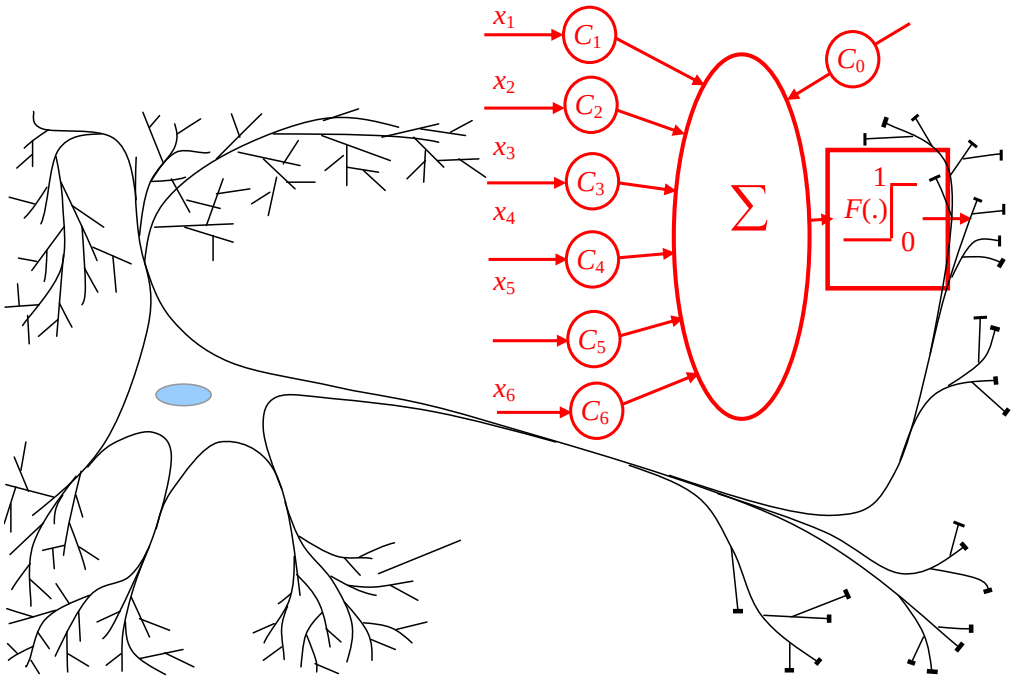
МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«ПЕНЗЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

А. И. Иванов

СРЕДА МОДЕЛИРОВАНИЯ «БИОНЕЙРОАВТОГРАФ»

Учебно-методическое пособие



ПЕНЗА 2020

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Пензенский государственный университет» (ПГУ)

А. И. Иванов

СРЕДА МОДЕЛИРОВАНИЯ «БИОНЕЙРОАВТОГРАФ»

Учебно-методическое пособие

Пенза
Издательство ПГУ
2020

УДК 681.32 2
И20

Р е ц е н з е н т

доктор технических наук, профессор,
ученый секретарь
Научно-производственного предприятия «Рубин», г. Пенза,
М. М. Бутаев

Иванов, А. И.

И20 Среда моделирования «БиоНейроАвтограф» : учеб.-метод. пособие / А. И. Иванов. – Пенза : Изд-во ПГУ, 2020. – 60 с.

Представлены материалы для теоретической подготовки и практического выполнения лабораторных работ в среде моделирования «БиоНейроАвтограф» по дисциплинам «Нейросетевые технологии в защите информации», «Биометрия и защита информации». Рассмотрены вопросы быстрого автоматического обучения искусственных нейронных сетей по требованиям ГОСТ Р 52633.5–2011 и их тестирования по ГОСТ Р 52633.3–2011.

Издание подготовлено на кафедре «Технические средства информационной безопасности» ПГУ и предназначено для обучающихся по специальностям 10.05.02 «Информационная безопасность телекоммуникационных систем», 10.05.03 «Информационная безопасность автоматизированных систем», для работы на практических занятиях и внеаудиторной самостоятельной подготовки, а также может быть полезно аспирантам и докторантам, изучающим высокоразмерные искусственные нейронные сети в рамках научных исследований по применению нейронных сетей для защиты информации.

УДК 681.32 2

© Пензенский государственный
университет, 2020

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

В настоящее время активно идут процессы информатизации современного общества. Практически все государства декларируют свое стремление создавать электронное правительство для оказания электронных услуг своим гражданам. Интернет-портал электронного правительства страны, области, района, города, поселка будут отличаться от обычных веб-сайтов только тем, что к ним должно быть обеспечено высокое доверие. Гражданин должен быть уверен в том, что он обращается именно к своему электронному правительству, а электронное правительство должно быть уверено в том, что оно отсылает свои ответы не роботу, а именно тому гражданину, который ранее обратился. Эта задача не является тривиальной, простой «капчей» здесь не обойтись.

В обычном мире представитель правительства проверяет полномочия гражданина, пользуясь его паспортом. В виртуальном мире все оказывается сложнее, там пользоваться сканом своего паспорта нельзя, даже если это биометрический паспорт с радиочитаемой микросхемой. Персональные данные в микросхеме о биометрии человека нельзя применять в Интернет-пространстве. Для Интернета и других открытых информационных пространств необходимо создавать специальные Интернет-паспорта или иные Интернет-удостоверения личности, которые, с одной стороны, являются биометрическими, а с другой стороны, являются некоторыми криптографическими конструкциями, защищающими конфиденциальность персональных данных своих хозяев.

Одной из важнейших особенностей всех электронных документов является то, что они могут иметь неограниченное число равноценных копий. Такое понятие, как «оригинал» электронного документа, является анахронизмом. Любая копия электронного документа – это его оригинал, который может быть проверен по содержащейся в нем цифровой подписи, т.е. в ближайшем будущем мы с вами столкнемся с огромным числом электронных Интернет-паспортов и удостоверений личности, созданных для авторизованного взаимодействия гражданина с государственными и негосударственными средствами предоставления ему значимых Интернет-услуг.

Скорее всего, роль паспортов и удостоверений личности будут играть личные электронные кабинеты, созданные взаимными усилиями гражданина и организациями, предоставляющими ему значимые Интернет-услуги. Каждый из нас имеет банковскую карту и счет в банке. С 2014 г. все банки России авторизованно оповещают своих клиентов о снятых с их счета платежах, сделать это безопасно можно только через личный электронный кабинет. Оперативно оповестить о факте списания

(поступления) денег можно SMS-сообщением или электронным письмом, но по этим открытым каналам связи нельзя указывать суммы списаний (поступлений) и адреса получателей (источников) денежных средств.

То же самое относится к «Федеральной налоговой службе» и «Федеральной Службе судебных приставов». Эти государственные структуры обязаны своевременно оповещать граждан, попавших в их поле зрения, о всех значимых для них правовых действиях в ближайшее время и уже существующих претензиях. Сделать это проще всего через личные кабинеты каждого налогоплательщика или должника.

Все сказанное выше относится к любому значимому источнику информации, будь то городское или поселковое электронное правительство, избирательная комиссия, поликлиника, отделение ГИБДД, военкоматы, университеты или обычные школы. Все источники значимой для личности, оперативно меняющейся информации вынуждены будут создавать виртуальные личные электронные кабинеты гражданина и отвечать за их информационную безопасность. Так же, как сегодня мы получаем кассовый чек при каждой покупке, завтра мы будем извещаться о всех действиях, касающихся каждого из нас, проведенных банком, медицинским учреждением, налоговым органом или ГИБДД. Возникает некоторая специфическая информационная система множества личных электронных кабинетов, отображенная на рис. В.1.

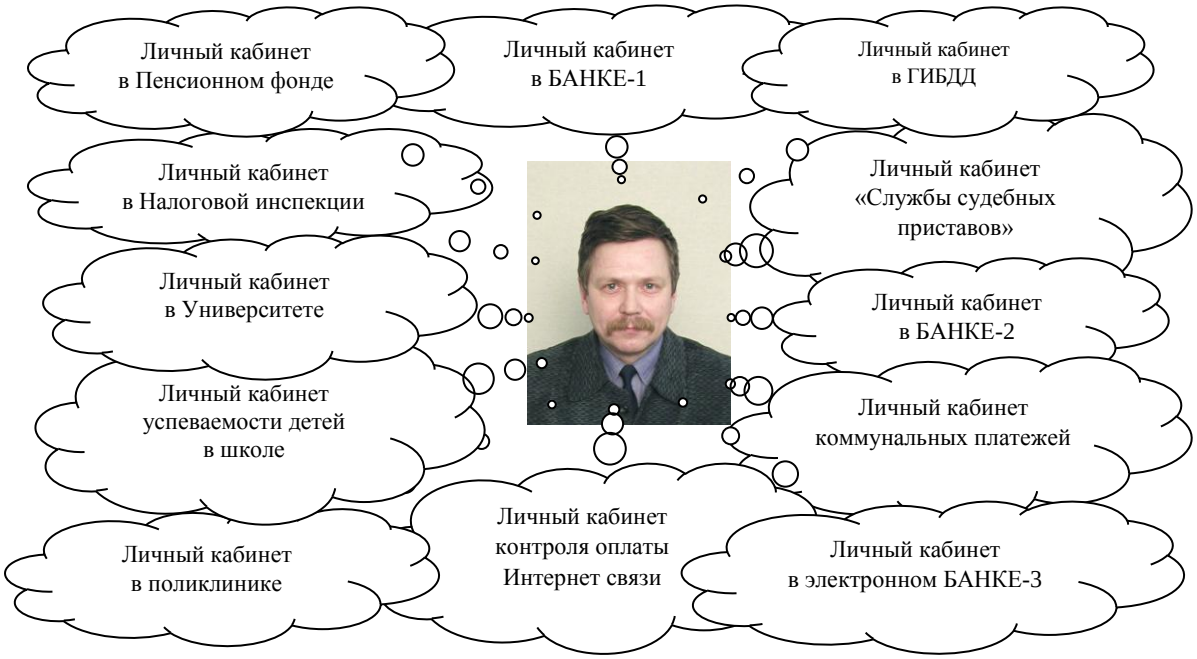


Рис. В.1. Множество правительственных и частных личных электронных кабинетов, поддерживающих оперативную авторизованную Интернет-связь с личностью потребителя значимых виртуальных услуг

Основной проблемой системы электронных кабинетов является их слабая защита: как правило, они защищены короткими легко запоминаемыми паролями. Такие пароли легко подбираются злоумышленниками. Короткие пароли удобны для запоминания, однако они порождают только иллюзию защищенности, не являясь на самом деле серьезным препятствием для хакеров.

1. ОБЩИЕ ПОЛОЖЕНИЯ БИОМЕТРИЧЕСКОЙ ИНТЕРНЕТ-АУТЕНТИФИКАЦИИ

1.1. Необходимость в биометрической авторизации значимых Интернет-действий

Безопасность личной информации электронных кабинетов может быть обеспечена только совместными усилиями владельца кабинета и владельца сервера, на котором этот кабинет размещен. Для того, чтобы обезопасить свою информацию от посягательств Интернет-злоумышленников, владелец личного кабинета должен использовать длинный пароль доступа, состоящий из 32 случайных знаков. Запоминать длинный пароль из 32 случайных знаков необязательно, можно его записать в хранителе паролей, который хранится на локальной машине и закрыт коротким паролем. Можно пойти по другому пути, дополнительно закрыв доступ к хранителю паролей биометрией. При этом программный хранитель длинных паролей может ориентироваться на уже имеющиеся в вашем компьютере средства ввода биометрических данных (чувствительный экран компьютера, встроенную видеокамеру, манипулятор «мышь», встроенный микрофон).

Для нас крайне важно осознать возможность организации биометрической защиты, личной информации, используя даже манипулятор «мышь». Это можно сделать, выполнив соответствующую лабораторную работу (п. 4.2 данного издания). Даже один рукописный символ «а» удается связать с любым случайным паролем из 32 символов. Если Вы пишете своей рукой нужный графический символ, Вы получаете свой пароль. Если Вы не знаете, какой графический образ следует воспроизвести, воспроизвести правильный пароль не получится. Эта ситуация отображена на рис. 1.1. В верхней части этого рисунка отображена ситуация ввода манипулятором «мышь» верного мини-пароля из одного графического символа. В нижней части рисунка отображена ситуация ввода случайного графического символа, приводящая к ошибке в 80 битах из 256 бит.

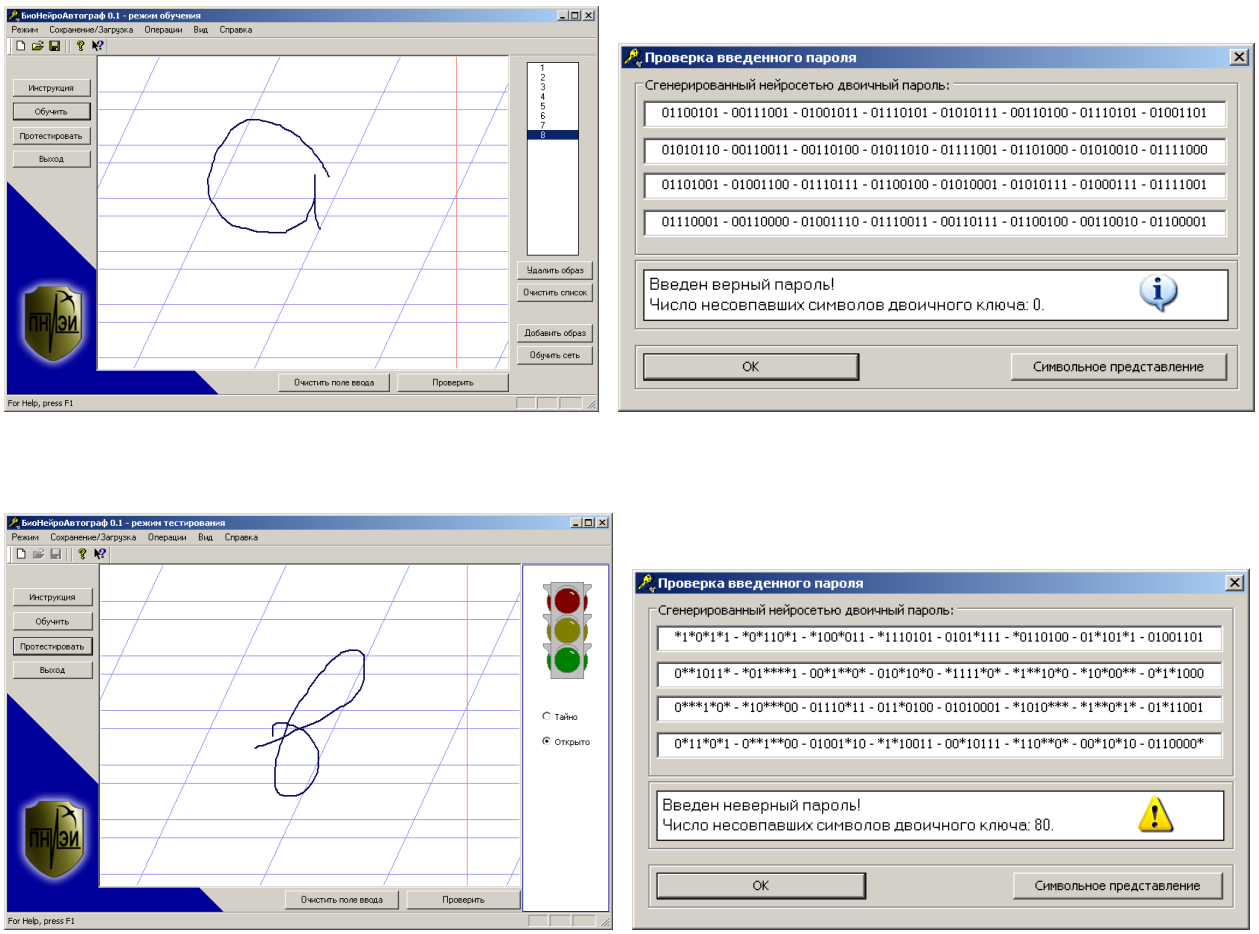


Рис. 1.1. Пример нейросетевого связывания графики простейшего рукописного образ *«а»* со случайным паролем длиной 256 бит (32 случайных знака)

Получается, что простейшей биометрической защиты длинных паролей доступа вполне достаточно для защиты своих личных кабинетов от Интернет-хакеров. Хакеры атакуют пароли доступа, ожидая, что они короткие. Если хакеру известно, что пароль длинный или защита данных осуществлена на длинном криптографическом ключе, атаки на них прекращаются.

