doi.org/10.21683/1729-2646-2020-20-2-28-34

Для цитирования:

Банных, А. Г. Нейросетевые эквиваленты новых статистических критериев для проверки гипотезы симметричности распределения данных малой выборки / А. Г. Банных // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 209–213.

УДК 519.24

Ю. И. Серикова, К. А. Перфилов, А. Ю. Малыгин,
С. А. Полковникова

ДВУХКРИТЕРИАЛЬНЫЙ СТАТИСТИЧЕСКИЙ АНАЛИЗ МАЛЫХ БИОМЕТРИЧЕСКИХ ВЫБОРОК

Аннотация. При переходе к использованию двухкритериального статистического анализа удается получать решения с более высокой достоверностью. Двухкритериальный статистический анализ по параллельной проверке двух альтернативных статистических гипотез о нормальном и равномерном распределениях позволяет снизить вероятность ошибок пропорционально произведению вероятностей проверки каждой частной гипотезы, что снижает требования к объему тестовой выборки в несколько раз.

Y. I. Serikova, K. A. Perfilov, A. Ya. Malygin, S. A. Polkovnikova

TWO-CRITICAL STATISTICAL ANALYSIS OF SMALL BIOMETRIC SAMPLES

Abstract. When you move to the use of two-critical statistical analysis, you can get solutions with greater reliability. Two-critical statistical analysis by parallel validation of two alternative statistical hypotheses on normal and even distributions reduces the probability of errors in proportion to the probabilities of each private hypothesis, which reduces the requirements for the size of the test sample several times.

Одним из наиболее популярных при статистическом анализе данных является критерий Пирсона. Хи-квадрат критерию Пирсона полностью посвящена первая часть рекомендаций Госстандарта [1], тогда как все остальные критерии описаны во второй части рекомендаций [2]. Подробное описание критерия Пирсона в первой части рекомендаций Госстандарта [1], отражает факт высокой востребованности именно этого критерия промышленностью. Методики, построенные на использовании хи-квадрат критерия, предполагают проверку некоторой статистической гипотезы о наблюдаемом законе распределения значений $\tilde{p}(x)$. Расчеты ведутся по классической формуле:

$$\chi^2 = n \cdot \sum_{i=1}^{k} \frac{\left(\frac{b_i}{n} - \tilde{p}_i \right)^2}{\tilde{p}_i}, \qquad (1)$$

где b_i – число опытов, попавших i-й интервал гистограммы; $\tilde{p}_i$ – ожидаемая теоретическая вероятность попадания в i-й интервал гистограммы; n – число опытов в тестовой выборке; k – число столбцов гистограммы.

К сожалению, стандартные методики статистических расчетов (1) при анализе биометрических данных дают недостоверные результаты. Для того, чтобы добиться вероятностей ошибок на уровне 0.05 приходится использовать порядка 100 опытов в тестовой выборке.

Главной причиной ошибок при анализе биометрических данных является недостаточный объем данных в исследуемых тестовых выборках [3–5]. Эта ситуация характерна не только для тестирования средств биометрической защиты информации. Та же самая ситуация возникает и при обработке любых данных: медицинских, спортивных, биологических и т.д.

В настоящее время наметилась тенденция решать проблему плохих данных искусственным заполнением пробелов в пустых интервалах гистограммы, так называемым «бутстрап методом» [6], который разрушает естественные корреляционные связи в существенно зависимых биометрических данных. Примерно такого же эффекта удается добиться цифровым сглаживанием гистограмм реальных данных [7–9].

Популярность использования хи-квадрат критерия Пирсона в промышленности во многом обусловлена тем, что при $n \to \infty$ его распределение описывается через гамма функцию с $m = k - 1$ числом степеней свободы:

$$p_{\chi^2}(n=\infty, m=k-1, x) = \frac{1}{2^{\frac{m}{2}} \cdot \Gamma\left(\frac{m}{2}\right)} \cdot x^{\frac{m}{2}-1} \cdot e^{\frac{-x}{2}}. \qquad (2)$$

Аналитическое описание (2) получено Пирсоном в 1904 г. и играло крайне важную роль в первой половине XX в., когда вычислительные возможности, используемые при статистической обработке данных, были весьма ограниченными.

К сожалению, традиционное применение хи-квадрат критерия для биометрических данных дает неудовлетворительные результаты. Одной из причин является ошибка, возникающая из-за конечной тестовой выборки. Практика показывает, что при конечной тестовой выборке (например, для $n = 81$) число степеней свободы

у хи-квадрат распределения оказывается не целым (дробным) и именно из-за этого возникает значительное расхождение:

$$p_{\chi^2}(n=81, m\neq k-1, x) \neq p_{\chi^2}(n=\infty, m=k-1, x). \qquad (3)$$

Ошибку из-за конечности тестовой выборки можно учесть путем численного эксперимента. Сегодня повторить эксперимент на компьютере 1 000 000 раз вполне возможно, что дает значения функции распределения значений с приемлемой для практического применения погрешностью.

При организации численного эксперимента будем исходить из того, что должны проверяться две статистические гипотезы. Первая гипотеза состоит в том, что данные тестовой выборки имеют нормальный закон распределения значений. Вторая гипотеза состоит в том, что данные этой же выборки могут иметь нормальный закон распределения значений. Далее значения хи-квадрат критерия должны сравниваться с некоторым порогом квантователя. Если значение хи-квадрат менее порога, то принимается решение о нормальности исследуемых входных данных. Если значение хи-квадрат критерия (1) оказывается выше или ниже порога, то принимается решение о наибольшей справедливости одной из гипотез.

Проверка первой гипотезы о нормальном законе распределения значений для конечной тестовой выборки. Будем исходить из того, что по критерию хи-квадрат требуется распознать ситуацию появление данных, соответствующих серии из 81 отсчетов, полученных от нормального генератора Γ_1. Для этой цели будем вычислять математическое ожидание тестовой выборки – E(x) и ее среднеквадратическое отклонение - $\sigma(x)$. Далее построим гистограмму, состоящую из $k = 9 = \sqrt{81}$ столбцов, равномерно покрывающих интервал от минимального значения – $(E(x)-3\cdot\sigma(x))$ до максимального значения – $(E(x)+3\cdot\sigma(x))$. При этом значения критерия хи-квадрат будем вычислять следующим образом:

$$\chi^2(\Ф) = 81\cdot\sum_{i=1}^{9}\frac{\left(\frac{b_i}{81}-\frac{1}{\sigma(x)\sqrt{2\pi}}\int\limits_{x_i}^{x_{i+1}}\exp\left\{\frac{-(E(x)-u)^2}{2\cdot(\sigma(x))^2}\right\}du\right)^2}{\frac{1}{\sigma(x)\sqrt{2\pi}}\int\limits_{x_i}^{x_{i+1}}\exp\left\{\frac{-(E(x)-u)^2}{2\cdot(\sigma(x))^2}\right\}du}, \qquad (4)$$

где пределы интегрирования x_1, x_2, …, x_{10} – это границы равномерных интервалов, на которых строится гистограмма частот появления данных в тестовой выборке.

На рис. 1 приведены кривые гистограмм распределения значений хи-квадрат критерия для данных, полученных от двух программных генераторов.

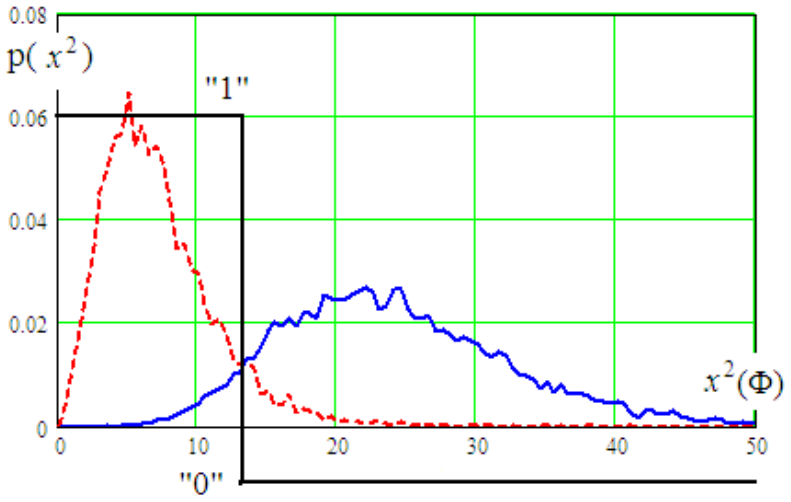


Рис. 1. Выделение данных с нормальным законом распределения значений (пунктирная линия) при проверке первой гипотезы

Из данных рис. 1 видно, что компаратор, принимающий решение об обнаружении входной нормальной последовательности должен давать состояние «1» в интервале значений от 0 до 14. Порог переключения компаратора в состояние «0» – 14. В этом случае вероятности ошибок первого и второго рода оказываются одинаковыми $P_1 = P_2 = P_{EE} = 0.054$.

Проверка второй гипотезы о равномерном законе распределения значений для конечной тестовой выборки. Будем исходить из того, что по критерию хи-квадрат требуется распознать ситуацию появление данных, соответствующих серии из 81 отсчетов, полученных от генератора данных с равномерным законом – Γ_2. Для этой цели будем находить max(x) и min(x) в тестовой выборке. Далее будем строить гистограмму, состоящую из $k = 9 = \sqrt{81}$ столбцов, равномерно покрывающих интервал от min(x) до max(x). При этом значения критерия хи-квадрат будем вычислять следующим образом:

$$\chi^2(\text{const}) = 81 \cdot \sum_{i=1}^{9} \frac{\left(\frac{b_i}{81} - \frac{1}{9}\right)^2}{\frac{1}{9}}, \qquad (5)$$

где границы интервалов гистограммы находятся следующим образом:

$$x_i = \min(x) + \frac{(\max(x) - \min(x)) \cdot i}{10}. \qquad (6)$$

На рис. 2 приведены кривые гистограмм распределения значений хи-квадрат критерия для данных, полученных от двух программных генераторов.

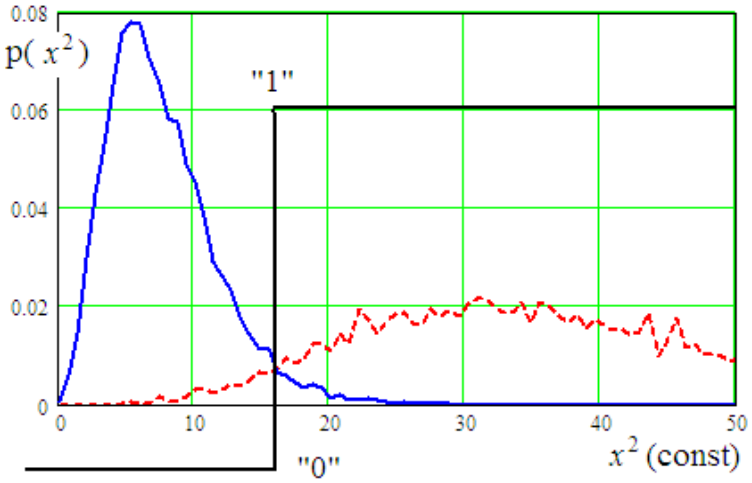


Рис. 2. Выделение данных с нормальным законом распределения значений (пунктирная линия) при проверке второй гипотезы

Из рис. 2 видно, что компаратор, принимающий решение об обнаружении входной нормальной последовательности должен давать состояние «1» в интервале значений от 17 и выше. Порог переключения компаратора в состояние «0» – 16. В этом случае вероятности ошибок первого и второго рода оказываются одинаковыми $P_1 = P_2 = P_{EE} = 0.054$.

Обобщенный критерий хи-квадрат, учитывающий параллельную проверку двух гипотез. Критерий хи-квадрат, построенный под поверку первой гипотезы (4) и хи-квадрат критерий, построенный под проверку второй гипотезы (5) – это две разных нелинейных функций преобразования, имеющих два выходных компаратора, настроенных по-разному. Так как эти критерии дополняют друг друга, обобщим их с использованием логической функции «или»:

$$\begin{cases} p(x) = \dfrac{1}{\sigma(x)\sqrt{2\pi}} \cdot \exp\left\{ \dfrac{-(E(x)-x)^2}{2 \cdot (\sigma(x))^2} \right\}, \text{ если } (\chi^2(\Phi) \le 14) \wedge (\chi^2(\text{const}) \ge 16); \\ p(x) = \text{const}, \text{ если } (\chi^2(\Phi) \ge 14) \vee (\chi^2(\text{const}) \le 16). \end{cases} \qquad (7)$$

Практика показала, что для обобщенного решающего правила (7) по сравнению с более простыми решающими правилами происходит значительное снижение вероятностей ошибок первого и второго рода $P_1 = P_2 = P_{EE} \approx 0.0025$. То есть, частные критерии хи-квадрат (4) и (5) обладают высоким уровнем согласованности принятия ими правильных решений, при этом их ошибочные решения, оказываются слабо коррелированы.

Если пользоваться хи-квадрат критерием по стандартным методикам [1], то для тестовой выборки в 81 опыт получим вероятности ошибок на уровне 0.054. Однако, как только перейти к учету ошибок, возникающих из-за конечности тестовой выборки (3) и применить обобщенный критерий хи-квадрат (7), то вероятность ошибок снижается примерно в 20 раз. Столь существенное снижение вероятностей ошибок может быть достигнуто только при размерах тестовой выборки в 800 отсчетов, это эквивалентно 10-ти кратному снижению требований к размерам тестовой выборки.

Библиографический список

1. Р 50.1.037–2002. Рекомендации по стандартизации. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Ч. I. Критерии типа χ^2. Госстандарт России. – Москва, 2002. – 140 с.

2. Р 50.1.037–2002. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Часть II. Непараметрические критерии. Госстандарт России. – Москва, 2002. – 123 с.

3. Волчихин, В. И. Нейросетевое преобразование биометрического образа человека в код его личного криптографического ключа / В. И. Волчихин, А. И. Иванов, А. Ю. Малыгин и др. – Москва : Радиотехника, 2008. – Кн. № 29. – (Сер. «Нейрокомпьютеры и их применение»). – 87 с.

4. Малыгин, А. Ю. Быстрые алгоритмы тестирования высоконадежных нейросетевых механизмов биометрической защиты информации : монография / А. Ю. Малыгин, В. И. Волчихин, А. И. Иванов, В. А. Фунтиков. – Пенза : Изд-во ПГУ, 2006. – 161 с.

5. Кухарев, Г. А. Биометрические системы: методы и средства идентификации личности человека / Г. А. Кухарев. – Санкт-Петербург : Политехника, 2001. – 240 с.

6. Руководство по биометрии / Р. М. Болл, Дж. Х. Коннел, Ш. Панканти, Н. К. Ратха, Э. У. Сеньор. – Москва : Техносфера, 2007. – 368 с.

7. Серикова, Н. И. Биометрическая статистика: «сглаживание» гистограмм, построенных на малой обучающей выборке / Н. И. Серикова, А. И. Иванов, С. В. Качалин // Вестник СибГАУ. – 2014 – № 3 (15). – С. 146–150.

8. Волчихин, В. И. Эффект снижения размера тестовой выборки за счет перехода к многомерному статистическому анализу биометрических данных /

В. И. Волчихин, А. И. Иванов, Н. И. Серикова, Ю. В. Фунтикова // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2015. − № 1 (33). − С. 50–59.

9. Кобзарь, А. И. Прикладная математическая статистика. Для инженеров и научных работников / А. И. Кобзарь. – Москва : ФИЗМАТЛИТ, 2006. – 816 с.

Для цитирования:

Серикова, Ю. И. Двухкритериальный статистический анализ малых биометрических выборок / Ю. И. Серикова, К. А. Перфилов, А. Ю. Малыгин, С. А. Полковникова // Безопасность информационных технологий : сб. науч. ст. по материалам II Всерос. науч.-техн. конф. – Пенза : Изд-во ПГУ, 2020. – С. 214–220.

СВЕДЕНИЯ ОБ АВТОРАХ

Авдонина Любовь Александровна, к.т.н., доцент кафедры техносферной безопасности, Пензенский государственный университет, г. Пенза.

Афанасьев Андрей Алексеевич, д.т.н., доцент, Академия ФСО, г. Орел.