Стойкость биометрической защиты не имеет значения, если Вы доверяете окружающим Вас людям. Только если Вы не до конца доверяете людям, имеющим доступ к Вашему компьютеру, должна использоваться достаточно сильная биометрическая защита, использующая специальные средства ввода биометрического образа «Свой». Могут использоваться сканеры рисунка отпечатка пальца, сканеры кровеносных сосудов руки, сканер радужной оболочки глаза, специальный графический планшет, сканер геометрии трехмерной маски лица человека.

1.2. Корпоративные биометрические удостоверяющие центры и центры биометрической авторизации шаговой доступности

Банк или медицинское учреждение должны быть уверены в том, что свой личный кабинет на их сайтах посещает только его подлинный хозяин. Та же самая уверенность должна быть и у электронного правительства в лице чиновников Министерства здравоохранения и социального развития. Формула по вычислению будущей пенсии усложняется, однако у человека должна оставаться возможность зайти на сайт своей региональной пенсионной службы и на калькуляторе этого сайта вычислить размер своей будущей пенсии. Технологически гарантированная авторизация доступа к информации личного электронного кабинета осуществима, если при заключении договора с банком его клиент выполнил биометрическую регистрацию своего образа в удостоверяющем центре банка.

Не следует путать корпоративный биометрический удостоверяющий центр банка с обычным удостоверяющим центром. У них совершенно разные функции. Обычный удостоверяющий центр компенсирует свои будущие расходы заранее, взимая авансом платеж за свои услуги по поддержанию сертификата открытого ключа в течение трех следующих лет (сегодня такой платеж составляет порядка 3000 руб.). При этом, будет ли пользоваться своим сертификатом открытого ключа человек (будет ли он пользоваться своей электронной подписью), удостоверяющий центр не волнует (он уже получил свою прибыль). Как показал опыт, подобная бизнес-модель малоэффективна: желающих зарегистрировать сертификат своего открытого ключа мало. Классические удостоверяющие центры имеют мало пользователей, и именно по этой причине их услуги оказываются дороги, что опять-таки приводит к снижению востребованности этих услуг.

У корпоративного биометрического удостоверяющего центра банка совершенно иные бизнес-модель и цель существования. Он должен помочь связать личный ключ формирования цифровой подписи с биометрией его владельца, тогда банк будет иметь уверенность в том, что подписал цифровой подписью платежное поручение именно владелец счета. При этом публиковать сертификат открытого ключа своего клиента банку нет необходимости, банку достаточно иметь этот сертификат и надежно хранить его в именованной или обезличенной форме. У корпоративного биометрического центра банка нет цели получить прибыль, его цель – обеспечить биометрико-криптографическую защиту банковских платежных операций. Бизнес-модель банковского биометрического

удостоверяющего центра, видимо, будет строиться на микроплатежах по каждой из поддерживаемых им операций биометрико-криптографической защиты платежей и доступа к личному кабинету клиента банка. При этом модель поддержания высокого уровня безопасности должна сохраниться: биометрический удостоверяющий центр банка не должен знать личный ключ своего клиента и связанный с ним его биометрический образ.

Как будут развиваться биометрические удостоверяющие центры, сегодня сказать трудно, однако это единственный путь гарантировать банкам, медицинским учреждениям, органам государственной власти то, что информация из личного электронного кабинета гражданина будет надежно защищена. Вполне возможно, что появятся биометрические удостоверяющие центры шаговой доступности, оказывающие услуги платной регистрации связки биометрического образа человека с его одноразовой парой ключей (открытого и личного). Будущее покажет, как будут развиваться события, однако уже сейчас понятно, что все значимые для человека Интернет-операции должны иметь надежную биометрико-криптографическую поддержку и некоторую систему центров доверия, осуществляющих безопасное связывание биометрии и криптографии.

1.3. Система требований к средствам биометрической авторизации

На сегодняшний день каждая из развитых стран старается пользоваться собственной криптографией и собственной безопасной связкой биометрии и криптографии. Очевидными лидерами этих процессов являются США и Россия. В США в период с 1992 по 2013 г. создано порядка 100 национальных стандартов, регламентирующих требования к биометрическим технологиям. Часть из этих стандартов переведена в ранг международных через международный комитет по стандартизации ISO/IEC JTC SC37 (Biometric), который создан в 2002 г. Стандартизацию безопасной связи биометрии и криптографии осуществляет ISO/IEC JTC SC27 (Security techniques).

В России активно развивается технология нейросетевого связывания биометрии с кодом ключа доступа [1]. В период с 2005 по 2012 г. под нейросетевую технологию создано восемь стандартов (номера и названия этих стандартов даны в табл. 1.1).

Таблица 1.1

Номера и названия стандартов

№	Номер и полное название стандарта
1	**ГОСТ Р 52633.0–2006** «Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации»
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»
3	**ГОСТ Р 52633.2–2010** «Защита информации. Техника защиты информации. Требования к формированию синтетических биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации».
4	**ГОСТ Р 52633.3–2011** «Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора»
5	**ГОСТ Р 52633.4–2012** «Защита информации. Техника защиты информации. Интерфейсы взаимодействия с нейросетевыми преобразователями биометрия-код»
6	**ГОСТ Р 52633.5–2011** «Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа»
7	**ГОСТ Р 52633.6–2012** «Защита информации. Техника защиты информации. Требования к индикации близости предъявленных биометрических данных образу "Свой"»
8	**ГОСТ Р 52633.7–2014** «Защита информации. Техника защиты информации. Высоконадежная мультибиометрическая аутентификация»

США и страны Евросоюза идут по пути связывания биометрических данных человека с его криптографическим ключом через применение так называемых «нечетких экстракторов» [2–7]. По своей сути «нечеткие экстракторы» строятся путем бинарного квантования «сырых» биометрических данных и последующей корректировки ошибок квантования классическими избыточными кодами с обнаружением ошибок. «Нечетким экстракторам» полностью посвящен второй раздел данного учебно-методического пособия.

Применительно к использованию «нечетких экстракторов» на сегодняшний день созданы только стандарты самого верхнего уровня. По уровню стандартизации эти технологии существенно уступают отечественным технологиям биометрико-нейросетевой аутентификации. Названия и номера соответствующих международных стандартов приведены в табл. 1.2.

Таблица 1.2

Международные стандарты

№	Номер и полное название стандарта
1	**ISO/IEC 24745:2011** Information technology – Security techniques – Biometric information protection
2	**ISO/IEC 24761:2009** Information technology – Security techniques – Authentication context for biometrics
3	**ISO/IEC 19792:2009** Information technology. Security techniques. Security evaluation of biometrics

Роль стандартизации биометрических технологий трудно переоценить. Соотношения взаимных усилий США и России в этом направлении стандартизации отображены на рис. 1.2.

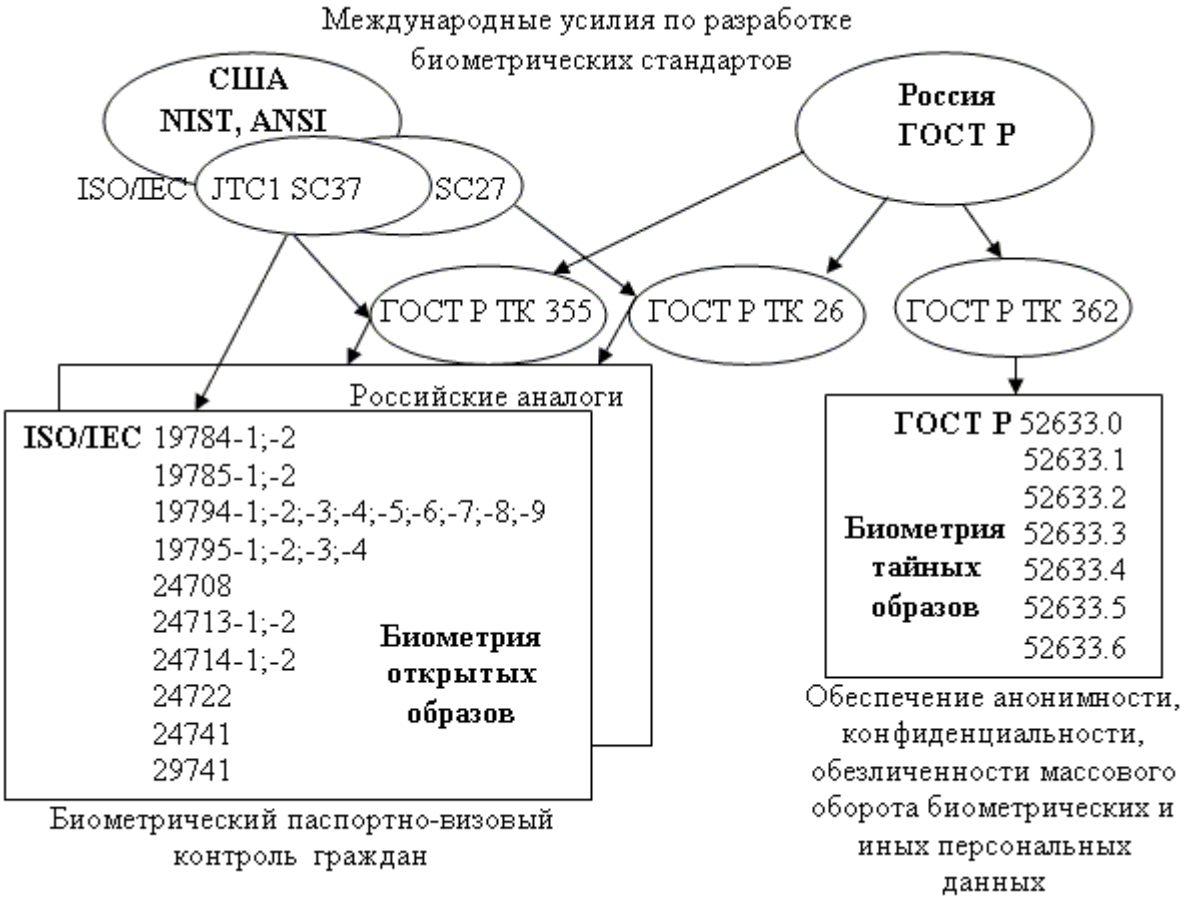


Рис. 1.2. Совместные усилия США и России
по созданию биометрических стандартов

Из данного рисунка видно, что США на сегодняшний день являются страной, лидирующей по биометрическим технологиям паспортно-визового контроля, в том числе международный Российский паспорт

с биометрической микросхемой создан, опираясь на международные биометрические стандарты, которые в 90-х гг. прошлого века разрабатывались как национальные стандарты США. Однако биометрию паспортно-визового контроля граждан нельзя использовать при Интернет-авторизации личности.

Россия занимает лидирующее положение по стандартизации биометрических приложений Интернет-авторизации. В ближайшее время национальные стандарты России, видимо, будут взяты за основу при создании международных стандартов. Третий раздел данного пособия целиком посвящен отечественным технологиям нейросетевого преобразования биометрии в код ключа доступа.

2. ПРОСТЕЙШИЕ ПРЕОБРАЗОВАТЕЛИ С ПРЯМЫМ КВАНТОВАНИЕМ БИОМЕТРИЧЕСКИХ ДАННЫХ

2.1. Бинарные монотонные квантователи «сырых» биометрических параметров

Биометрические технологии строятся на сканировании биометрических образов человека и извлечении из них контролируемых параметров. Основной задачей биометрической аутентификации является предоставить доступ носителю образа «Свой» к его информационному ресурсу, преобразовав биометрические параметры – ν_i в разряды кода ключа доступа – «c_i». Второй задачей биометрической аутентификации является преобразование случайных параметров – ξ_i случайного образа «Чужой» в случайные разряды случайного кода – «x_i». Преобразование непрерывных биометрических параметров в дискретные бинарные состояния в «нечетких экстракторах» осуществляется пороговыми функциями бинарного квантования:

$$\begin{cases} L_i \ \ \xi_i \ \ = "x_i" \\ L_i \ \ \nu_i \ \ = "c_i". \end{cases} \qquad (2.1)$$

Простейшие бинарные квантователи (2.1) выполняют сравнение анализируемых данных со значением единственного порога – θ_i. Квантователь может давать скачок из состояния «0» в состояние «1» при переходе возрастающей переменной через порог – θ_i. Возможно использование инверсных квантователей, которые при переходе возрастающей переменной через порог – θ_i дают инверсное изменение состояний из «1» в «0». Настройка каждого квантователя осуществляется выбором его типа и выбором порога квантования. Целью настройки квантователей является как можно большее снижение энтропии разрядов кода «Свой» при как можно большем увеличении энтропии состояний разрядов кода «Чужой»:

$$\mathrm{var}(\theta_i) \Rightarrow \begin{cases} \max(H("x_i")) \le 1{,}0 \ \text{бит} \\ \min(H("c_i")) \ge 0{,}0 \ \text{бит}. \end{cases} \qquad (2.2)$$

Идеальная настройка порога квантователя должна приводить к нулевому значению энтропии разряда кода «Свой».

$$\mathrm{var}(\theta_i) \Rightarrow \begin{cases} \max(H("x_i")) \le 1,0 \ \text{бит} \\ H("c_i") = 0,0 \ \text{бит}. \end{cases} \qquad (2.3)$$

В этом случае вероятность ошибки первого рода (отказ в доступе «Своему») становится нулевой, преобразователь биометрия-код безошибочно узнает образ «Свой», обеспечивая максимально возможное значение энтропии для кодов «Чужой». Поясним процедуру настройки однопороговых квантователей рис. 2.1.

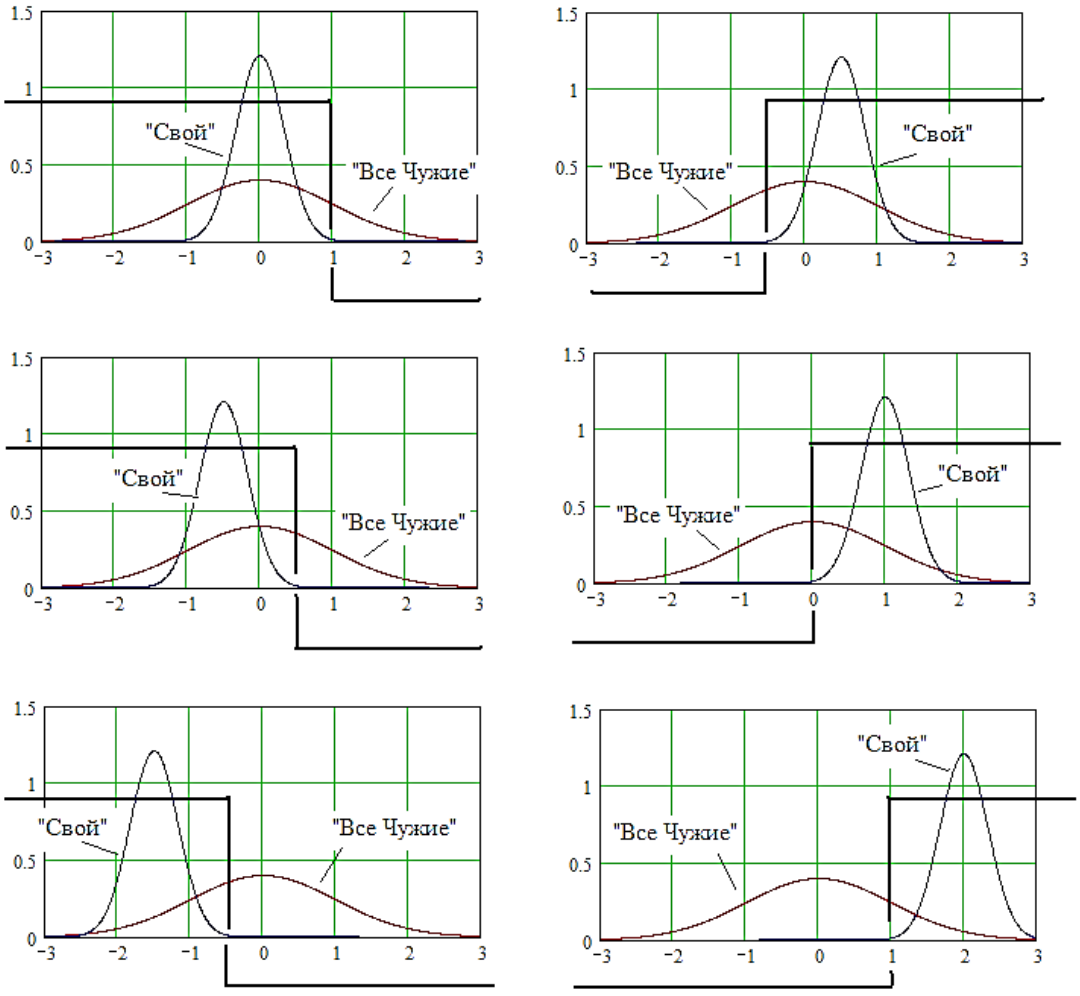


Рис. 2.1. Одностороннее выделение образа «Свой» состоянием «1» по порогу θ

Обычно распределение данных «Свой» много у́же распределения данных «Все Чужие». На сколько у́же распределение «Свой», зависит от стабильности воспроизведения биометрического образа. На сегодняшний день самыми стабильными являются данные рисунка радужной оболочки глаза [2]. Данные рис. 2.1 соответствуют рукописному почерку, для которого вариации параметров образа «Свой» оказываются при-

мерно в три раза меньше вариаций того же биометрического параметра для образов «Все Чужие».