Ахметов Бахытжан Сражатдинович, д.т.н., профессор, директор Института информационных и телекоммуникационных технологий Казахского национального технического университета имени К. И. Сатпаева, Республика Казахстан.

Ахметов Берик Бахытжанович, к.т.н., ректор Каспийского государственного университета технологий и инжиниринга им. Ш. Есенова, Республика Казахстан.

Банных Андрей Григорьевич, аспират кафедры технических средств информационной безопасности, Пензенский государственный университет, г. Пенза.

Безяев Александр Викторович, к.т.н., ведущий научный сотрудник, Пензенский филиал НТЦ «Атлас», г. Пенза.

Боровский Александр Сергеевич, д.т.н., доцент, проректор по научной работе, заведующий кафедрой управления и информатики в технических системах, Оренбургский государственный университет, г. Оренбург.

Боршевников Алексей Евгеньевич, аспирант, Дальневосточный федеральный университет, г. Владивосток.

Вершинин Николай Николаевич, д.т.н., профессор кафедры техносферной безопасности, Пензенский государственный университет, г. Пенза.

Волчихин Владимир Иванович, д.т.н., профессор, президент Пензенского государственного университета, заслуженный деятель науки РФ, г. Пенза.

Добржинский Юрий Вячеславович, профессор кафедры информационной безопасности Школы естественных наук Дальневосточного федерального университета, г. Владивосток.

Елфимов Андрей Владимирович, системный архитектор обособленного подразделения ОАО «ИнфоТеКС», г. Пенза.

Золотарева Татьяна Александровна, старший преподаватель, кафедра информатики, информационных технологий и защиты информации, Липецкий государственный педагогический университет имени П. П. Семенова-Тян-Шанского, г. Липецк.

Иванов Александр Иванович, д.т.н., доцент, научный консультант АО «ПНИЭИ», г. Пенза.

Иванов Алексей Петрович, к.т.н., доцент, заведующий кафедрой технических средств информационной безопасности, Пензенский государственный университет, г. Пенза.

Иванова Надежда Александровна, специалист Российского отделения компании "ABBYY", г. Москва.

Карпов Артем Павлович, специалист Пензенского филиала НТЦ «Атлас», г. Пенза.

Качайкин Евгений Иванович, специалист по информационной безопасности Минтруда России, г. Москва.

Качалин Сергей Викторович, к.т.н., заместитель начальника отдела, АО «НПП "Рубин"», г. Пенза.

Киселёв Владимир Юрьевич, студент кафедры информационной безопасности систем и технологий, Пензенский государственный университет, г. Пенза.

Князьков Владимир Сергеевич, д.т.н., профессор, главный научный сотрудник Научно-исследовательского института фундаментальных и прикладных исследований, Пензенский государственный университет, г. Пенза.

Крохин Игорь Алексеевич, инженер-программист АО «ПНИЭИ», г. Пенза.

Куприянов Евгений Николаевич, аспират кафедры технических средств информационной безопасности, Пензенский государственный университет, г. Пенза.

Малыгин Александр Юрьевич, д.т.н., профессор, кафедра РСС Военного учебного центра при Пензенском государственном университете, г. Пенза.

Малыгина Елена Александровна, к.т.н., докторант кафедры технических средств информационной безопасности, Пензенский государственный университет, г. Пенза.

Махсудов Суннатулла Рим Угли, студент кафедры информационной безопасности систем и технологий, Пензенский государственный университет, г. Пенза.

Перфилов Константин Александрович, аспират кафедры технических средств информационной безопасности, Пензенский государственный университет, г. Пенза.

Платонов Вадим Дмитриевич, студент кафедры информационной безопасности систем и технологий, Пензенский государственный университет, г. Пенза.

Полковникова Светлана Андреевна, аспирант кафедры вычислительной техники, Пензенский государственный университет, г. Пенза.

Постников Николай Андреевич, инженер АО «ПНИЭИ», г. Пенза.

Потапов Алексей Владимирович, научный сотрудник ФГУП «18 Центральный научно-исследовательский институт» Министерства обороны Российской Федерации, г. Москва.

Ратников Кирилл Андреевич, аспират кафедры технических средств информационной безопасности, Пензенский государственный университет, г. Пенза.

Савинов Константин Николаевич, старший преподаватель Военного учебного центра при Пензенском государственном университете, г. Пенза.

Серикова Юлия Игоревна, аспирант кафедры информационно-вычислительных систем, Пензенский государственный университет, г. Пенза.

Солопов Александр Иванович, к.т.н., научный сотрудник, ФГУП «18 Центральный научно-исследовательский институт» Министерства обороны Российской Федерации, г. Москва.

Строков Алексей Валерьевич, специалист компании «Организационно-технические решения», г. Москва.

Сулавко Алексей Евгеньевич, к.т.н., доцент кафедры комплексной защиты информации, Сибирская государственная автомобильно-дорожная академия, г. Омск.

Туреев Сергей Васильевич, к.т.н., заместитель генерального директора по научно-техническому развитию, Научно-исследовательский институт систем связи и управления, г. Москва.

Урнев Иван Васильевич, д.т.н., старший научный сотрудник Научно-исследовательского института фундаментальных и прикладных исследований, Пензенский государственный университет, г. Пенза.

Фунтиков Вячеслав Александрович, к.т.н., генеральный директор АО «ПНИЭИ», г. Пенза.

Хворостухин Сергей Павлович, к.т.н., старший научный сотрудник, АО «ПНИЭИ», г. Пенза.

Цимбал Владимир Анатольевич, заслуженный деятель науки РФ, д.т.н., профессор кафедры автоматизированных систем управления филиала Военной академии Ракетных войск стратегического назначения им. Петра Великого, г. Серпухов Московской области.

Чистова Галина Константиновна, д.т.н., профессор, профессор кафедры автономных информационных и управляющих систем, Пензенский государственный университет, г. Пенза.

Юнин Алексей Петрович, ведущий специалист АО «ПНИЭИ», г. Пенза.

СОДЕРЖАНИЕ

АО «Пензенский научно-исследовательский электротехнический институт» (АО «ПНИЭИ»)

В АО»ПНИЭИ» в настоящее время разрабатываются и серийно выпускаются комплексы и технические средства криптографической защиты информации, средства специальной связи, обеспечивающие конфиденциальность, достоверность, целостность информации при передаче ее по различным каналам связи. Активно развиваются такие направления как

– создание средств управления защищенными информационно-телекоммуникационными сетями;

– создание специальных систем передачи данных;

– создание средств электронного документооборота;

– развитие и внедрение биометрико-нейросетевых технологий.

Учеными и специалистами ПНИЭИ создаются и внедряются новые поколения аппаратуры и комплексов технических средств для обработки и защиты мультимедийной информации, передаваемой по разнородным каналам и информационно-телекоммуникационным системам связи, созданным на базе современных международных протоколов.

Акционерное общество
ПЕНЗЕНСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ ИНСТИТУТ

440000, г. Пенза, ул. Советская, 9
info@pniei.penza.ru
сайт: пниэи.рф

(841-2) 59-33-50
приемная (841-2) 59-33-35
служба маркетинга (841-2) 59-33-43

Комплекс технических средств ПОРТАЛ

КТС ПОРТАЛ предназначен для организации защищенного корпоративного (ведомственного) портала с мультимедийными сервисами, имеющего собственную логическую инфраструктуру управления, не зависящую от зарубежных ресурсов логического управления публичной сетью Интернет, и комплексно реализующего современные технологии безопасности доверенных вычислений на основе отечественной элементной базы.

В основе системных решений лежит разработка собственной облачной криптографически защищенной среды, реализующей идеологию «интернет в интернете», и предоставляющей набор как стандартных, так и узкоспециализированных веб-сервисов.

Комплекс разворачивается на базе существующих ведомственных локальных сетей и не требует установки и настройки программного обеспечения на рабочих станциях. Работа пользователей осуществляется также, как если бы они работали через Интернет, но реальный выход в глобальную сеть пользователям будет недоступен, и наоборот, доступ в ведомственную сеть со стороны открытой сети Интернет также невозможен.

КТС включает в себя серверную составляющую, аппаратные средства криптографической защиты информации, мобильное приложение для доступа с Android-устройств и программное обеспечение взаимосвязанных и объединенных между собой прикладных сервисов.

Пользователь получает доступ к защищенной электронной почте, защищенному мессенджеру мгновенных сообщений с различный устройств (смартфон, планшет, ноутбук и т.д.), организует защищенные аудио и видеоконференции между разнородными техническими средствами, ведет защищенные переговоры с помощью сервиса виртуальной АТС, получает доступ к защищенному облачному хранилищу файлов и защищенному сервису справочной информации. При этом действия пользователя мало отличаются от привычных ему действий при работе в интернете через стандартный браузер.

Акционерное общество
ПЕНЗЕНСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ ИНСТИТУТ

440000, г. Пенза, ул. Советская, 9
info@pniei.penza.ru
сайт: пниэи.рф

(841-2) 59-33-50
приемная (841-2) 59-33-35
служба маркетинга (841-2) 59-33-43

ПОРТАЛ-СЕРВЕР

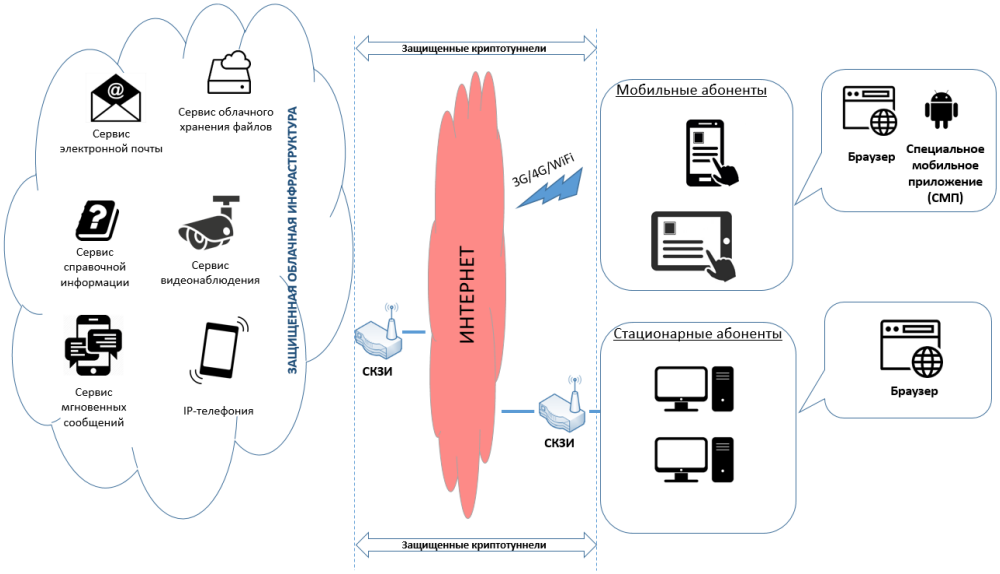
Продукт Портал-сервер представляет собой облачную криптографически замкнутую среду информационного взаимодействия и входит в линейку изделий комплекса "Портал".

Состав системных служб

- DNS-служба
- служба синхронизации времени (NTP)
- сетевая служба
- служба маршрутизации

Состав защищенных прикладных сервисов

- сервис электронной почты
- сервис обмена мгновенными сообщениями
- сервис видеоконференций
- сервис справочной информации (Вики-страницы)
- сервис облачного хранения файлов (Диск)
- видеонаблюдение
- голосовая и ведеосвязь



Логическая организация защищенной облачной системы реализована так как это делается в глобальной сети Интернет – т.е. с собственной внутренней структурой доменных имен для доступа к ресурсам. По аналогии с сетью Интернет, пользователи могут использовать облачные сервисы через привычные для них браузеры без каких-либо дополнительных условностей.

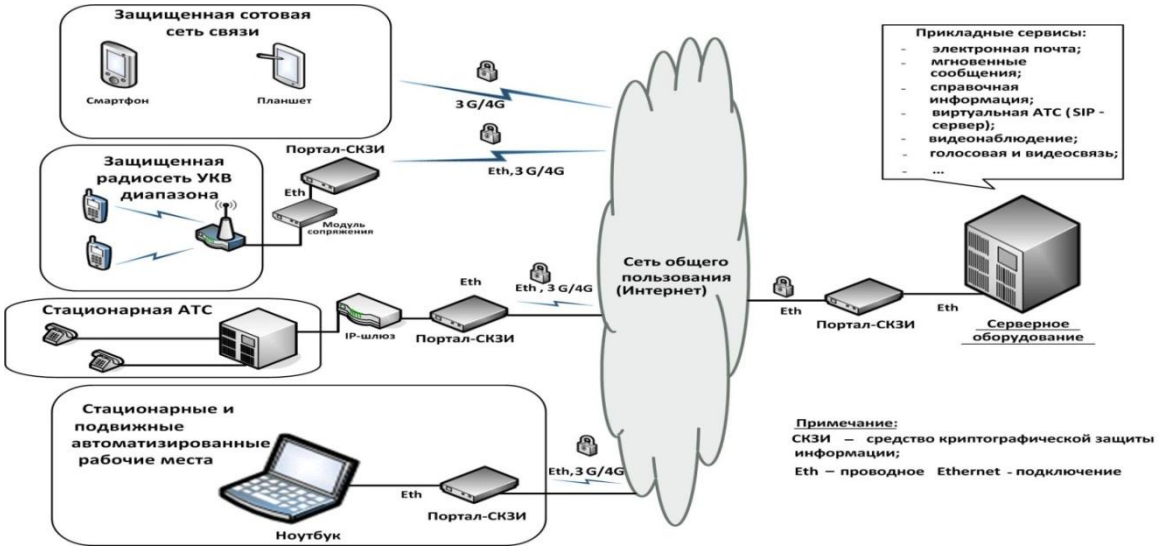
Акционерное общество
ПЕНЗЕНСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ ИНСТИТУТ

440000, г. Пенза, ул. Советская, 9
info@pniei.penza.ru
сайт: пниэи.рф

(841-2) 59-33-50
приемная (841-2) 59-33-35
служба маркетинга (841-2) 59-33-43

Доступ к информационным ресурсам может осуществляться с любого клиентского устройства сети, находящегося за СКЗИ. Разграничение доступа осуществляется с помощью механизмов аутентификации.

Также возможна организация следующей схемы связи разнородных информационных систем и отдельных технических средств в единой криптографически защищенной сервис-ориентированной среде:



Доступна организация системным администратором с помощью виртуальной АТС аудиоконференции между разнородными техническими средствами (смартфон, планшет, радиостанция, стационарный телефон, стационарное и подвижное рабочее место).

Средства криптографической защиты (Портал-СКЗИ)

Криптографическую защиту передаваемых данных в этих средах возможно осуществлять на различных уровнях стека телекоммуникационных протоколов:

- на канальном уровне (в проработке)
- на прикладном уровне (Портал-ПО, Портал-1-SD)
- на сетевом уровне (Портал-10, М-687, Швейцар-М, Портал-1000)

Акционерное общество
ПЕНЗЕНСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ ИНСТИТУТ

440000, г. Пенза, ул. Советская, 9
info@pniei.penza.ru
сайт: пниэи.рф

(841-2) 59-33-50
приемная (841-2) 59-33-35
служба маркетинга (841-2) 59-33-43

Таблица 1 – Возможные типы СКЗИ для использования в предлагаемой архитектуре

Наимено-вание	**Класс защиты**	**Скорость**	**Габариты, мм**	**Примечание**
Портал-ПО	длина ключа до 56 бит	Ограничена производи-тельностью устройства	–	Реализуется на прикладном уровне
Портал-1-SIM	длина ключа до 56 бит, для класса КС1 – в разработке	1 Мбит/с	форм-фактор SIM-карты	Совместная работа с программным СКЗИ
Портал-1-SD	длина ключа до 56 бит, для класса КС1 – в разработке		форм-фактор SD-карты, microSD-карты	Совместная работа с Портал-10, Портал-1000
Портал-10	длина ключа до 56 бит, для класса КС1 – в разработке	10 Мбит/с	165×115×25	Поддержка Wi-Fi, работа в динамических IP- адресах Совместная работа с Портал-1-SD, Портал-1000
Портал-1000	длина ключа до 56 бит, для класса КС1, КА – в разработке	600 Мбит/с	440×380×58	Совместная работа с Портал-1-SD, Портал-10, М-687А, Швейцар-М
Швейцар-М	КА, КВ	35 Мбит/с	230×165×55	Совместная работа с М-687, Портал-1000, Швейцар-М
М-687 (М-687А, М-687В)	КА, КВ, гостайна	95 Мбит/с	392x316x53	Совместная работа с Швейцар-М, Портал-1000

Акционерное общество
ПЕНЗЕНСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ ИНСТИТУТ
440000, г. Пенза, ул. Советская, 9
info@pniei.penza.ru
сайт: пниэи.рф
(841-2) 59-33-50
приемная (841-2) 59-33-35
служба маркетинга (841-2) 59-33-43

Изделие ПОРТАЛ-1-SD

Изделие является составной частью комплекса ПОРТАЛ и обеспечивает шифрование информации с длиной ключа 56 бит.