Так как первой задачей является хорошее распознавание биометрических данных «Свой», при настройке порога – θ выбирают максимум или минимум значений распределения параметра «Свой». Выбор осуществляют, исходя их близости правой или левой границы распределения «Свой» к центру распределения «Все Чужие». Чем ближе граница «Свой» в центре распределения параметра образов «Все Чужие», тем выше значение энтропии данных «Чужие».

Из рис. 2.1 видно, что в первом приближении распределения данных «Свой» достаточно хорошо описываются нормальными законами. В связи с этим при настройке порогов квантователей можно пользоваться гипотезой нормальности, считая, что практически все данные «Свой» попадают в интервал $[E(\nu) - 3\cdot\sigma(\nu); E(\nu) + 3\cdot\sigma(\nu)]$, т.е. для настройки значения порога квантования достаточно знать среднеквадратическое отклонение данных «Свой» – $\sigma(\nu)$, математическое ожидание данных «Свой» – $E(\nu)$, а также математическое ожидание данных «Все Чужие» – $E(\xi)$, которое обычно оказывается нулевым $E(\xi) = 0$.

2.2. Энтропия биометрических кодов «Чужой»

Параллельно с высоким уровнем узнавания образов «Свой» при настройке квантователей нужно обеспечить высокий уровень отвержения «Чужих» (высокий уровень энтропии кодов «Все Чужие»). В этом отношении описанный выше алгоритм настройки квантователей «нечетких экстракторов» весьма конструктивен, так как каждый из квантователей настраивается независимо. Оценить энтропию i-го разряда после настройки можно по Шеннону через учет вероятностей появления короткого алфавита, состоящего всего из двух состояний «0» и «1»:

$$H("x_i") = -P("0_i")\cdot\log_2(P("0_i")) - P("1_i")\cdot\log_2(P("1_i"))\,. \qquad (2.4)$$

Очевидно, что мы не встретим проблем и при вычислении совместной энтропии двух разрядов биокода, так как алфавит их возможных состояний остается коротким:

$$H("x_1, x_2") = -\sum_{i=1}^{4} P_i("x_1, x_2")\cdot\log_2(P_i("x_1, x_2"))\,. \qquad (2.5)$$

Проблемы вычисления энтропии по Шеннону начинаются для биокодов с большим числом разрядов. Так биокод, полученный Даугманом [2]

из рисунка радужной оболочки глаза, имеет 2048 разрядов. Для того, чтобы оценить энтропию столь длинного кода, придется использовать огромное число опытов. Задача вычисления энтропии по Шеннону кодов длиной 2048 бит:

$$H("x_1, x_2, ..., x_{2048}") = -\sum_{i=1}^{2^{2048}} P_i("x_1, x_2, ..., x_{2048}") \times \times \log_2(P_i("x_1, x_2, ..., x_{2048}")), \quad (2.6)$$

является технически не выполнимой.

Обойти проблему вычисления энтропии длинных кодов удается, если отказаться от процедур Шеннона, предполагающих поиск редких событий при обработке больших массивов исходных кодов. Необходимо от наблюдения редких событий перейти к предсказанию вероятности их появления. Для этой цели перейдем из пространства кодов «Все Чужие» –"$\bar{x}$" в пространство расстояний Хэмминга между кодом аутентификации «Свой» и кодами «Все Чужие»:

$$h("\bar{x}") = \sum_{i=1}^{2048} "x_i" \oplus "c_i". \quad (2.7)$$

Пример распределения расстояний Хэмминга между кодом аутентификации «Свой» и кодами «Все Чужие» дан на рис. 2.2 для кодов длиной 2048 бит. Корректно настроенные квантователи всегда дают практически нормальный закон распределения расстояний Хэмминга до кодов «Все Чужие» с математическим ожиданием, равным половине длины биокода.

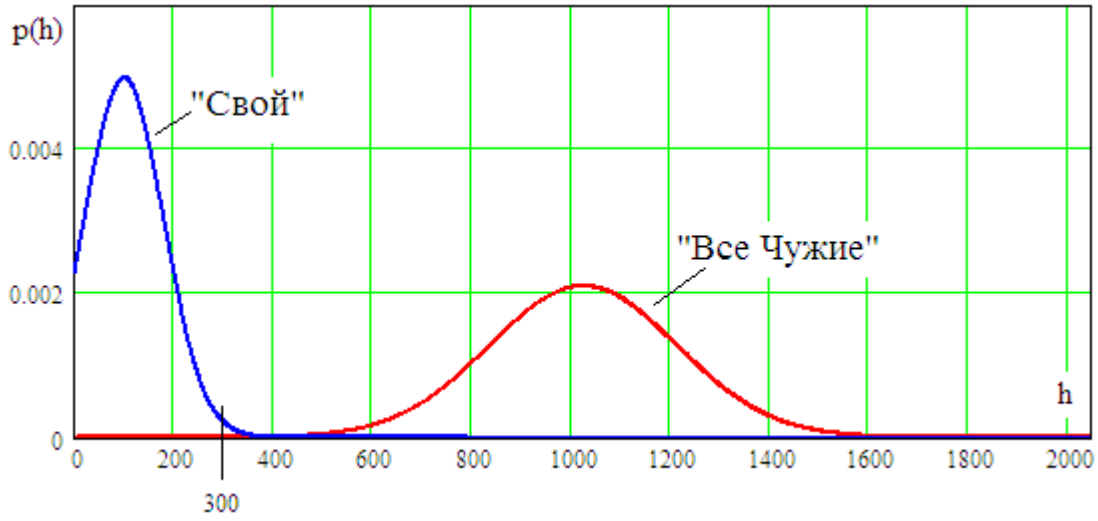


Рис. 2.2. Распределения расстояний Хэмминга между кодом «Свой» и «Все Чужие», а также между кодом «Свой» и ошибочными кодами «Свой»

Для настройки квантователей необходимо знать математическое ожидание и среднеквадратическое отклонение распределения данных «Свой» по каждому из биометрических параметров. Если обучать «нечеткий экстрактор» на 20 примерах образа «Свой», относительная ошибка вычисления математического ожидания параметра с высокой вероятностью составит примерно 20 %. Примерно такие же результаты получаются и при оценке среднеквадратического отклонения данных «Свой».

Из-за малого числа используемых при обучении примеров образа «Свой» не удается точно настроить квантователи контролируемых биометрических параметров. По этой причине коды-отклики на образ «Свой» не совпадают с кодом аутентификации «Свой». На рис. 2.2 эта ситуация отображена появлением распределения расстояний Хэмминга между кодом аутентификации «Свой» и ошибочными кодами «Свой».

Обычно ошибки кодов «Свой» в «нечетких экстракторах» правят классическими избыточными самокорректирующимися кодами [8–10]. Предположим, что 97 % неверных бит биокода не будут превышать значения 300 бит и будут правиться за счет избыточности самокорректирующимся кодом. В этом случае вероятность ошибок второго рода «нечеткого экстрактора» составит

$$P_2 = \frac{1}{\sigma(h("\bar{x}"))\sqrt{2\pi}} \int_{-\infty}^{300} \exp\left\{\frac{-(E(h("\bar{x}")-u)^2}{2\cdot\sigma^2(h("\bar{x}"))}\right\} du. \qquad (2.8)$$

Зная вероятность ошибок второго рода, мы можем оценить энтропию кодов длиной 2048 бит:

$$H("x_1, x_2, ..., x_{2048}") = -\log_2(P_2). \qquad (2.9)$$

Заметим, что при вычислениях (2.8) и (2.9) нет необходимости использовать миллиарды биокодов «Чужие». Вполне достаточно 200 кодов «Все Чужие». Этих данных вполне достаточно для оценки математического ожидания и среднеквадратического отклонения распределения расстояний Хэмминга «Все Чужие» с относительной погрешностью порядка 5 %.

2.3. Применение классических кодов с избыточностью для корректировки биометрических ошибок

Следует подчеркнуть, что длина ключа криптографической аутентификации не должна быть большой. Так длина криптографического

ключа для шифрования по отечественному стандарту ГОСТ 28147–89 составляет 256 бит, такую же длину ключа использует ГОСТ Р 34.10–94 при формировании цифровой подписи. Получается, что биокод Даугмана [2] длиной 2048 бит оказывается в восемь раз длиннее кода ключа доступа. Следует воспользоваться этим обстоятельством для обнаружения и корректировки ошибок в биокоде. Общий принцип обнаружения и корректировки ошибок избыточными кодами иллюстрируется рис. 2.3.

Повторение	1*	1	0	1*	1	0	1	1
	0	0*	0	0	1	1*	1	0*
	0	1	0	0	1	1*	1	1
	0	1	0	0	0*	0	1	1
	1*	1	0	0	1	0	0*	0*
ИТОГО	0	1	0	0	1	0	1	1

Рис. 2.3. Обнаружение ошибок и их исправление простейшим кодом с пятикратным дублированием

Как видно из данного рисунка, простейший код с пятикратным дублированием информации длиной 5 × 8 = 40 бит способен править до двух ошибок в каждом из восьми столбцов, т.е. код правит до 16 ошибок (до 40 %) при избыточности в 400 %. Код строится на анализе частоты появления состояния «0» и «1» в столбцах. Верным считается наиболее вероятное состояние анализируемых разрядов в столбце. По этой причине код не способен обнаруживать и править три и более ошибок в одном столбце, что является его главным недостатком.

Этот недостаток ослабляется, если пользоваться более сложными кодами с более сложными правилами обнаружения и корректировки ошибок [8–10]. Обычно классические самокорректирующиеся коды имеют явно выраженную информационную часть и избыточную часть, которая получается из информационной части. Структура получения подобных кодов и их применения приведена на рис. 2.4.

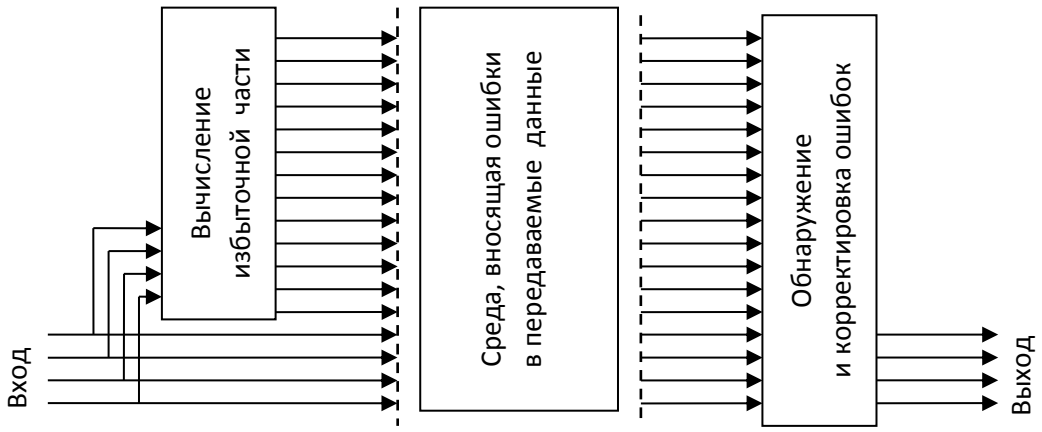


Рис 2.4. Ввод и устранение избыточной информации при использовании самокорректирующихся кодов

Независимо от типа используемого правила обнаружения ошибок все классические самокорректирующиеся коды требуют большой избыточности и не могут эффективно исправлять более половины ошибок в коде. Для корректировки 4 % ошибок требуется 50 % избыточности кода. Для полной корректировки 6 % ошибок в любой их последовательности требуется код со 100 % избыточности. Размер избыточной части кода экспоненциально растет с ростом необходимого числа обнаруживаемых и исправляемых ошибок. Для того, чтобы править порядка 50 % ошибок, самокорректирующийся код должен практически полностью состоять из избыточной части, имея очень короткую информативную часть.

Кроме того, классические самокорректирующиеся коды строились для каналов связи, где хорошо работает гипотеза равновероятного распределения ошибок. Для биометрических кодов гипотеза равновероятного распределения ошибок в разрядах не работает. Квантование биометрических данных обычно приводит к появлению стабильных и нестабильных разрядов со строго фиксированным положением. Информацию о показателе стабильности каждого из разрядов кода «Свой» классические самокорректирующиеся коды не способны учитывать, и, как следствие, их корректирующая способность уступает нейросетевым корректорам биометрических ошибок. Преимущества нейросетевых корректоров обусловлены тем, что во время обучения они становятся способны учитывать местоположение и показатель стабильности корректируемых разрядов.

2.4. Компрометация биометрических данных образа «Свой» в пространстве расстояний Хэмминга

Важным элементом биометрических технологий защиты информации является тайна применяемого при аутентификации образа «Свой». Проще всего сохранить в тайне данные биометрического образа «Свой», разместив их в аппаратной части средств биометрической аутентификации и исключив возможность программного доступа к «нечеткому экстрактору». Подобное техническое решение дороже программной защиты «нечетких экстракторов», в связи с этим биометрическая и криптографическая общественность значительное внимание уделяет исследованию вопросов трудоемкости извлечения биометрической информации из «нечетких экстракторов».

Из теории известно, что обратные задачи всегда намного сложнее прямых задач, т.е. настроить квантователи «нечеткого экстрактора» и синтезировать для него соответствующий самокорректирующийся код много проще, чем извлечь данные из настроенного «нечеткого экстрактора».

Попытаемся оценить сложность задачи восстановления неизвестного биометрического образа «Свой», если известен биометрический код аутентификации «Свой». Очевидно, что в этом случае мы можем восстановить избыточную часть самокорректирующегося кода «Свой» и экспериментально найти коды-отклики тестовой базы образов «Чужие». В итоге мы получим гистограмму распределения расстояний Хэмминга между кодами «Свой» и «Чужие» (рис. 2.5). Практика показала, что для проведения подобного численного эксперимента вполне достаточно тестовой базы примерно из 10 000 биометрических образов «Чужой».

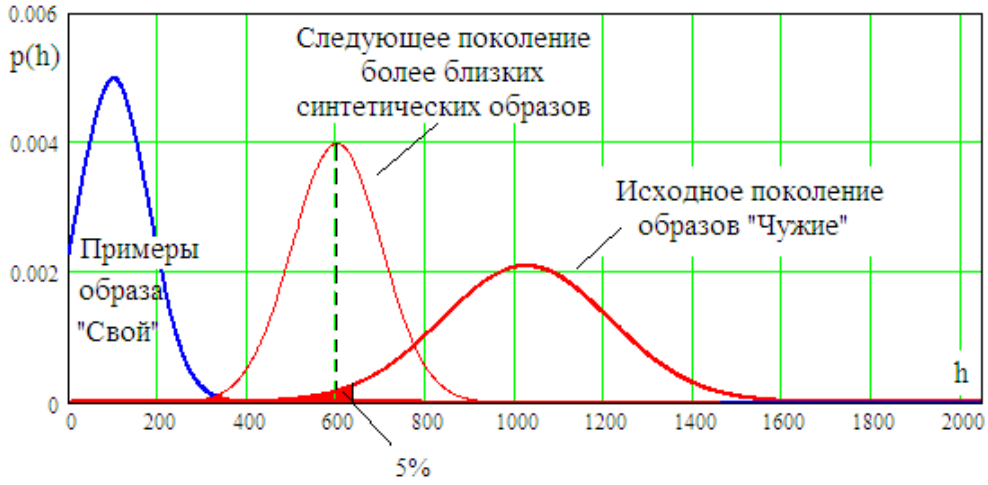


Рис. 2.5. Генетический алгоритм направленной селекции образов «Чужой», наиболее близких к образу «Свой»

Обычные для настоящего времени вычислительные машины позволяют проводить тестирование порядка 100 биометрических образов в секунду, т.е. гистограмму распределения расстояний Хэмминга удастся получить примерно через интервал в 15 мин. Далее мы можем отсортировать полученные данные, выбрав из них 5 % наиболее близких кодов, соответствующих 500 образов «Чужой», наиболее похожих на образ «Свой». В итоге мы получаем генетический материал для получения следующего поколения синтетических биометрических образов, близких к образу «Свой». На рис. 2.5 использованные 5 % кодов помечены темной заливкой на плотности распределения расстояний Хэмминга «Чужие».