Обеспечивает

- **обмен с изделиями КТС *ПОРТАЛ***
- чтение ключевой, служебной и пользовательской информации из встроенной FLASH-памяти
- криптографическую обработку информации в соответствии с заданным алгоритмом специального преобразования

Основные технические характеристики

- максимальная тактовая частота – 20 МГц
- ряд напряжений питания – 1,8 В ± 10%; 3,0 В ± 10%; 5,0 В ± 10 %
- интерфейс ввода-вывода информации – последовательный в соответствии со стандартом SD и микроSD (SPI)
- объем встроенной FLASH-памяти – 16 Гбайт
- количество циклов стирания/записи FLASH-памяти – не менее 100 000
- время сохранности информации во FLASH-памяти – не менее 10 лет
- встроенный аппаратный ускоритель операций для криптографических алгоритмов DES
- диапазон рабочих температур: от минус 25 °С до плюс 85 °С.

Производится на базе отечественного микроконтроллера «Курган» с доверенным загрузчиком нулевого уровня.

Акционерное общество
ПЕНЗЕНСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ ИНСТИТУТ

440000, г. Пенза, ул. Советская, 9
info@pniei.penza.ru
сайт: пниэи.рф

(841-2) 59-33-50
приемная (841-2) 59-33-35
служба маркетинга (841-2) 59-33-43

Изделие ПОРТАЛ-1-SIM

Изделие является составной частью комплекса ПОРТАЛ и обеспечивает шифрование информации с длиной ключа 56 бит по интерфейсу стандарта ISO/IEC 7816-3.

Обеспечивает

- обмен с изделиями КТС *ПОРТАЛ*
- запись, хранение ключевой, служебной и пользовательской информации во встроенной FLASH-памяти
- чтение ключевой, служебной и пользовательской информации из встроенной FLASH-памяти
- криптографическую обработку информации в соответствии с заданным алгоритмом специального преобразования

Основные технические характеристики

- максимальная тактовая частота – 20 МГц
- ряд напряжений питания – 1,8 В ± 10 %; 3,0 В ± 10 %; 5,0 В ± 10 %
- интерфейс ввода-вывода информации – последовательный в соответствии со стандартом ISO/IEC 7816-3
- протокол информационно-логического взаимодействия – оригинальный на основе протокола T0 стандарта ISO/IEC 7816-3
- объем встроенной FLASH-памяти – 384 Кбайт
- количество циклов стирания/записи FLASH-памяти не менее 100 000
- время сохранности информации во FLASH-памяти не менее 10 лет
- встроенный аппаратный ускоритель операций для криптографических алгоритмов – ГОСТ 28147–89 и DES
- встроенный сопроцессор модульной арифметики
- форм-фактор – SIM
- диапазон рабочих температур: от минус 25 до плюс 85 °С

Акционерное общество
ПЕНЗЕНСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ ИНСТИТУТ

440000, г. Пенза, ул. Советская, 9
info@pniei.penza.ru
сайт: пниэи.рф

(841-2) 59-33-50
приемная (841-2) 59-33-35
служба маркетинга (841-2) 59-33-43

Изделие ПОРТАЛ-10

Изделие является составной частью комплекса ПОРТАЛ и обеспечивает шифрование и имитозащиту конфиденциальной информации с длиной ключа 56 бит

Обеспечивает

- встречную работу с аналогичным изделием ПОРТАЛ-10, а так же с изделием ПОРТАЛ-1000
- криптографическую защиту IP-пакетов методом полной инкапсуляции
- прозрачное автоматическое шифрование/расшифрование информации с заданной стойкостью по алгоритму шифрования – ГОСТ 28147-89
- контроль целостности пакетов данных – имитозащиту по ГОСТ 28147–89
- аутентификацию источника данных
- ключевую систему – полносвязную ключевую матрицу с индивидуальными ключами на каждом направлении обмена, ключевая структура предусматривает работу с 5000 абонентами
- гибкую полнофункциональную настройку изделия (с ПЭВМ)
- возможность встречной работы через NATP преобразователи (через маршрутизаторы, межсетевые экраны) в сетях с «серой IP адресацией»
- возможность встречной работы через сети операторов сотовой связи 3G (HSDPA, HSUPA,UMTS), EDGE, GPRS, LTE при наличии сервера маршрутизации мобильного трафика;
- возможность подключения USB-модема непосредственно к изделию
- возможность встречной работы по каналам Ethernet (100BASE-T), Wi-Fi
- возможность работы в режиме сервера маршрутизации мобильного трафика
- контроль технического состояния готовности к работе
- контроль наличия действующих и очередных ключей
- контроль целостности программного обеспечения
- контроль меток точного времени
- функцию дистанционного конфигурирования (реконфигурирования)
- дистанционное управление ключами (ввод ключевой информации, переход с действующего ключа на очередной, полное и выборочное стирание ключевой информации)
- круглосуточную необслуживаемую работу

Электропитание изделия ПОРТАЛ-10 осуществляется от:

- сети переменного тока напряжением 220 В (+ 22 В; – 33 В) с частотой 50 Гц
- сети постоянного тока напряжением 5 В (стандартный USB интерфейс)

Акционерное общество
ПЕНЗЕНСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ ИНСТИТУТ

440000, г. Пенза, ул. Советская, 9
info@pniei.penza.ru
сайт: пниэи.рф

(841-2) 59-33-50
приемная (841-2) 59-33-35
служба маркетинга (841-2) 59-33-43

Изделие подключается к локальной сети, а также к оборудованию транспортной сети по интерфейсу Ethernet (100BASE-T на скорости 100 Мбит/с), Wi-Fi или к сети операторов сотовой связи 3G (HSDPA, HSUPA,UMTS), EDGE, GPRS, LTE через модемное оборудование и поддерживает протокол Ethernet 802.3 на портах, не внося ограничений в работу протоколов верхних уровней.

Габаритные размеры изделия: 165x110x30 мм; масса 0,7 кг.

Акционерное общество
ПЕНЗЕНСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ ИНСТИТУТ

440000, г. Пенза, ул. Советская, 9
info@pniei.penza.ru
сайт: пниэи.рф

(841-2) 59-33-50
приемная (841-2) 59-33-35
служба маркетинга (841-2) 59-33-43

Комплекс «Швейцар»

В АО «ПНИЭИ» создан комплекс технических средств, позволяющий решить задачу по защите информации, передаваемой по протоколам IP. Комплекс предназначен для организации защищенной связи и обмена информацией между сегментами телекоммуникационных систем ведомств, а также для решения задач автоматизации управления безопасностью, включая функции дистанционного управления средствами криптографической защиты.

На базе комплекса предусматривается построение подсистем имеющих в составе до 5000 объектов (объект – локальная сеть или отдельный пользователь): архитектура комплекса позволяет строить подсистему криптографической защиты с единым центром управления безопасностью либо с иерархической структурой управления и контроля, содержащей до 200 подсетей.