Имея 500 биометрических образов родителей – близкий «Чужой», можно получить следующее поколение синтетических еще более близких образов «Чужой», воспользовавшись процедурами скрещивания образов родителей и получения от них потомков. Процедуры скрещивания биометрических образов родителей регламентируются ГОСТ Р 52633.2–2010 (см. табл. 1.1). Простейшим вариантом скрещивания является получение одного потомка, одинаково похожего на своих родителей, когда биометрические параметры образа потомка получают, складывая биометрические параметры образов родителей в равной пропорции. Из 500 исходных образов этот способ позволяет получить 500! вариантов биометрических образов потомков. При размножении образов потомков получается новое поколение с распределением расстояний Хэмминга, расположенным ближе к распределению расстояний кодов «Свой» (см. рис. 2.5).

Достаточно работы обычной вычислительной машины в течение примерно 30 мин для того, чтобы получить второе поколение синтетических биометрических образов, при этом несколько из образов следующего поколения будут давать биокод, точно совпадающий с биокодом аутентификации. Если этого не происходит, то нужно повторить процедуру селекции и получить следующее еще более близкое поколение биометрических образов «Почти Свой».

2.5. Компрометация кода «Свой» в пространстве расстояний Хэмминга

Приведенная в предыдущем параграфе процедура восстановления биометрических данных «Свой» по коду аутентификации многими англоязычными специалистами не воспринимается серьезно. Недоверие к подобным процедурам основано на том, что злоумышленник не может

заранее знать код аутентификации. Если же он знает код аутентификации, то он им воспользуется напрямую. При скомпрометированном коде криптографической аутентификации нет смысла выполнять дополнительную работу по восстановлению биометрических параметров образа «Свой». Для чистой криптографии подобная аргументация верна, для стыка биометрии и криптографии код аутентификации может быть скомпрометирован атакой на слабо защищенную биометрию. Покажем это на примере «нечетких экстракторов».

Так как мы не знаем код аутентификации "$\overline{c}$", будем подавать на «нечеткий экстрактор» поочередно 10 000 образов «Чужие» из тестовой базы. В итоге мы получим реальные коды-отклики "$\overline{x}_1$","$\overline{x}_2$",....,"$\overline{x}_{10000}$". Из теории нам известно, что энтропия кодов «Чужой» много больше энтропии кода «Свой»:

$$H("\overline{x}_i") \rangle\rangle H("\overline{c}"). \qquad (2.10)$$

Это означает, что мы можем подобрать код "$\overline{c}$", опираясь на реальные коды-отклики "$\overline{x}_1$","$\overline{x}_2$",....,"$\overline{x}_{10000}$" и фундаментальное условие (2.10). Для этой цели воспользуемся генератором случайных целых чисел с равномерным распределением в интервале от 0 до 2^{2048} и получим от этого генератора 1000 кодов "$\overline{z}_1$","$\overline{z}_2$",....,"$\overline{z}_{1000}$". Очевидно, что коды "$\overline{z}_i$" будут иметь разное расстояние Хэмминга до кода "$\overline{c}$". Чем ближе код "$\overline{z}_i$" будет к коду "$\overline{c}$", тем ниже значение будет у его энтропии. Если эти коды совпадут, то их энтропия может оказаться близка к нулю.

Для сортировки кодов "$\overline{z}_i$" по их энтропии относительно кодов "$\overline{x}_1$","$\overline{x}_2$",....,"$\overline{x}_{10000}$" необходимо вычислить соответствующие расстояния Хэмминга:

$$h("\overline{z}_i","\overline{x}_j") = \sum_{k=1}^{2048} "z_{k,i}" \oplus "x_{k,j}". \qquad (2.11)$$

Повторив вычисления для j:= 1,…,10000, мы получим множество распределений расстояний Хэмминга, примеры которых приведены на рис. 2.6. Далее, располагая множеством распределений Хэмминга для разных кодов, мы оказываемся способны оценить их статистические моменты и найти энтропию этих кодов по формулам (2.8) и (2.9). Обладая оценками энтропии кодов $H("\overline{z}_1"), H("\overline{z}_2"),...., H("\overline{z}_{1000}")$, мы можем осуществить их упорядочивание по возрастанию значения энтропии. Теперь, взяв коды с минимальными значениями энтропии, мы отфильтруем из всех кодов ближайшие к коду аутентификации. Цифры

на рис. 2.6 соответствуют возрастанию значений энтропии распределений расстояний Хэмминга. Чем выше среднеквадратическое отклонение расстояний Хэмминга, тем ниже энтропия соответствующего кода "$\overline{z}_i$".

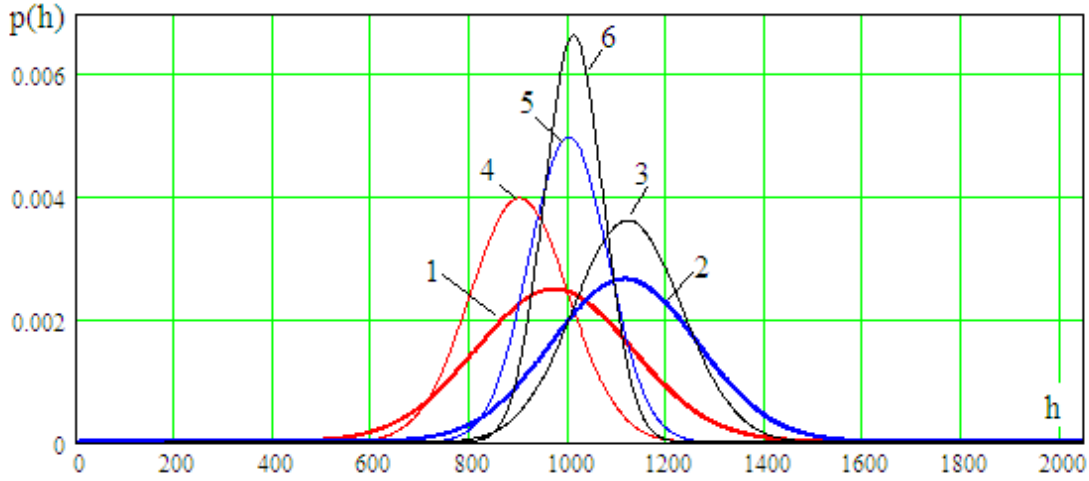


Рис. 2.6. Распределения расстояний Хэмминга между случайными кодами "$\overline{z}_i$" и кодами «Все Чужие»

Если теперь взять код "$\overline{z}_i$" с минимальным значением энтропии и относительно него вычислить расстояния Хэмминга для других кодов с малым значением энтропии, мы будем наблюдать две хорошо разделимые группы расстояний. Одна из этих групп будет близка к коду аутентификации, а другая – к его инверсии. На рис. 2.7 даны получающиеся расстояния Хэмминга для двух хорошо разделимых групп кодов.

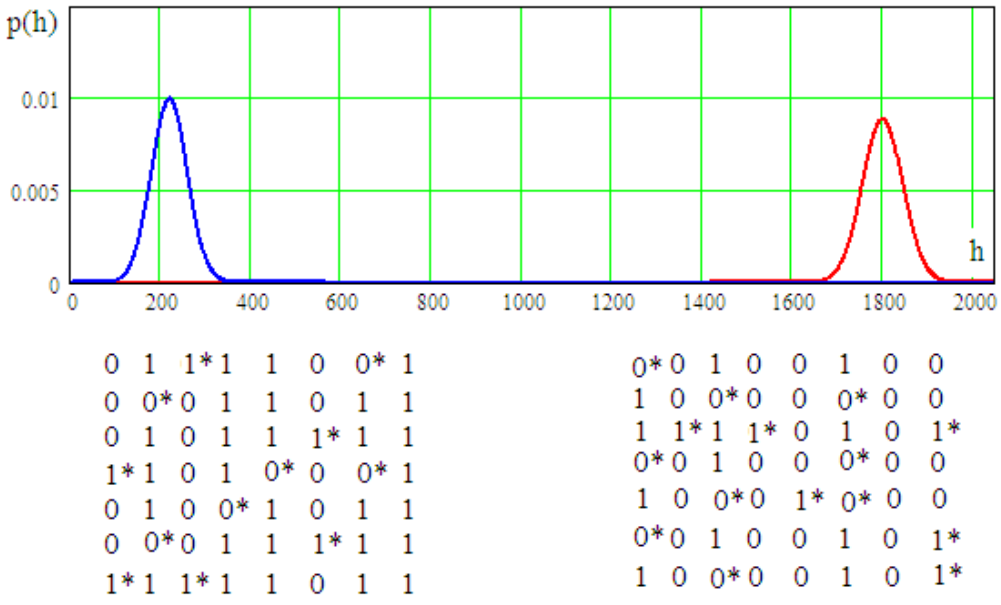


Рис. 2.7. Распределения двух групп кодов, близких к коду "$\overline{c}$" и его инверсии

В нижней части рис. 2.7 даны примеры одинаковых фрагментов кодов, принадлежащих к разным группам распределений расстояний Хэмминга. Внутри каждой из групп кодов можно восстанавливать значение кода по большинству встречающихся в столбце состояний. Ошибки в каждой из групп кодов помечены символом *.

Практика показала, что на выборке в 10000 кодов-откликов "$\overline{x}_i$" и выборке в 1000 случайных кодов "$\overline{z}_i$" удается верно восстановить до 98 % разрядов кода аутентификации. Столь высокие показатели эффективности атаки статистической компрометации биокодов «нечетких экстракторов» требуют принятия мер по дополнительной защите биометрических данных. Положение осложняется тем, что вычисления, на которых строится операция по компрометации ключа аутентификации, осуществляются быстро. Они не требуют работы с биометрическими образами, обрабатываются только числа. Описанный в предыдущем параграфе генетический алгоритм размножения биометрических образов-родителей много медленнее, чем алгоритм генерации и оценки энтропии чисел "$\overline{z}_i$" с последующим восстановлением из них кода аутентификации.

2.6. Защита биокода простым наложением гаммы

Одним из приемов защиты биометрических кодов аутентификации, активно применяемых в США и странах Евросоюза [2–4], является наложение гаммы. Гамма, например, может быть получена из ключа аутентификации, накрытого избыточным самокорректирующимся кодом, как это показано на рис. 2.8.

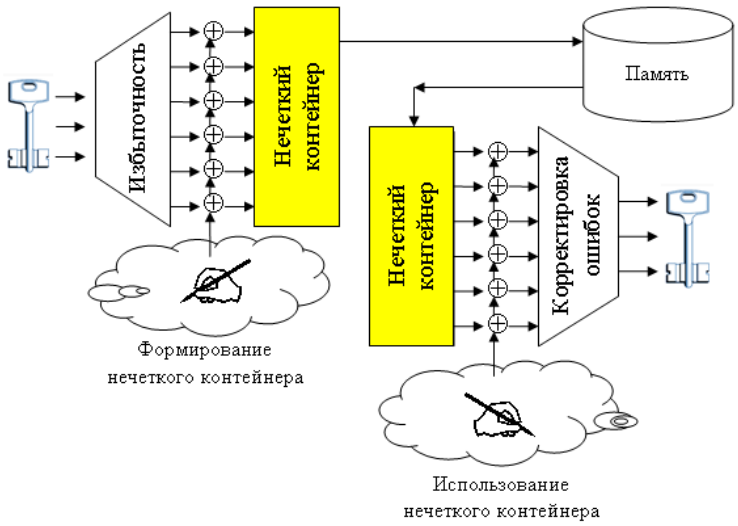


Рис. 2.8. Формирование и использование нечетких контейнеров, защищенных гаммированием

Проблема «нечетких экстракторов» состоит в том, что им необходимо одновременно и корректировать биокод с ошибками, и скрывать данные. Для того, чтобы одновременно и скрывать данные, и корректировать ошибки в них, ключ аутентификации размещают в информационной части самокорректирующегося кода. Далее вычисляют его избыточную часть в форме синдромов для обнаружения и правки ошибок (см. рис. 2.4). Так как код ключа аутентификации сохраняется в тайне, объединение его с синдромами ошибок можно рассматривать как некоторую защищающую информацию гамму. Если биокод накрыть этой гаммой, то мы получим некоторый «нечеткий контейнер», который может быть относительно безопасно размещен в памяти программного средства аутентификации.

Когда нужно воспользоваться «нечетким контейнером», его извлекают из памяти и складывают по модулю два с проверяемым биокодом. При этом восстанавливается самокорректирующийся код ключа аутентификации. Пользуясь информационной частью самокорректирующегося кода и синдромами ошибок, удается исправить ошибки. Восстановленный этим способом биоключ далее можно применять в каком-либо из известных протоколов криптографической аутентификации.

К сожалению, столь простой способ защиты не эффективен. Реализации атак, описанных в двух предыдущих параграфах, не усложняются. Высокая уязвимость защиты биометрических данных простым гаммированием обусловлена тем, что одинаковое гаммирование разных кодов не влияет на значение расстояний Хэмминга между ними:

$$h("\overline{x}_i","\overline{c}") = \sum_{k=1}^{2048} "x_{i,k}" \oplus "c_k" = \sum_{k=1}^{2048} ("g_k" \oplus "x_{i,k}") \oplus ("g_k" \oplus "c_k"), \quad (2.12)$$

где $"g_k"$ – состояния «0» или «1» k-го разряда, маскирующей биокод гаммы.

В силу коммутативности операции сложения по модулю два при вычислении расстояний Хэмминга двух кодов происходит снятие одной и той же гаммы, маскирующей сравниваемые коды. Это означает, что описанные ранее атаки работают и для схемы, отображенной на рис. 2.8. Гаммирование данных дает не более чем иллюзию защищенности, если атаковать «нечеткие экстракторы» в пространстве расстояний Хэмминга.

2.7. Перечень эффективных метрик, используемых при компрометации биоданных «Свой», защищенных гаммированием

Чистые математики делят алгоритмы перебора на полиномиальные (P-алгоритмы) и неполиномиальные (NP-алгоритмы) [11]. Полиноми-

альные алгоритмы считаются инженерно реализуемыми, и, как следствие, задачи, решаемые алгоритмом полиномиальной вычислительной сложности, не могут рассматриваться как серьезная защита, такие алгоритмы не являются криптографическими.

Для нас важно то, что операция гаммирования (длина гаммы равна длине бинарного сообщения) в классической криптографии считается вычислительно сложной, т.е. NP-полной. Как показано в предыдущем параграфе, та же самая процедура гаммирования при защите биокодов, оказывается, дает полиномиальную сложность перебора. Это следствие изменения природы защищаемых данных. На защищенный гаммой неизвестный текст мы никак не можем повлиять (нет дополнительных входов влияния), и задача снятия неизвестной гаммы с неизвестного сообщения имеет экспоненциальную вычислительную сложность.

Если мы имеет дело с преобразователями биометрия-код, то мы накрываем непрерывные данные (континуумы) после их квантования гаммой. Объект защиты коренным образом изменился, и, соответственно, появились новые уязвимости и новые возможности для атаки. Важно сделать так, чтобы все новые уязвимости были понятны, а затем важно подобрать условия, в которых новые уязвимости могут быть надежно защищены. Данное пособие как раз и создано для описания нового объекта защиты (преобразователей биометрия-код), описания их уязвимостей, методов защиты от новых атак на новые уязвимости.

Если оставаться в рамках классических методов исследования кодов, то «нечеткие экстракторы» кажутся надежными. Например, если «нечеткий экстрактор» имеет реальную стойкость 10^{10} попыток подбора, мы должны создать базу образов «Чужой» не менее чем из 10^{10} образов и пытаться найти коллизию образов «Свой» и «Чужой». При этом с очень высокой вероятностью мы не найдем опасной коллизии и сделаем ложный вывод о стойкости «нечеткого экстрактора» выше 10^{10} к попыткам подбора. Почему это произойдет? Потому, что мы занимались «слепым» перебором. «Слепой» перебор – это как раз и есть иллюзия NP-полноты или иллюзия экспоненциальной сложности задачи.

На самом деле внутри исследуемых биокодов (откликов «нечетких экстракторов») есть некоторая статистическая информация, используя которую, мы перестаем быть слепыми и становимся способны осуществлять направленный перебор. Очевидно, что фундаментальной метрикой сравнения биокодов образов «Чужой» является их энтропия – $H("\bar{x}_i")$. Чем меньше энтропия образа «Чужой» тем он ближе к образу «Свой». Появление метрики энтропии сразу же делает задачу компрометации данных образа «Свой» полиномиальной. Однако для того, чтобы

быстро вычислять энтропию длинных кодов, необходимо переходить в пространство расстояний Хэмминга. Только благодаря тому, что расстояние Хэмминга не чувствительно к гаммированию, мы можем технически реализовать атаку поиска биометрических образов «Чужой» с минимальной энтропией.