Разработка велась с учетом потребности средств защиты как в сегменте защиты информации, содержащей сведения, составляющие государственную тайну со степенью секретности «секретно», так и защиты конфиденциальной информации:

– изделия М-687 (гос. тайна), М-687А и М-687В (конфиденциальный контур) с пропускной способность до 100 Мбит/с со стыками Ethernet, обеспечивающие шифрование и имитозащиту информации и режим ввода ключевой информации по каналам связи и удаленного мониторинга, обеспечивают взаимодействие с изделием Швейцар-М (при защите конфиденциальной информации);

– изделие Швейцар-М (конфиденциальный контур) с пропускной способность до 40 Мбит/с, обеспечивающее шифрование и имитозащиту информации и режим ввода ключевой информации по каналам связи и удаленного мониторинга, обеспечивает взаимодействие с изделиями М-684А и М-687В;

– аппаратура М-684 (гос. тайна), М-684А и М-687В (конфиденциальный контур)- станции децентрализованного изготовления ключей и их распределения по каналам связи, с функциями удаленного мониторинга состояния технических средств комплекса, автоматизированного сбора и учета сведений о событиях безопасности в подсистеме криптографической защиты. Обеспечивают возможность организации многоуровневой иерархической подсистемы управления безопасностью в сетях IP, реализованных на базе изделий М-687 (М-687А, М-687В) и Швейцар-М.

Акционерное общество
ПЕНЗЕНСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ ИНСТИТУТ

440000, г. Пенза, ул. Советская, 9
info@pniei.penza.ru
сайт: пниэи.рф

(841-2) 59-33-50
приемная (841-2) 59-33-35
служба маркетинга (841-2) 59-33-43

– Все изделия комплекса удобны в эксплуатации, не требуют длительной специальной подготовки персонала, обеспечивают круглосуточную необслуживаемую работу, имеют относительно низкую стоимость по сравнению с аналогами.

– Комплекс является самостоятельной разработкой в полном объеме схемных решений и программного обеспечения, в нем отсутствует системное программное обеспечение сторонних разработчиков и не предъявляются требования к смежной аппаратуре.

Продукт прошел сертификацию на соответствие требованиям ФСБ по защите информации, содержащей сведения, составляющие государственную тайну, и на соответствие по защите информации, не содержащей сведений, составляющих государственную тайну.

Акционерное общество
ПЕНЗЕНСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ ИНСТИТУТ

440000, г. Пенза, ул. Советская, 9
info@pniei.penza.ru
сайт: пниэи.рф

(841-2) 59-33-50
приемная (841-2) 59-33-35
служба маркетинга (841-2) 59-33-43

Изделие М-687 (М-687А, М-687В)

ИЗДЕЛИЕ КРИПТОГРАФИЧЕСКОЙ ЗАЩИТЫ ДАННЫХ В IP-СЕТЯХ

Изделие обеспечивает работу

- М-687 в режиме шифрования и имитозащиты информации, содержащей сведения, составляющие государственную тайну с грифом не выше «секретно» (встречная работа с аналогичным изделием и изделием М-641)
- М-687А в режиме шифрования и имитозащиты конфиденциальной информации, класс КА (встречная работа с аналогичным изделием и изделиями М-641К
- М-687В в режиме шифрования и имитозащиты конфиденциальной информации, класс КВ (изделие работает встречно с аппаратурой Швейцар-Я)

Примечание – изделия изготавливаются по единой документации, различие – ключевые документы, вводимые на объектах эксплуатации.

Изделие имеет два исполнения

- аппаратура М-687 (М-687А, М-687В) РИВУ.465644.01 – с пультом управления ПБ090 РИВУ.468381.010
- аппаратура М-687 (М-687А, М-687В) РИВУ.465644.031-01 – без пульта управления ПБ090 РИВУ.468381.010

Изделие обеспечивает

- криптографическую защиту IP-пакетов методом полной инкапсуляции
- прозрачное автоматическое шифрование/расшифрование информации с заданной стойкостью по алгоритму шифрования – ГОСТ 28147–89
- контроль целостности пакетов данных – имитозащиту по ГОСТ 28147–89
- аутентификацию источника данных
- поддержку фрагментации пакетов
- возможность генерации «ложного трафика» и выравнивание размеров передаваемых пакетов (нормализацию трафика). Режим аналогичный режиму работы изделия «Сито»
- ключевую систему – полносвязную ключевую матрицу с индивидуальными ключами на каждом направлении обмена, ключевая структура предусматривает работу с 5000 абонентов
- гибкую полнофункциональную настройку изделия (с ПЭВМ)
- межсетевое экранирование информационных потоков с выполнением следующих требований:

Акционерное общество
ПЕНЗЕНСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ ИНСТИТУТ

440000, г. Пенза, ул. Советская, 9
info@pniei.penza.ru
сайт: пниэи.рф

(841-2) 59-33-50
приемная (841-2) 59-33-35
служба маркетинга (841-2) 59-33-43

- возможность задания правил фильтрации IP-пакетов для обоих направлений передачи (LAN-WAN, WAN-LAN), с не менее чем 100 правил для каждого направления передачи
- возможность протоколирования событий межсетевого экранирования
- поддержку классификации трафика на основе IP-адресов, номеров протоколов, номеров портов транспортных протоколов TCP и UDP, полей ToS или DiffServ и поддерживает маркировку и перемаркировку трафика по полям ToS или DiffServ в соответствии с заданными правилами.
- возможность назначения IP-адреса «вручную» (статическая адресация) и динамически по протоколу DHCP. Аппаратура с динамически назначенным IP-адресом WAN обеспечивает возможность встречной работы только с аппаратурой с «вручную» назначенным IP-адресом WAN
- возможность дистанционного мониторинга и управления ключевой информациият аппаратуры децентрализованного изготовления ключей (М-684):
  - контроль технического состояния готовности к работе
  - контроль наличия действующих и очередных ключей
  - контроль целостности программного обеспечения
  - контроль меток точного времени
  - функцию дистанционного конфигурирования (реконфигурирования)
  - дистанционное управления ключами (ввод ключевой информации, переход с действующего ключа на очередные, полное и выборочное стирание ключевой информации)
- защиту от НСД при вскрытии корпуса
- круглосуточную необслуживаемую работу
- пропускную способность 94 Мбит/с при длине передаваемых пакетов 1400 байт

М-687 имеет оригинальный, разработанный специалистами АО «ПНИЭИ» конструктив, выполняющий функции экранирования и теплоотвода с возможностью установки в 19” стойку (высота-1U).

Электропитание изделия осуществляется от сети переменного тока напряжением 220 В (+ 22 В; – 33 В) с частотой 50 Гц ± 2,5 Гц. Мощность, потребляемая изделием от сети переменного тока, не превышает 15 В·А.

Изделие подключается к локальной сети (или отдельной станции), а также к оборудованию транспортной сети по интерфейсам Ethernet (10BASE-T, 100BASE-TX, RJ-45 на скоростях 10 и 100 Мбит/с) и поддерживает протокол Ethernet 802.3 на портах, не внося ограничений в работу протоколов верхних уровней.

Габаритные размеры изделия: 392x316x52,5 мм, масса-5,3 кг.

Акционерное общество
ПЕНЗЕНСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ ИНСТИТУТ

440000, г. Пенза, ул. Советская, 9
info@pniei.penza.ru
сайт: пниэи.рф

(841-2) 59-33-50
приемная (841-2) 59-33-35
служба маркетинга (841-2) 59-33-43

Швейцар-М

ИЗДЕЛИЕ КРИПТОГРАФИЧЕСКОЙ ЗАЩИТЫ ИНФОРМАЦИИ В СЕТЯХ ПЕРЕДАЧИ ДАННЫХ

Предназначено для обеспечения безопасности конфиденциальной информации в IP-сетях стандарта IEEE 802.3/802.3u

Обеспечивает

- встречную работу с изделиями Швейцар-Я, Швейцар-М, М-687А, М-687В
- криптографическую аутентификацию изделий встречной работы
- криптографическую защиту IP-пакетов методом полной инкапсуляции
- прозрачное шифрование информации в режиме гаммирования с обратной связью по ГОСТ 28147-89
- контроль целостности пакетов данных – имитозащиту по ГОСТ 28147–89
- создание не менее 50 криптографически защищенных туннелей
- создание 10 новых криптографически защищенных туннелей в секунду
- защиту от кодирования открытой информации (выравнивание трафика, генерация ложного трафика, маркировка поля ToS)
- межсетевое экранирование сетевого трафика на основе пакетной фильтрации
- наличие механизма QoS на сетевом уровне
- наличие ключевой системы – полносвязной ключевой матрицы с индивидуальными ключами на каждом направлении обмена
- возможность встречной работы с 5000 изделий в сети
- ввод ключевой информации с использованием пульта ПБ090
- взаимодействие со станцией генерации и распределения ключей
- функциональную настройку с использованием пульта ПБ090, USB-flash накопителей и ПЭВМ, а также со стороны станции генерации и распределения ключей
- мониторинг работы изделия на ПЭВМ, подключаемой к управляющему порту изделия
- регистрация событий безопасности
- ведение статистики межсетевого экранирования
- контроль целостности программного обеспечения
- защиту от НСД при вскрытии корпуса
- круглосуточную необслуживаемую работу

По условиям эксплуатации изделие удовлетворяет требованиям групп 1.1, 1.3. Диапазон рабочих температур: от – 10 до +50 °С.