Заметим, что хэммингова расстояния между сравниваемыми кодами – это не единственная метрика, не чувствительная к гаммированию биокодов. Метрика стабильности разрядов биометрических кодов также не чувствительна к защите биокодов гаммой. Метрика стабильности i-го разряда определяется через вероятности появления в разряде состояний «1» или «0»:

$$\gamma_i = 2 \cdot \left| 0{,}5 - P("1_i") \right| = 2 \cdot \left| 0{,}5 - P("0_i") \right|. \qquad (2.13)$$

В предельной ситуации $P(«0_i») = P(«1_i») = 0{,}5$ показатель стабильности оказывается нулевым. Если же значение разряда не меняется $P(«0_i») = 1$ или $P(«1_i») = 1$, то показатель стабильности оказывается единичным. Примеры гистограмм распределения показателей стабильности образа «Свой» и образов «Чужой» приведены на рис. 2.9.

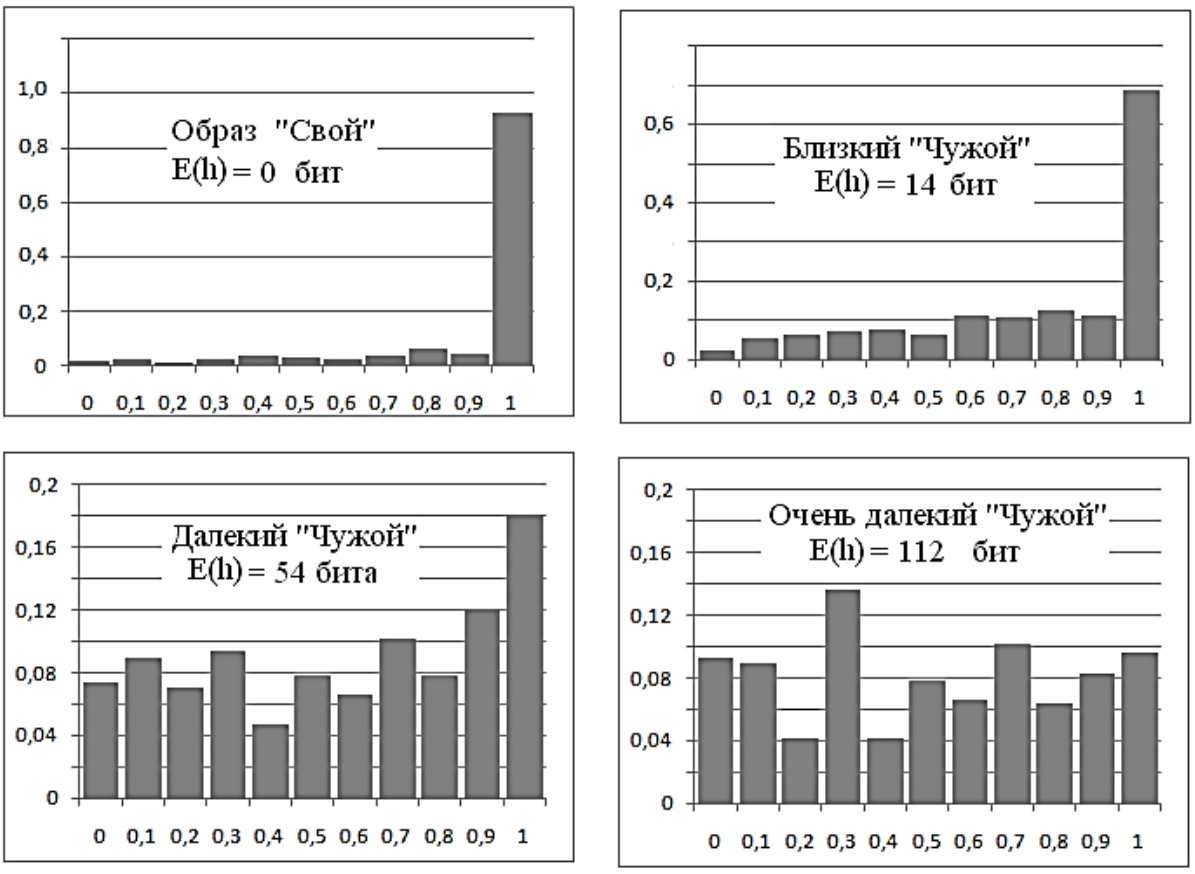


Рис. 2.9. Падение стабильности разрядов кодов «Чужой» по мере удаления образа «Чужой» от образа «Свой»

Из данного рисунка видно, что по мере увеличения расстояния Хэмминга образа «Чужой» от образа «Свой» стабильность разрядов биокодов падает. Такой показатель, как средняя стабильность всех разрядов биокода – $E(\gamma)$, может использоваться для определения направления движения в сторону образа «Свой». При таком направленном переборе следует стремиться увеличивать показатель стабильности разрядов при генетической селекции близких образов «Чужой».

Очевидным является то, что гаммирование биокода не может повлиять на показатели стабильности разрядов кода. Гаммирование – это детерминированная операция, которая приводит только к смене наиболее вероятного состояния (состояние «0» может смениться на состояние «1»). Это никак не влияет на значение показателя стабильности (2.13), так как его можно вычислять и через вероятность $P(«0_i»)$, и через вероятность $P(«1_i»)$. Результат вычислений будет одинаковым.

Так как показатель стабильности разрядов биокода и метрика расстояний Хэмминга между ними не чувствительны к гаммированию, их комбинация также является не чувствительной, т.е. атаки компрометации «нечетких экстракторов» могут строиться на использовании взвешенной метрики Хэмминга:

$$\gamma h = \sum_{i=1}^{2048} \gamma_i \cdot ("x_i" \oplus "c_i"). \qquad (2.14)$$

Применение взвешенной метрики Хэмминга позволяет значительно экономить вычислительные ресурсы при реализации генетических алгоритмов направленного перебора. Причина выигрыша в том, что при вычислениях большее внимание уделяется значимым (стабильным) разрядам биокодов. Влияние нестабильных разрядов биокодов уменьшается, растет стабильность результатов статистических оценок.

Последним известным на сегодняшний день показателем выбора направления перебора является среднее значение модулей коэффициентов парной корреляции между разрядами биокода – $E(|r_{i,j}|)$. Гаммирование исследуемых биокодов может приводить только к смене знака коэффициентов парной корреляции, но не влияет на модуль этого показателя. В связи с этим среднее значение модулей коэффициентов парной корреляции между разрядами биокода $E(|r_{i,j}|)$ может использоваться для выбора направления перебора. При этом следует выбирать те образы «Чужой» для генетической селекции, у которых выше внутренние корреляционные связи. У разрядов биокода «Свой» «нечеткого экстрактора» модули корреляционных связей максимальны.

Корреляционная метрика среднего значения модулей коэффициентов парной корреляции играет важную роль в теории нейросетевых преобразователей биометрия-код. Опираясь на эту метрику, удается вычислять энтропию длинных кодов, минуя промежуточную процедуру оценки расстояний Хэмминга, например, пользуясь номограммой, приведенной на рис. 3.6. При вычислении этой метрики вообще не нужны данные в виде примеров образа «Свой» и данные о биокоде «Свой».

3. НЕЙРОСЕТЕВЫЕ ПРЕОБРАЗОВАТЕЛИ БИОМЕТРИЯ-КОД ДОСТУПА

3.1. Общие положения нейросетевой биометрической аутентификации

Основной проблемой «нечетких экстракторов» является то, что их квантователи работают с «сырыми» биометрическими данными, а обогащение данных осуществляется в цифровой форме. Одновременно обогащать данные и защищать их шифрованием технически трудно. Желательно разделить операции, обогащая «сырые» биометрические данные еще в континуальной форме до квантования. Тогда после квантования дискретные данные будут иметь достаточно высокое качество, что снимает ряд проблем последующей защиты биокодов.

На рис. 3.1 дается общая схема нейросетевой аутентификации и ее подготовки, сводящейся к обучению искусственной нейронной сети преобразователя биометрия-код.

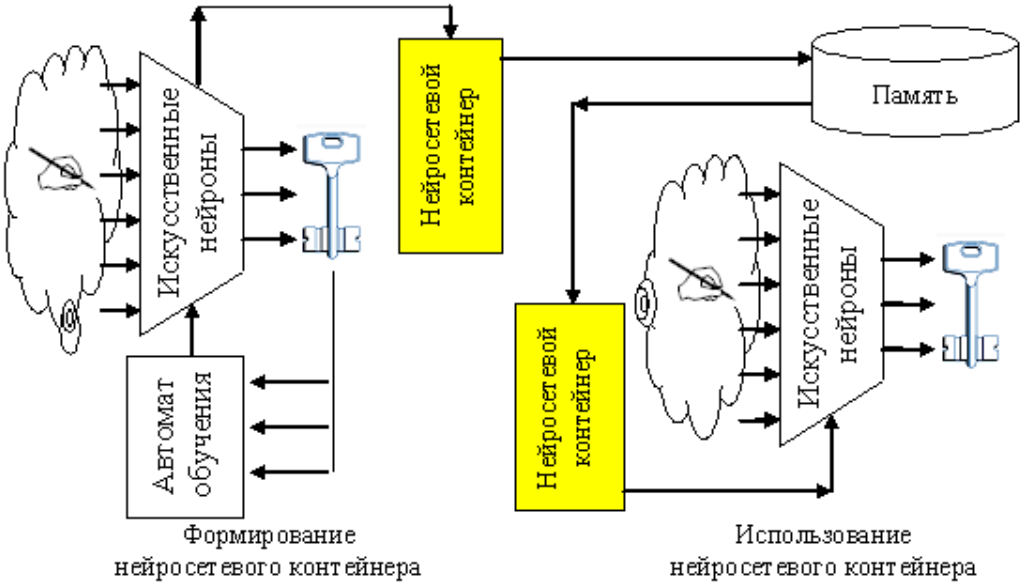


Рис. 3.1. Блок-схема нейросетевой аутентификации и ее подготовки

Нейронные сети требуют их обучения. Во время обучения должны использоваться биометрические образы «Свой» и образы «Чужие», также должен быть задан код ключа, соответствующий образу «Свой». Очевидно, что процедура обучения является потенциально опасной, так как требуется присутствие конфиденциальной информации (образа «Свой» и кода ключа). Так как обучение опасно, оно должно занимать

минимальное время и осуществляться автоматически. После обучения информация об образах «Свой» и кода ключа уничтожается. Остаются только данные об обученной нейронной сети в виде таблицы связей нейронов и таблицы весовых коэффициентов сумматоров нейронов. Две эти таблицы дают полную информацию о нейросети, удобно эти таблицы хранить отдельно в виде нейросетевого контейнера, фактически содержащего всю информацию о биометрии пользователя и его ключе.

Нейросетевой контейнер может храниться в памяти информационной системы, если есть гарантии сохранения конфиденциальности этой информации. Если нет гарантий сохранения конфиденциальности данных нейросетевого контейнера, то его необходимо защитить. Защита нейросетевого контейнера, например, может осуществиться самошифрованием, когда часть выходного биокода нейросетевого преобразователя используется для шифрования части еще не использованных данных нейросетевого контейнера. Блок-схема защиты нейросетевого контейнера самошифрованием приведена на рис. 3.2 (СКЗИ – средства криптографической защиты информации).

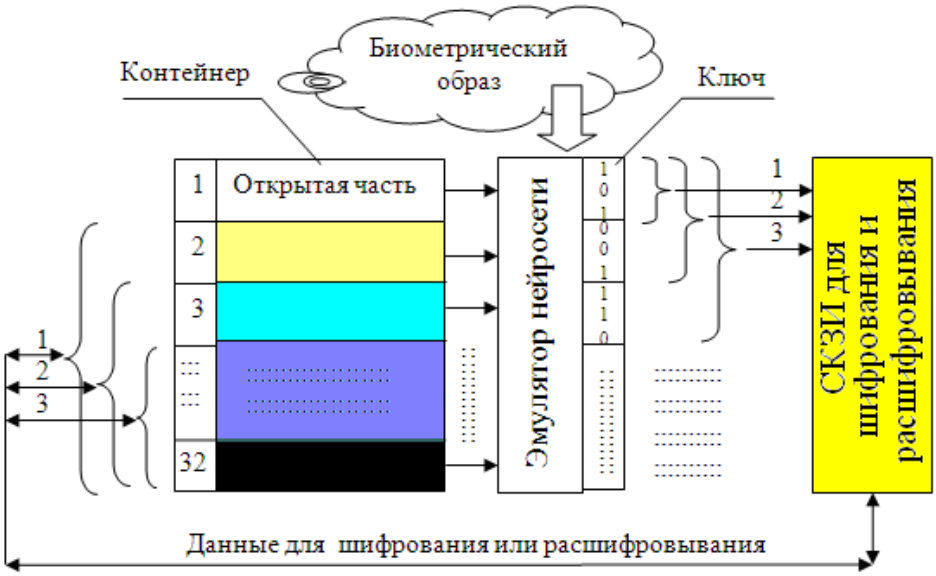


Рис. 3.2. Защита нейросетевого контейнера самошифрованием

Если защита биометрических данных осуществляется стандартными симметричными алгоритмами шифрования, то ее можно считать надежной, так как стандартные алгоритмы (ГОСТ 28147–89, DES, ASE) построены с применением гаммирования и перемешивания данных. Из-за перемешивания возникает эффект размножения биометрических ошибок, который мешает наблюдать статистики расстояний Хэмминга, реальные значения энтропии образов «Чужой», а также показатели ста-

бильности и коррелированности разрядов биокодов. Все это делает нейросетевые преобразователи биометрия-код гораздо более защищенными в сравнении с «нечеткими экстракторами».

При биометрической аутентификации происходят обратные процессы. Нейросетевой контейнер извлекается из памяти, далее по его данным формируется искусственная нейронная сеть. Если была осуществлена защита контейнера, то открытой оказывается только первая часть нейронов. Далее производится подача биометрических данных проверяемой личности на открытые нейроны. Если предъявлен образ «Свой», то на выходе у первой части нейронов появляется верная часть кода «Свой», которая расшифровывает следующую часть нейронов. Для образа «Свой» процедуры шифрования и расшифровывания данных оказываются симметричными и не мешают аутентификации.

Иная ситуация возникает, когда предъявленный образ оказывается «Чужим». В этом случае первая и последующие части кода на выходах нейронной сети оказываются случайными, верного расшифровывания данных нейросетевого контейнера не происходит. Возникает эффект хэширования выходных данных, разрушающий корреляционные связи и другие статистики образов «Чужие». Злоумышленник уже не может наблюдать статистические закономерности образов «Чужие», которые могут позволить ему осуществлять направленный перебор.

Хэширование данных нейросети после их неверного расшифровывания приводит к тому, что дисперсия распределения расстояний Хэмминга сжимается, как это показано на рис. 3.3.