Изделие имеет специально разработанный малогабаритный экранированный, теплоотводящий корпус.

Габаритные размеры изделия: 230×165×30 мм.

Вес: ~1кг.

Акционерное общество
ПЕНЗЕНСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ ИНСТИТУТ

440000, г. Пенза, ул. Советская, 9
info@pniei.penza.ru
сайт: пниэи.рф

(841-2) 59-33-50
приемная (841-2) 59-33-35
служба маркетинга (841-2) 59-33-43

М-684

АППАРАТУРА ДЕЦЕНТРАЛИЗОВАННОГО ИЗГОТОВЛЕНИЯ, РАСПРЕДЕЛЕНИЯ КЛЮЧЕЙ И ОРГАНИЗАЦИИ МНОГОУРОВНЕВОЙ СИСТЕМЫ ДИСТАНЦИОННОГО МОНИТОРИНГА В IP-СЕТЯХ

Предназначена для децентрализованного изготовления и распределения шифрключей по каналам связи и организации многоуровневой системы дистанционного мониторинга в сетях передачи данных IP.

Обеспечивает

- децентрализованное изготовление ключевых документов
- распределение и доведение ключей до изделий Швейцар-М, М-687 (М-687А, М-687В) по каналам связи согласно заданной схемы распределения, а также до М-684, находящейся на нижележащих уровнях управления
- дистанционное управление ключами в изделиях Швейцар-М, М-687 (М-687А, М-687В) и М-684, находящихся на нижележащих уровнях управления, по каналам связи, включая управление сменой, стиранием (сбросом) ключей, контроль их наличия и состояния
- запись ключевой и служебной информации, в том числе и контроль правильности осуществленной записи на носители ДК-6 в целях доставки ключевой информации и ее непосредственного ввода в изделия Швейцар-М, М-687 (М-687А, М-687В), Швейцар-Я согласно заданной схемы распределения
- запись больших массивов ключевой и служебной информации, в том числе и контроль правильности осуществления записи на ВНИ в целях доставки до М-684, находящейся на нижележащих уровнях управления, при отсутствии канала связи
- поэкземплярный учет ключей, сроков их действия, стирания, а также отображение сведений о наличии действующих и очередных ключей в обслуживаемых изделиях Швейцар-М, М-687 (М-687А, М-687В), Швейцар-Я, а также в М-684, находящихся на нижележащих уровнях управления
- своевременную доставку очередных ключей для обслуживаемых изделий Швейцар-М, М-687 (М-687А, М-687В), а также для М-684, находящихся на нижележащих уровнях управления, с использованием каналов связи

Акционерное общество
ПЕНЗЕНСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ ИНСТИТУТ

440000, г. Пенза, ул. Советская, 9
info@pniei.penza.ru
сайт: пниэи.рф

(841-2) 59-33-50
приемная (841-2) 59-33-35
служба маркетинга (841-2) 59-33-43

- выполнение режима удаленного конфигурирования (реконфигурирования) и мониторинга изделий Швейцар-М, М-687 (М-687А, М-687В) в зашифрованном и имитозащищенном виде по каналам связи и отображение его результатов на мониторе:
  - контроль технического состояния готовности к работе
  - контроль наличия действующих и очередных ключей
  - контроль целостности программного обеспечения
  - контроль меток точного времени
  - функцию дистанционного конфигурирования (реконфигурирования)
  - дистанционное управления ключами (ввод ключевой информации, переход с действующего ключа на очередные, полное и выборочное стирание ключевой информации)
- выполнение следующих функций по управлению безопасностью:
  - ведение баз данных, обеспечивающих ввод, хранение и редактирование сведений о криптографической связанности абонентов, а также служебной информации об абонентах
  - ввод сведений о произошедших компрометациях, рассылка и доведение команд восстановления связи в целях исключения скомпрометированных абонентов из сети связи
- круглосуточную работу

Габаритные размеры: 370 х 313 х 70 мм

Акционерное общество
ПЕНЗЕНСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ ИНСТИТУТ

440000, г. Пенза, ул. Советская, 9
info@pniei.penza.ru
сайт: пниэи.рф

(841-2) 59-33-50
приемная (841-2) 59-33-35
служба маркетинга (841-2) 59-33-43

Изделие БиоЗамок

БиоЗамок-БВ

БиоЗамок-К

Изделие БиоЗамок предназначено для управления электромеханическим замком или защелкой входной двери с помощью смартфона. Электронное управление замком удобно, тем что позволяет отказаться от использования связки ключей. Это экономит время так как смартфон всегда под рукой.

Модификация БиоЗамок-БВ позволяет дополнительно проверить пользователя посредством биометрической аутентификации. В качестве биометрических характеристик могут использоваться изображение лица или отпечаток пальца. Изделие БиоЗамок поддерживает считывание бесконтактных карт и брелоков RFID, обеспечивает удаленный доступ к встроенной видеокамере.

Управление и конфигурирование изделия БиоЗамок может осуществляться через Web-интерфейс, как на ПК, так и с помощью смартфона.

Монтаж изделия БиоЗамок может производиться:

– на внешнюю сторону двери (БиоЗамок-БВ);

– в полость внутри каркаса двери (БиоЗамок-К);

– в стену (БиоЗамок-БВ).

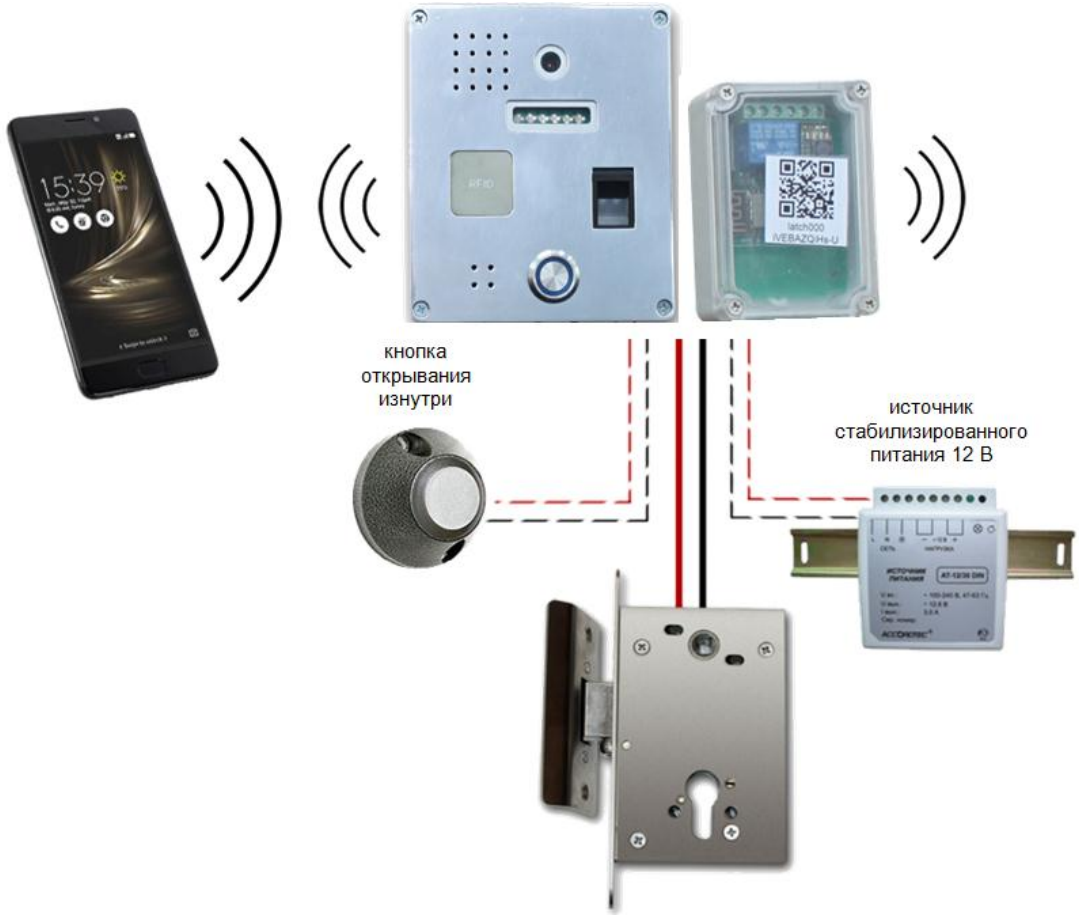
Акционерное общество
ПЕНЗЕНСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ ИНСТИТУТ