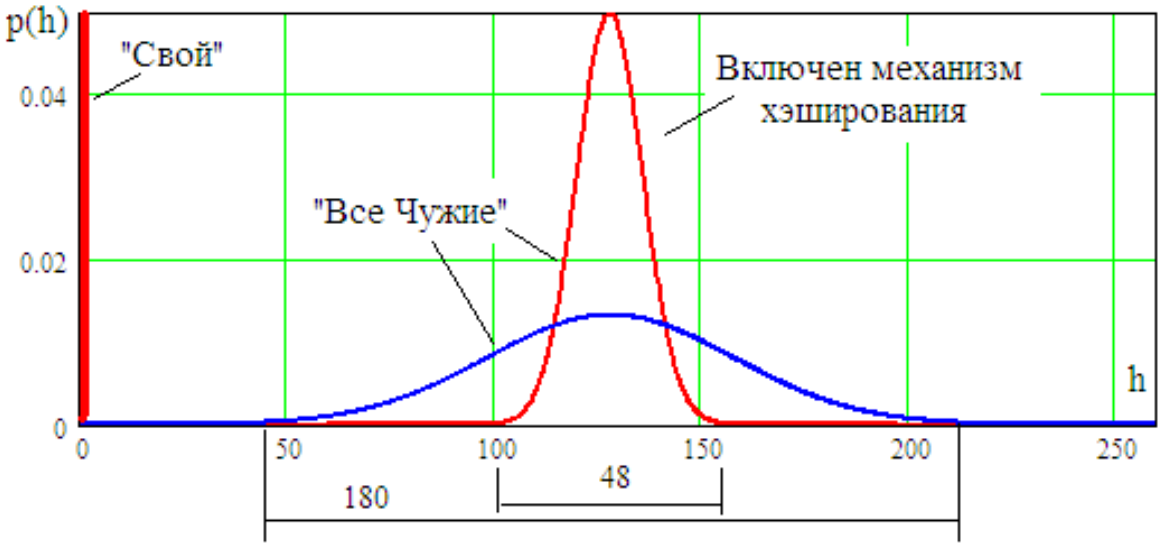


Рис. 3.3. Эффект сжатия распределения расстояний Хэмминга между кодом «Свой» и кодами «Все Чужие» после включения механизма размножения ошибок

Из данного рисунка видно, что при включении механизма размножения ошибок (хэширования данных) происходит примерно трехкратное сжатие среднеквадратического отклонения распределений расстояний Хэмминга. Нейросеть без хэширования имеет $\sigma(h) = 30$ бит, нейросеть с включенным механизмом хэширования имеет $\sigma(h) = 8$ бит. Распределение расстояний Хэмминга для образов «Свой» у хорошо обученных нейросетевых преобразователей находится в интервале от 0 до 1 (ошибок первого рода нет). Это дает нам право оценивать вероятность ошибок второго рода по следующей формуле:

$$P_2 = \frac{1}{\sigma(h("\bar{x}"))\sqrt{2\pi}} \int_{-\infty}^{1} \exp\left\{\frac{-(E(h("\bar{x}") - u)^2}{2 \cdot \sigma^2(h("\bar{x}"))}\right\} du. \qquad (3.1)$$

Подстановка данных распределений рис. 3.3 дает для не защищенного нейросетевого контейнера $P_2 = 10^{-5}$, что соответствует энтропии биокода 16,5 бита. Если же мы попытаемся оценить вероятность ошибок второго рода для защищенного контейнера, то получим $P_2 \approx 10^{-77}$, что соответствует энтропии в 256 бит как у идеального криптографического ключа длиной 256 разрядов, т.е. включение механизма размножения ошибок препятствует наблюдению реальных статистик распределений расстояний Хэмминга кодов «Все Чужие». Вместо реального распределение расстояний Хэмминга мы видим идеальное распределение, соответствующее идеальной криптографической защите.

3.2. Однослойный нейросетевой преобразователь биометрии

Нейросетевой преобразователь биометрии в код подобен «нечеткому экстрактору». Отличие между ними состоит лишь в том, что нейрон перед операцией квантования биометрических данных осуществляет их суммирование с некоторыми весовыми коэффициентами. В свою очередь, весовые коэффициенты при обучении подбираются так, чтобы сделать энтропию каждого разряда биокода максимальной, сохранив высокую узнаваемость образа «Свой». Алгоритм обучения нейронов может быть любым, главное, чтобы он был полностью автоматизирован и получал приемлемый результат за короткий интервал времени.

При предъявлении примеров образа «Свой» обученный однослойный нейросетевой преобразователь описывается следующей системой нейросетевых функционалов:

$$\overline{L}\left\{\begin{bmatrix} \mu_{1,1} & \mu_{1,2} & :::: & \mu_{1,N} \\ \mu_{1,2} & \mu_{2,2} & :::: & \mu_{2,N} \\ :::: & :::: & :::: & :::: \\ \mu_{1,n} & \mu_{2,n} & :::: & \mu_{n,N} \end{bmatrix} \times \begin{bmatrix} v_1 \\ v_2 \\ v_3 \\ :::: \\ v_N \end{bmatrix}\right\} = \begin{bmatrix} "c_1" \\ "c_2" \\ :::: \\ "c_n" \end{bmatrix}. \qquad (3.2)$$

Если подавать на входы обученной однослойной нейронной сети биометрические параметры примеров образа «Чужой», то преобразователь должен давать разные выходные коды на данные разных примеров:

$$\overline{L}\left\{\begin{bmatrix} \mu_{1,1} & \mu_{1,2} & :::: & \mu_{1,N} \\ \mu_{1,2} & \mu_{2,2} & :::: & \mu_{2,N} \\ :::: & :::: & :::: & :::: \\ \mu_{1,n} & \mu_{2,n} & :::: & \mu_{n,N} \end{bmatrix} \times \begin{bmatrix} \xi_1 \\ \xi_2 \\ \xi_3 \\ ::::: \\ \xi_N \end{bmatrix}\right\} = \begin{bmatrix} "x_1" \\ "x_2" \\ :::: \\ "x_n" \end{bmatrix}. \qquad (3.3)$$

Настройка вектора квантователей нейронов $\overline{L}$ ::: осуществляется так, чтобы изменение выходного состояния нейрона происходило в центре распределения данных «Все Чужие», как это показано на рис. 3.4. В этом случае энтропия выходных кодов образов «Все Чужие» оказывается максимальной.

Для того, чтобы преобразователь биометрия-код хорошо узнавал образ «Свой», осуществляют подбор весовых коэффициентов нейрона – μ_i – таким образом, чтобы распределение данных «Свой» сместилось из центра распределения «Все Чужие» на его периферию. На рис. 3.4 отображено смещение распределения данных «Свой» в правую сторону, так как образ «Свой» по условиям обучения должен давать состояние «1» на выходе у нейрона. Если бы требовалось иметь иное состояние, то при обучении распределение «Свой» смещалось бы в другую сторону.

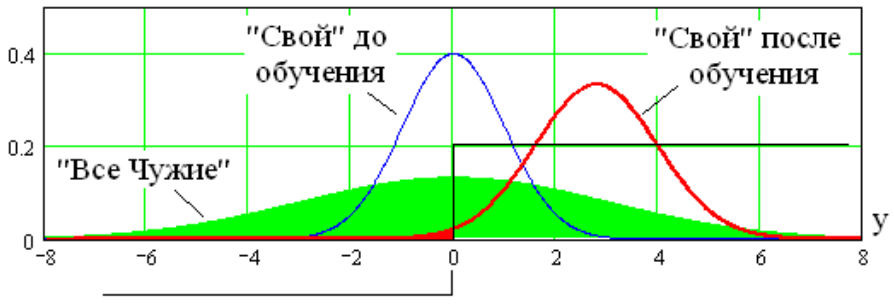


Рис. 3.4. Эффект выталкивания распределения данных образа «Свой» из области распределения данных «Все Чужие» на выходе сумматора нейрона

Если биометрические параметры на входах обучаемого нейрона имеют высокое качество, то можно использовать обычный итерационный алгоритм обучения. Блок-схема алгоритма итерационного обучения дана на рис. 3.5.

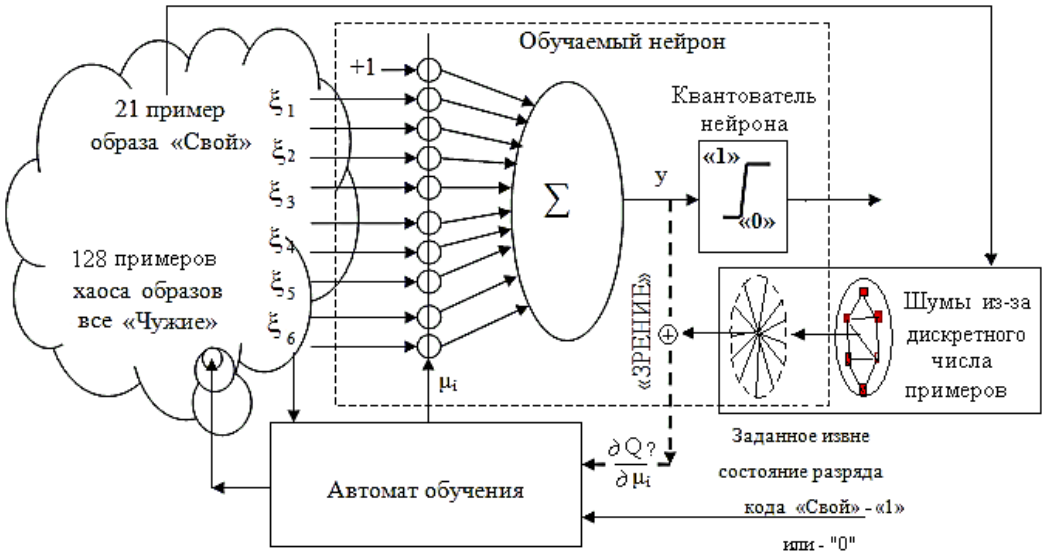


Рис. 3.5. Блок-схема итерационного алгоритма обучения одного нейрона на биометрических данных «хорошего» качества

То, насколько далеко мы вытолкнем распределение образа «Свой», зависит от качества работы нейрона. Для контроля текущего качества обучения используется следующий показатель:

$$Q = \frac{\left|E(y_v) - E(y_\xi)\right|}{\sqrt{\sigma_v \cdot \sigma_\xi}}, \qquad (3.4)$$

где $E(y_v)$ – математическое ожидание примеров «Свой»; $E(y_\xi)$ – математическое ожидание примеров «Все Чужие»; σ_v – среднеквадратическое отклонение примеров «Свой»; σ_ξ – среднеквадратическое отклонение примеров «Все Чужие».

Смысл показателя качества (3.4) достаточно прост. Он растет при разнесении центров разделяемых множеств «Свой» и «Все Чужие» и сохранении их среднеквадратических отклонений. Также растет показатель качества (3.4), если при обучении уменьшается среднее геометрическое среднеквадратических отклонений разделяемых множеств.

Пользуясь показателем качества (3.4), можно построить итерационную процедуру обучения нейрона по критерию движения в сторону повышения качества. Для этой цели необходимо вычислять частные

производные по качеству обучения $\frac{\partial Q}{\partial \mu_i}$ по каждому из настраиваемых весовых коэффициентов. Если производная положительна, то следует увеличивать значение регулируемого весового коэффициента – μ_i. При отрицательной производной качества следует уменьшать регулируемый весовой коэффициент.

Казалось бы, что, вычисляя частные производные по качеству обучения, мы легко можем построить автомат итерационного обучения. Это действительно так, когда речь идет о обучении нейронов с малым числом входов при биометрических данных «хорошего» качества. Если это не так, то обучать нейроны с большим числом «плохих» входных данных оказывается трудно. Проблема состоит во внутренних шумах процедуры обучения, которые обусловлены малым дискретным числом примеров обучения «Свой». Заранее создать очень много примеров «Все Чужие» технически не сложно. Много сложнее заставить пользователей прикладывать свои усилия, создавая десятки или даже сотни примеров образа «Свой».

Вычислительная проблема обучения состоит в неустойчивости вычисления частных производных качества $\frac{\partial Q}{\partial \mu_i}$? при малом числе примеров «Свой». Погрешности $\Delta E(y_v)$, $\Delta\sigma(y_v)$, возникающие из-за малого числа примеров «Свой», усиливаются при вычислении производных $\frac{\partial Q}{\partial \mu_i}$. Влияние погрешностей «забивает» полезную компоненту определения направления движения в сторону $\max\{Q\}$. Возникает хаотическое блуждание по многомерной поверхности ошибок вместо направленного движения в сторону максимума качества обучения. Эта ситуация иллюстрируется рис. 3.6.

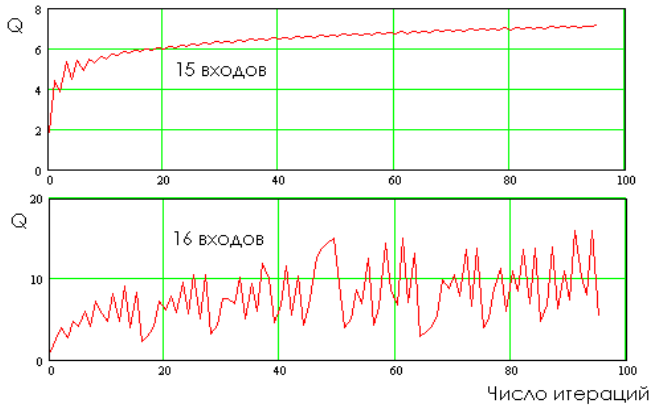


Рис. 3.6. Утрата устойчивости процедуры итерационного обучения одного нейрона

Из данного рисунка видно, что при 15 входах нейрона итерационное обучение происходит почти устойчиво, однако при увеличении числа входов нейрона до 16 устойчивость итерационного обучения утрачивается. Появляется множество так называемых локальных минимумов и максимумов качества обучения, являющихся не чем иным, как усилением собственного шума вычислительных процедур, обусловленного конечным числом примеров континуума данных образа «Свой».

Для того, чтобы сделать алгоритм обучения быстрым, ГОСТ Р 52633.5–2011 (см. табл. 1.1) рекомендует вообще отказаться от итерационного поиска оптимального решения. Вместо вычисления производных по качеству обучения стандарт рекомендует вычислять модули весовых коэффициентов нейрона по следующей формуле:

$$|\mu_i| = \frac{|E(\xi_i) - E(\nu_i)|}{\sigma(\nu_i)} \tag{3.5}$$

для нормированных биометрических данных $\sigma(\xi_i) = 1$.

Выбор знаков при весовых коэффициентах зависит от направления выталкивания распределения данных «Свой». При выталкивании в правую часть знаки выбираются следующим образом:

$$\text{sign}(\mu_i) = -\text{sign}(E(\xi_i) - E(\nu_i)). \tag{3.6}$$

При выталкивании данных «Свой» в левую часть распределения данных «Все Чужие» используют обратные знаки при весовых коэффициентах:

$$\text{sign}(\mu_i) = +\text{sign}(E(\xi_i) - E(\nu_i)). \tag{3.7}$$

Эти почти вычисления несколько снижают качество обучения, но делают алгоритм обучения очень быстрым и абсолютно устойчивым. Автомат обучения, осуществляющий вычисления по формулам (3.5) – (3.7), является абсолютно устойчивым и позволяет обучать большие нейронные сети за малый интервал времени порядка 0.1 с на обычных вычислительных машинах.

Если же мы попытаемся реализовать один из известных алгоритмов итерационного обучения нейронов [12, 13], то получим неустойчиво работающий автомат, осуществляющий обучение за неопределенное время.

3.3. Многослойные нейросетевые преобразователи биометрии в код

Теоретически нейросетевой преобразователь биометрия-код может иметь любое число слоев нейронов. Обучать такие преобразователи следует послойно. Каждый слой должен обучаться самостоятельно, используя данные выходов предыдущего слоя как примеры для обучения следующего слоя.

Матричное описание многослойных нейронных сетей осуществляется по аналогии с описанием однослойной сети. Например, двухслойная нейронная сеть при воздействии на нее образами «Свой» и «Чужой» будет описываться следующими матричными уравнениями:

$$\bar{B}\left\{\begin{bmatrix}\beta_{1,1} & \beta_{1,2} & :::: & \beta_{1,n}\\ \beta_{1,2} & \beta_{2,2} & :::: & \beta_{2,n}\\ :::: & :::: & :::: & ::::\\ \beta_{1,n} & \beta_{2,n} & :::: & \beta_{n,n}\end{bmatrix}\times\bar{L}\left\{\begin{bmatrix}\mu_{1,1} & \mu_{1,2} & :::: & \mu_{1,N}\\ \mu_{1,2} & \mu_{2,2} & :::: & \mu_{2,N}\\ :::: & :::: & :::: & ::::\\ \mu_{1,n} & \mu_{2,n} & :::: & \mu_{n,N}\end{bmatrix}\times\begin{bmatrix}v_1\\ v_2\\ v_3\\ ::::\\ v_N\end{bmatrix}\right\}\right\}=\begin{bmatrix}"c_1"\\ "c_2"\\ ::::\\ "c_n"\end{bmatrix}, \quad (3.8)$$

$$\bar{B}\left\{\begin{bmatrix}\beta_{1,1} & \beta_{1,2} & :::: & \beta_{1,n}\\ \beta_{1,2} & \beta_{2,2} & :::: & \beta_{2,n}\\ :::: & :::: & :::: & ::::\\ \beta_{1,n} & \beta_{2,n} & :::: & \beta_{n,n}\end{bmatrix}\times\bar{L}\left\{\begin{bmatrix}\mu_{1,1} & \mu_{1,2} & :::: & \mu_{1,N}\\ \mu_{1,2} & \mu_{2,2} & :::: & \mu_{2,N}\\ :::: & :::: & :::: & ::::\\ \mu_{1,n} & \mu_{2,n} & :::: & \mu_{n,N}\end{bmatrix}\times\begin{bmatrix}\xi_1\\ \xi_2\\ \xi_3\\ ::::\\ \xi_N\end{bmatrix}\right\}\right\}=\begin{bmatrix}"x_1"\\ "x_2"\\ ::::\\ "x_n"\end{bmatrix}, \quad (3.9)$$

где $\beta_{i,j}$ – весовые коэффициенты второго слоя нейронов; $\bar{B}$ ::: – вектор квантователей нейронов второго слоя.

Очевидно, что по аналогии с матричным описанием двухслойной сети нейронов мы можем сделать похожую запись с большим числом вложений для сети, содержащей 3, 4, 5,… слоев нейронов. Все подобные сети могут быть описаны аналогичными матричными континуально-квантовыми уравнениями. Во всех нейронных сетях биометрических приложений последний слой нейронов обязательно должен иметь идеальный (полный) квантователь с очень крутым фронтом переключения. Нейроны предшествующих слоев сети могут осуществлять неполное квантование биометрических данных. В частности, первый и второй слои нейронов трехслойной сети могут иметь нелинейные элементы с линейными участками [1]. Примеры нелинейных элементов с полным и частичным квантованием приведены на рис. 3.7.

Из данного рисунка видно, что при идеальном (100 %) квантовании все 100 % состояний распределения данных всего алфавита возможных образов «Все Чужие» будут иметь только два состояния, соответствующие «0» и «1». При 75 % квантования оставшиеся 25 % будут иметь неопределенное состояние, находящееся в интервале от нуля до единицы.

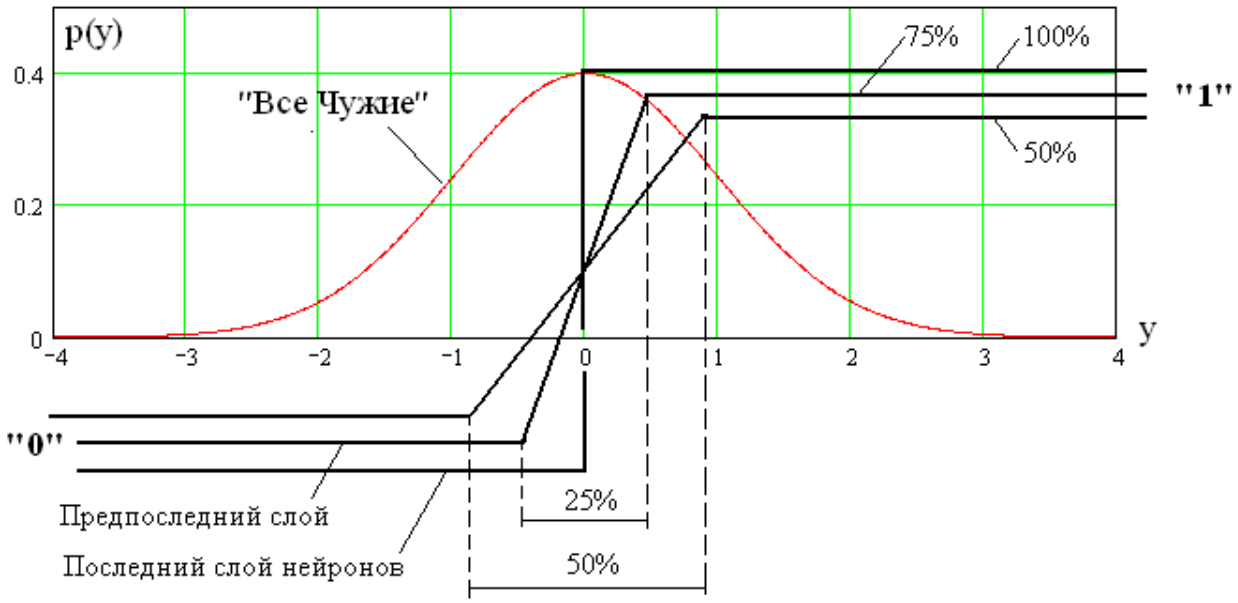


Рис. 3.7. Нелинейные функции нейронов, осуществляющие частичное (неполное) и полное (100 %) квантование выходов нейронов

В случае, если двухслойная нейронная сеть имеет в первом и втором слоях нейронов полные квантователи, обучение преобразователя осуществляется алгоритмами ГОСТ Р 52633.5–2011. Если квантователи у внутренних слоев неполные, то их обучение может быть осуществлено итерационными методами по критерию поиска максимума качества (3.4).

Следует подчеркнуть, что для преобразователей биометрия-код наиболее предпочтительными являются однослойные и двухслойные нейронные сети с полными квантователями данных в обоих слоях. Для таких нейронных сетей механизмы размножения биометрических ошибок достаточно хорошо отработаны (см. рис. 3.2), т.е. многослойные нейронные сети с неполными квантователями данных промежуточных нейронных сетей могут быть применены только в биометрических средствах аутентификации, размещенных в доверенной вычислительной среде (должна использоваться аппаратная защита доступа).

3.4. Выбор длины блоков шифрования, используемых при реализации механизма размножения ошибок биометрических данных

Важным вопросом является выбор длины блоков самошифрования (см. блок-схему рис. 3.2). Если блок окажется слишком длинным, например, 60 бит (как с алгоритмом DES), возможно организовать атаку обращения матриц нейросетевых функционалов алгоритмом генетического направленного подбора (алгоритм описан в параграфах 2.4, 2.5). Необходимо определить оптимальную длину блока самошифрования. Оптимизацию будем осуществлять через синтез номограммы, связывающей энтропию кодов разной длины со средним значением модулей коэффициентов корреляции между разрядами кодов.

Вполне очевидным является то, что полная независимость (некоррелированность) биометрических кодов «Чужой» дает последовательности кодов «белым шумом». Такие последовательности будут иметь энтропию, точно совпадающую с длиной биометрического кода.

В ином предельном случае, когда все разряды кодов абсолютно коррелированны, энтропия кодов «Чужой» оказывается единичной независимо от длины кодов:

$$\begin{cases} E(|r_{ij}|) = 1; \\ H(n) = 1. \end{cases} \qquad (3.10)$$

В общем случае энтропия выходных кодов «Чужой» описывается трехмерной функцией [1, 14]:

$$H(n, E(|r|), E(P_{2,i})) = \beta(n, E(|r|), E(P_{2,i}) \cdot (n-1) + 1) \cdot H(1), \qquad (3.11)$$

где $H(1)$ – среднее значение одномерной энтропии всех разрядов выходного кода; $P_{2,i}$ – вероятность ошибки второго рода для i-го разряда кода «Чужой».

Форма описания (3.11) удобна тем, что содержит почти мультипликативный трехмерный коэффициент связи многомерной энтропии и среднего значения одномерных частных энтропий по каждому из выходных разрядов биометрического кода. Для наиболее важного случая $E(P_{2,i}) = 0{,}5$ удалось построить номограмму двухмерного почти мультипликативного коэффициента связи [1]. Однако этот подход оказался не продуктивен, так как кривые подобных номограмм описываются очень сложными аналитическими зависимостями.

Возможен еще один путь описания многомерной энтропии, который строится, опираясь на так называемые многомерные корреляционные функции. Этот тип функций вводится через отношение многомерных функций энтропии:

$$R(n)=\left\{1-\frac{H(n)}{n\cdot H(1)}\right\}=\left\{1-\frac{H(n)}{\sum_{i=1}^{n}H_i(1)}\right\}, \qquad (3.12)$$

где $H_i(1)$ – одномерная энтропия i-го разряда.

Обычная двухмерная корреляция r совпадает с R(2). Для более высоких значений размерности r и R(n) оказываются связаны между собой номограммой, приведенной на рис. 3.8. Номограмма построена по данным, приведенным в табл. 3.1.

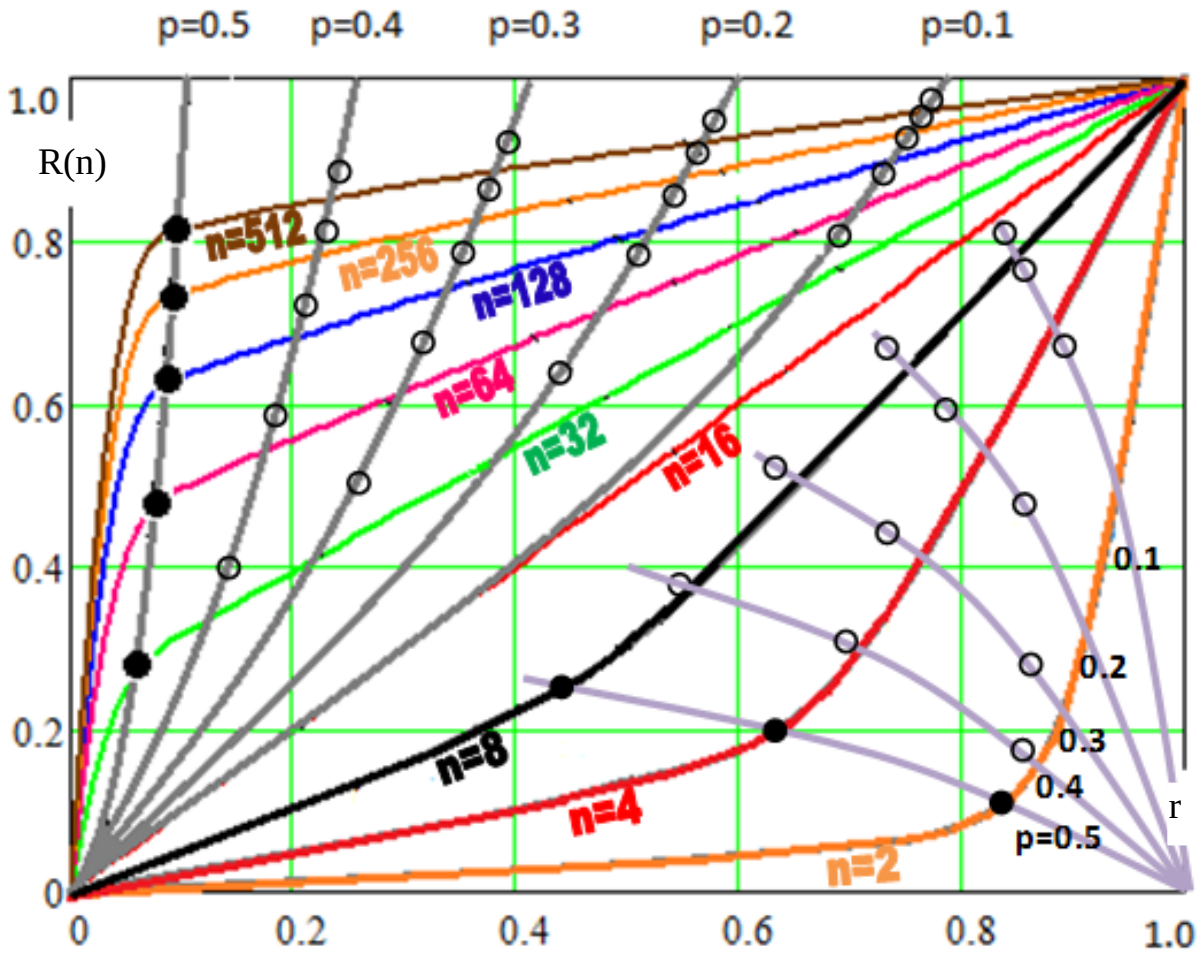


Рис. 3.8. Номограмма связи значений высокоразмерных корреляций R(n) с усредненными значениями модулей обычных корреляционных функций – r

Таблица 3.1

Значения кривых номограммы рис. 3.8 для параметра P = 0,5

Значения многомерной корреляции R(n)											
		Двухмерная корреляция r									
		0,00	0,10	0,20	0,30	0,40	0,50	0,60	0,70	0,80	0,90	1,00
Размерность n	2	0,00	0.01	0.02	0.03	0.04	0,05	0,06	0,07	0,09	0,30	1,00
	4	0,00	0.03	0.05	0.08	0,11	0,13	0,18	0,32	0,53	0,76	1,00
	8	0,00	0.06	0,10	0,15	0.23	0,33	0,43	0,57	0,71	0,86	1,00
	16	0,00	0,10	0,21	0,32	0,42	0,51	0,62	0,71	0,82	0,90	1,00
	32	0,00	0,32	0,39	0,47	0,55	0,62	0,70	0,77	0,85	0,92	1,00
	64	0,00	0,50	0,56	0,62	0,67	0,73	0,78	0,84	0,89	0,95	1,00
	128	0,00	0,63	0,68	0,72	0,77	0,81	0,84	0,88	0,92	0,96	1,00
	256	0,00	0,74	0,77	0,81	0,84	0,86	0,89	0,92	0,95	0,97	1,00
	512	0,00	0,81	0,84	0,87	0,89	0,91	0,93	0,95	0,96	0,98	1,00

Из данного рисунка видно, что кривые связи высокоразмерных корреляций R(n) с усредненными значениями модулей обычных корреляционных функций r имеют значительные почти линейные участки. Это связано со смысловой однородностью связываемых второй номограммой функций. Причина сложности номограммы энтропия-корреляция в том, что она связывает разнородные параметры [1, 14]. Номограмма рис. 3.8 упростилась из-за того, что связывает между собой однородные функции (функции корреляции разной размерности).

Из рис. 3.8 видно, что значительные участки кривых фактически образованы почти линейными функциями (кривые хорошо описываются кусочно-линейным приближением со сплайн-стыковкой линейных фрагментов). Крайне удобным является также то, что сплайн-сшивка линейных участков находится на параболе (точки отмечены темной заливкой на рис. 3.8). Для каждого из дополнительных условий P = 0,4, P = 0,3, P = 0,2, P = 0,1 получаются другие параболы и иные точки сплайн-сшивки. Эти данные приведены в табл. 3.2.

Таблица 3.2

Координаты точек сплайн-сшивки линейных участков, находящиеся на разных параболах

Координаты (r ; R(n))						
		Параметр p				
		0.5	0.4	0.3	0.2	0.1
Размерность n	2	(0.84;0.12)	(0.86;0.18)	(0.85;0.28)	(0.86;0.48)	(0.88;0.68)
	4	(0.63;0.21)	(0.71;0.32)	(0.73;0.43)	(0.78;0.61)	(0.86;0.77)
	8	(0.44;0.25)	(0.54;0.38)	(0.63;0.52)	(0.73;0.66)	(0.85;0.82)
	32	(0.05;0.28)	(0.14;0.40)	(0.26;0.50)	(0.43;0.63)	(0.68;0.80)
	64	(0.06;0.48)	(0.18;0.59)	(0.32;0.67)	(0.51;0.78)	(0.72;0.89)
	128	(0.07;0.63)	(0.22;0.71)	(0.35;0.78)	(0.54;0.84)	(0.75;0.92)
	256	(0.08;0.73)	(0.23;0.81)	(0.38;0.86)	(0.56;0.91)	(0.76;0.96)
	512	(0.09;0.82)	(0.24;0.89)	(0.39;0.92)	(0.57;0.95)	(0.77;0.98)

Следует отметить, что все координаты точек сшивки лежат на двух типах парабол. Для размерностей выше 16 параболы строятся из общей точки, расположенной в левом нижнем углу рис. 3.8 (точка R(n) = 0, r = 0). Для низких размерностей параболы имеют общую точку в правом нижнем углу рис. 3.8 (точка R(n) = 0, r = 1). Данные первой и второй парабол сведены в табл. 3.3.

Таблица 3.3

Параболы, описывающие положение точек сшивки линейных участков

Полиномы многомерной корреляции R(n)		
Параметр p	Двухмерная корреляция r	
	Размерность n=32, 64, 128, 256, 512	Размерность n=2, 4, 8
p=0.5	$56.19r^2 + 1.59r$	$-0.42r^2 + 0.15r + 0.27$
0.4	$5.86r^2 + 2.12r$	$-1.19r^2 + 0.98r + 0.20$
0.3	$2.54r^2 + 1.01r$	$-1.02r^2 - 0.11r + 1.13$
0.2	$1.47r^2 + 0.58r$	$-4.17r^2 + 4.51r - 0.33$
0.1	$0.83r^2 + 0.59r$	$-62.5r^2 + 112.5r - 50$

Из всего выше изложенного следует, что размерность n = 16 играет особую роль, она имеет линейную связь между 16-мерной и обычной двухмерной корреляцией:

$$R(16) = r\,. \qquad (3.13)$$

При размерностях до и после 16 энтропия и корреляции имеют связь: по-разному вогнутые (выпуклые). Размерность n = 16 – граница, разделяющая просто многомерные преобразования и высокоразмерные преобразования (точное значение границы n = 15,471).

Из всего выше сказанного следует, что максимальная сложность обращения матриц нейросетевых функционалов будет при выборе длины блока самошифрования до 16 бит. При малой длине блоков процедуры обращения матриц нейросетевых функционалов не работают. При длине более 16 появляется теоретическая возможность обращения матриц нейросетевых функционалов. Насколько длина блока шифрования может быть больше 16, покажут дальнейшие исследования.