440000, г. Пенза, ул. Советская, 9
info@pniei.penza.ru
сайт: пниэи.рф

(841-2) 59-33-50
приемная (841-2) 59-33-35
служба маркетинга (841-2) 59-33-43

Технические характеристики изделия БиоЗамок

Напряжение питания	+12 В
Мощность потребления	6 Вт
Температура эксплуатации	от 0 до +50 °С
Допустимая влажность воздуха	не более 80 %
Вес (нетто)	0,5 кг
Габаритные размеры (ШхВхГ)	130x150x40 мм
Формат бесконтактных карт и брелоков	EM4100
Интерфейс беспроводной сети Wi-Fi	IEEE 802.11 b/g/n
Поддержка браузеров	Firefox, Chrome, Internet Explorer
Разрешение изображения	не менее 320 x 480 пикселей
Вероятность ошибочного предоставления доступа	менее 0,001 %
Вероятность ошибочного отказа в доступе	менее 1%



Акционерное общество
ПЕНЗЕНСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ ИНСТИТУТ

440000, г. Пенза, ул. Советская, 9
info@pniei.penza.ru
сайт: пниэи.рф

(841-2) 59-33-50
приемная (841-2) 59-33-35
служба маркетинга (841-2) 59-33-43

БиоТокен

ПРОГРАММНО-АППАРАТНОЕ СРЕДСТВО ДЛЯ ПРОВЕРКИ И ФОРМИРОВАНИЯ ЭЛЕКТРОННОЙ ПОДПИСИ С БИОМЕТРИЧЕСКИМ ПОДТВЕРЖДЕНИЕМ ЛИЧНОСТИ

Область применения

– системы обезличивания персональных данных медицинских учреждений (в составе изделия БиоГарант)

– системы электронного документооборота, торговли и услуг

и

– системы контроля и управления доступом с аппаратно-программным модулем доверенных вычислений, как отдельный фактор идентификации или для связывания биометрии с паролем доступа

– серверы децентрализованной идентификации пользователей

Функциональные возможности

– получение биометрических данных с графического планшета или сканера отпечатков пальцев

– создание и/или загрузка пары ключей формирования ЭП

– генерация псевдослучайных чисел с использованием естественной нестабильности биометрических образов

– формирование ЭП под электронными документами после биометрической авторизации пользователя

– связывание биометрии с личным ключом в процессе настройки БиоТокен с учетом требований пакета стандартов ГОСТ Р 52633 без выхода введенной биометрии и ключа пользователя из доверенной вычислительной среды БиоТокен

– обучение преобразователя биометрия-код (ГОСТ Р 52633.0–2006) на числе примеров биометрических образов «Свой» от 8 до 32 «Свой»

– хранение параметров связывания в защищенном биометрическом контейнере (ГОСТ Р 52633.4–2011)

– подтверждение критических операций загрузки данных в БиоТокен авторизованным пользователем

Электропитание устройства осуществляется от USB порта ПЭВМ. Устройство потребляет не более 150 мА.

Средство выполнено в форм-факторе USB, по условиям эксплуатации удовлетворяет требованиям климатического исполнения УХЛ4.2 ГОСТ 15150 с ограничением предельной пониженной температуры окружающей среды до минус 10°С.

Средний срок службы – не менее 5 лет.

Средняя наработка на отказ – не менее 10 000 ч.

Габаритные размеры – 72х40х17 мм.

Масса – не более 50 г.

Акционерное общество
ПЕНЗЕНСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ ИНСТИТУТ

440000, г. Пенза, ул. Советская, 9
info@pniei.penza.ru
сайт: пниэи.рф

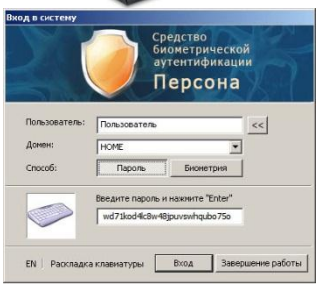
(841-2) 59-33-50
приемная (841-2) 59-33-35
служба маркетинга (841-2) 59-33-43

Персона

ПРОГРАММНОЕ СРЕДСТВО БИОМЕТРИЧЕСКОЙ ИДЕНТИФИКАЦИИ И АУТЕНТИФИКАЦИИ ПОЛЬЗОВАТЕЛЕЙ

Обеспечивает выполнение функций:

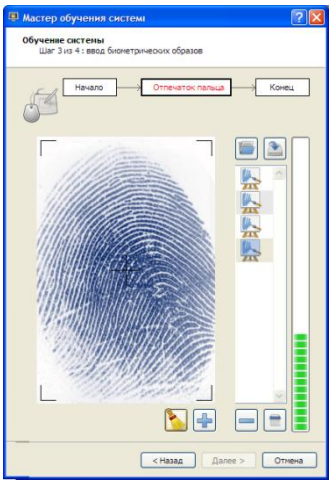
- защищѐнную биометрическую идентификацию субъектов доступа, проводимую путѐм создания нейросетевых биометрических контейнеров (НБК)
- простую защищѐнную и строгую биометрическую аутентификацию субъектов доступа с использованием НБК
- доступ к авторизованному запуску операционной системы Windows XP и контроль доступа к ее ресурсам с помощью средств криптографической защиты информации (СКЗИ)
- защиту файлов данных и контейнеров СКЗИ произвольного размера с помощью биометрических образов



Программное средство осуществляет преобразование легкозапоминаемого рукописного слова-пароля или отпечатка пальца в произвольный длинный пароль или ключ до 256 бит. Таким образом, пользователь избавлен от необходимости хранить надлежащим образом ключ или запоминать длинный случайный пароль. При подключении дополнительных модулей возможно связывание пароля (ключа) с голосовой фразой и другими биометрическими технологиями.

В программном средстве используются алгоритмы быстрого автоматического обучения искусственных нейронных сетей, параметры которых хранятся в нейросетевых биометрических контейнерах.

Преимуществом НБК является то, что сам ключ в них не хранится, не хранятся также биометрические образы пользователя.



Программное средство биометрической идентификации и аутентификации пользователей устанавливается на персональный компьютер с операционной системой семейства Windows, выполнено в соответствии с требованиями пакета стандартов ГОСТ Р 52633 и Федерального закона «О персональных данных».

Акционерное общество
ПЕНЗЕНСКИЙ НАУЧНО-ИССЛЕДОВАТЕЛЬСКИЙ ЭЛЕКТРОТЕХНИЧЕСКИЙ ИНСТИТУТ

440000, г. Пенза, ул. Советская, 9
info@pniei.penza.ru
сайт: пниэи.рф

(841-2) 59-33-50
приемная (841-2) 59-33-35
служба маркетинга (841-2) 59-33-43

Научное издание

БЕЗОПАСНОСТЬ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ

Сборник научных статей по материалам
II Всероссийской научно-технической конференции
(г. Пенза, 3 июня 2020 г.)

Статьи печатаются в авторской редакции.

Компьютерная верстка *Р. Б. Бердниковой*
Дизайн обложки *А. А. Стаценко*

Подписано в печать 25.11.2020. Формат $60\times84^{1}/_{16}$.
Усл. печ. л. 14,41.
Заказ № 412. Тираж 100.

Издательство ПГУ
440026, Пенза, Красная, 40
Тел./факс: (8412) 666-049, 666-777; e-mail: iic@pnzgu.ru