4. СРЕДА МОДЕЛИРОВАНИЯ «БИОНЕЙРОАВТОГРАФ»

Следует отметить, что задавать корректно биометрические данные достаточно сложно. Если пользоваться некоторым программным генератором данных, то всегда возникает вопрос о том, насколько эти данные корректны. Для того, чтобы снять вопрос о корректности получаемых биометрических данных, была создана среда моделирования «БиоНейроАвтограф»; на данный момент это единственный общедоступный инструмент для получения корректных биометрических данных как в русскоязычном сегменте Интернет, так и англоязычном сегменте Интернет. К сожалению, фирмы-разработчики биометрических продуктов отказываются предоставлять открытый доступ к содержанию своих технических решений. То, что данные в среде моделирования «БиоНейроАвтограф» корректны, каждый может убедиться самостоятельно, выполнив ряд лабораторных работ, размещенных на сайте АО «ПНИЭИ». То, что биометрические данные этого продукта общедоступны, обеспечивается их открытым хранением в директории DATA, среде моделирования «БиоНейроАвтограф», где в процессе работы сохраняются файлы:

– 3sigma.txt (содержит информацию о факте попадания биометрических данных каждого из 416 параметров в его допустимый интервал $\pm 3\sigma$);

– mean.txt (содержит запись среднего значения каждого из 416 контролируемых биометрических параметров);

– mera.txt (содержит меры Хэмминга между кодом «trainKey» и кодом «trainKey», полученные при инициализации режима «проверить»);

– params.txt (содержит 416 биометрических параметров, учитываемых нейронной сетью);

– stdev.txt (содержит 416 стандартных отклонений биометрических параметров, учитываемых нейронной сетью);

– testKeys.txt (содержит данные о коде длиной 256 бит, полученном на выходах нейронной сети при предыдущих проверках);

– trainKey.txt (содержит данные о коде длиной 256 бит, использованном при обучении искусственной нейронной сети);

– weights.txt (содержит данные о 24 весах 256 нейронов преобразователя биометрия-код, полученных в результате последнего обучения).

Для получения среды моделирования зайдите на страницу http://пниэи.рф/activity/science/noc.htm, скачайте архив bioneuroautograph.zip. Распакуйте архив и запустите находящийся в папке файл БиоНейроАвтограф.exe. При этом появится основная экранная форма с фотографией административного здания АО «ПНИЭИ» (г. Пенза, улица Советская, 9).

4.1. Как задать пароль доступа или криптографический ключ

Среда моделирования «БиоНейроАвтограф» предназначена для нейросетевого связывания рукописного пароля с обычным паролем доступа (набираемым на клавиатуре) или криптографическим ключом. Пароль доступа может быть изменен по усмотрению студента. Для того, чтобы задать или изменить пароль, необходимо выбрать пункт меню "Режим", подпункт "Задать пароль" в левом верхнем углу основного диалогового окна (рис. 4.1) или одновременно нажать комбинацию клавиш "Ctrl+P".

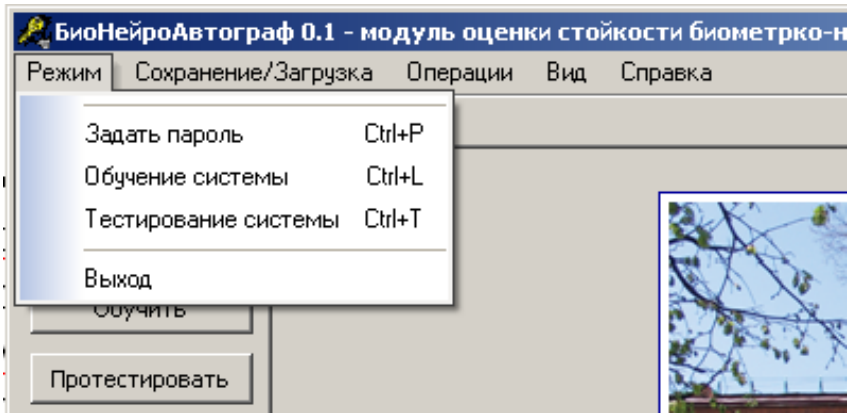


Рис. 4.1. Выбор пункта меню "Задать пароль"

При этом появится окно создания пользовательского пароля с двумя полями ввода (рис. 4.2). В верхнем поле введите имя пользователя (свой логин), в нижнем введите свой пароль, состоящий, например, из 32 символов "a" в латинской кодировке.

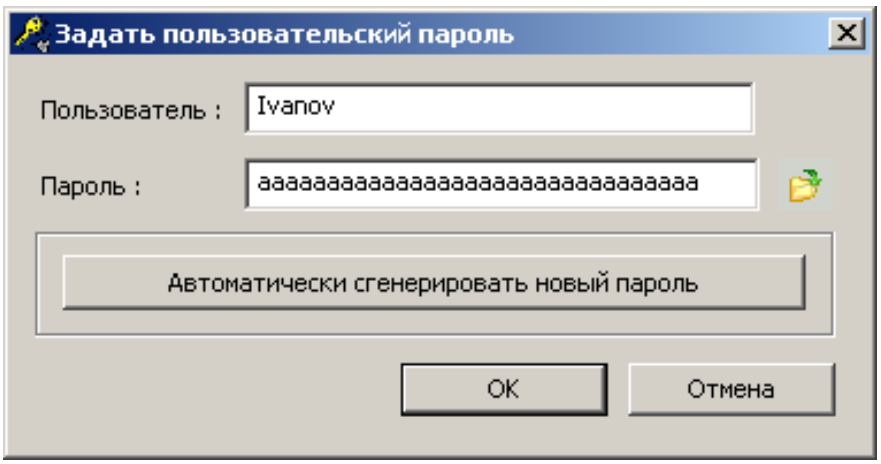


Рис. 4.2. Диалоговое окно создания пароля

Для создания длинного случайного пароля нажмите кнопку "Автоматически сгенерировать новый пароль". Сгенерированный 32-символьный пароль при обучении преобразуется в обучающий ключ длиной 256 бит (32 случайных символа в 8-битной кодировке). Сохранение введенного имени пользователя и пароля происходит после нажатия кнопки "ОК". В случае успешного сохранения данных можно приступать к обучению нейронной сети.

4.2. Как обучить нейронную сеть

Обучение нейронной сети осуществляется в режиме обучения (рис. 4.3), который вызывается с помощью нажатия кнопки "Обучить" основного меню, либо выбором пункта меню "Режим", подпункта "Обучение системы", либо одновременным нажатием комбинации клавиш "Ctrl+L".

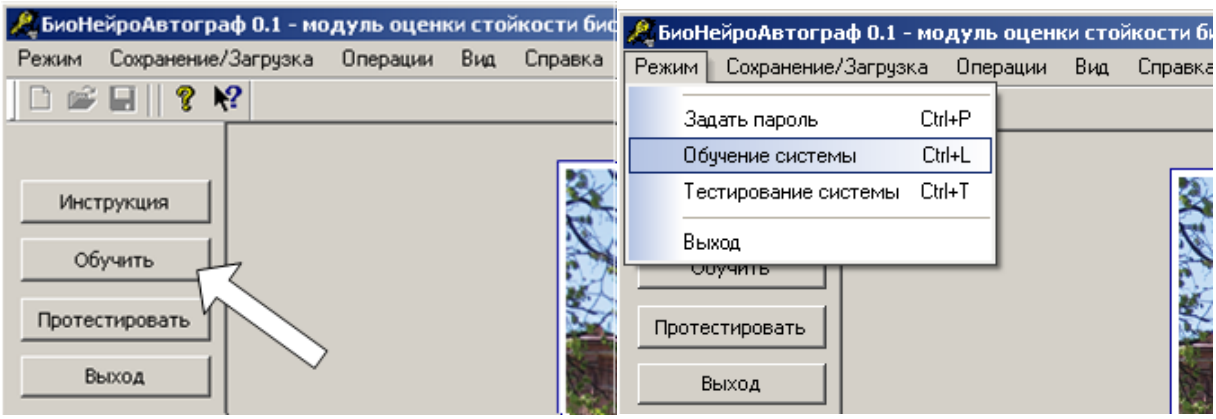


Рис. 4.3. Вызов режима обучения

Для того, чтобы обучить нейронную сеть, необходимо ввести несколько обучающих примеров той или иной рукописной буквы или рукописного слова. Если пользоваться манипулятором "мышь", то лучше вводить примеры отдельных букв либо короткие слова, так как писать с помощью "мышки" достаточно сложно. Если имеется графический планшет, то для обучения необходимо вводить рукописное слово, состоящее из трех и более букв. При выборе обучающего символа или слова необходимо знать, что чем длиннее вводимое слово и чем выше стабильность его написания, тем выше стойкость обученной сети к атакам подбора. Так как с помощью манипулятора "мышь" вводить длинные стабильные слова невозможно, то качество обучения будет низким, а вероятности появления ошибок первого и второго рода высокими.

После входа в режим обучения введите с помощью манипулятора "мышь" или графического планшета несколько примеров выбранного для обучения рукописного слова или рукописной буквы (см. рис. 4.4).

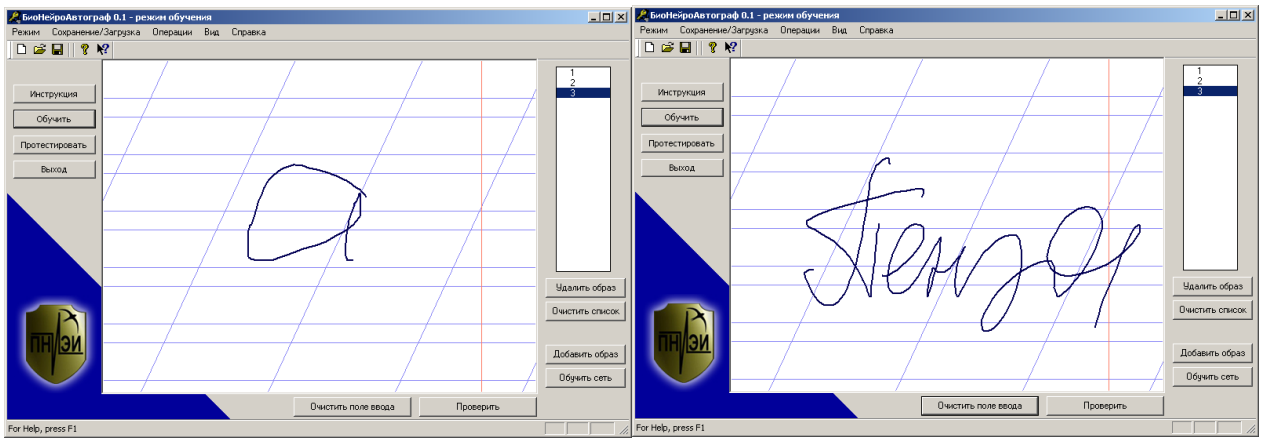


Рис. 4.4. Диалоговое окно обучения нейронной сети

После каждого ввода примера обучающего слова необходимо добавить введенный пример в список обучающих примеров (список в правом верхнем углу диалогового окна обучения). Добавление примера осуществляется нажатием кнопки "Добавить образ" в правом нижнем углу диалогового окна.

Все ранее введенные примеры могут быть просмотрены щелчком манипулятора "мышь" по соответствующему номеру примера в списке обучающих примеров. Если добавленный в список пример неудачен (например, дрогнула рука), то его можно удалить. Для этого необходимо выбрать неудачный пример в списке обучающих примеров и нажать кнопку "Удалить образ". Чтобы удалить все обучающие примеры из списка, нажмите кнопку "Очистить список".

Удаление текущего введенного рукописного образа и очистка поля ввода осуществляются с помощью кнопки "Очистить поле ввода".

После добавления в список обучающих примеров трех или более примеров рукописного образа запустите процесс обучения нейронной сети, нажав кнопку "Обучить сеть". Обучение сети из 256 нейронов длится менее 1 с. По окончании обучения появляется окно с результатами обучения (рис. 4.5), содержащее информацию о предполагаемой стабильности введенных биометрических примеров, вероятности узнавания образа "Свой" и ожидаемой стойкости системы к атакам подбора обучающего рукописного слова-пароля, т.е. вероятность пропуска образа "Чужой".

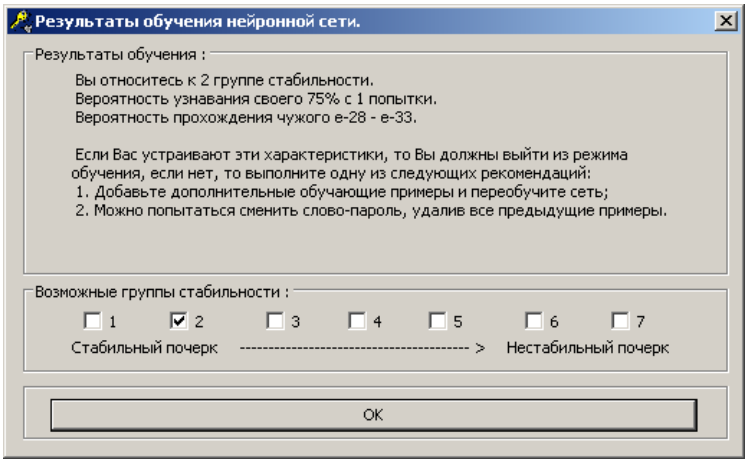


Рис. 4.5. Окно с результатами обучения

Можно попытаться изменить качество обучения, увеличивая или уменьшая число примеров обучения. Более точную оценку качества обучения можно дать только после тестирования искусственной нейронной сети. Доверие к результатам обучения, отображаемым в окне с результатами обучения, низкое, так как эти данные рассчитывались по обучающей выборке.

4.3. Как проверить обученную нейронную сеть

Для того, чтобы получить достоверную оценку качества обучения, необходимо в поле ввода рукописных образов ввести пример рукописного образа "Свой" и нажать кнопку "Проверить". При этом вычисляются параметры введенного рукописного образа, они подаются на входы искусственной нейронной сети, вычисляется выходной код и выводится окно с результатами сравнения полученного выходного кода с обучающим кодом (рис. 4.6).

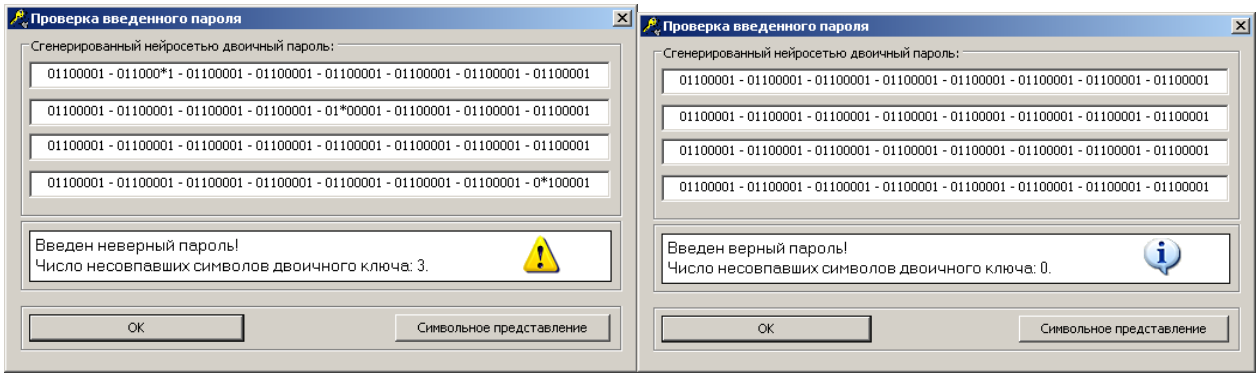


Рис. 4.6. Окно вывода полученного кода на примерах образа "Свой"

Сеть хорошо узнает образ "Свой", если мера Хэмминга на тестовых примерах равна нулю (все разряды полученного кода совпадают с заданным при обучении кодом "Свой"). Если несколько бит кодов не совпадают, то обучение нужно продолжить, добавив в обучающую выборку дополнительные примеры рукописного образа "Свой". Хорошо обученная нейронная сеть должна с высокой вероятностью узнавать образ "Свой".

Для проверки способности нейронной сети отказывать в доступе образам "Чужой" необходимо воспроизвести случайное рукописное слово (букву). При этом появляется случайный код, совпадающий с кодом "Свой" в случайных разрядах. В окне вывода полученного кода совпавшие разряды отображаются верными состояниями "0" и "1", а несовпавшие разряды отображаются символом "*" (рис. 4.7).

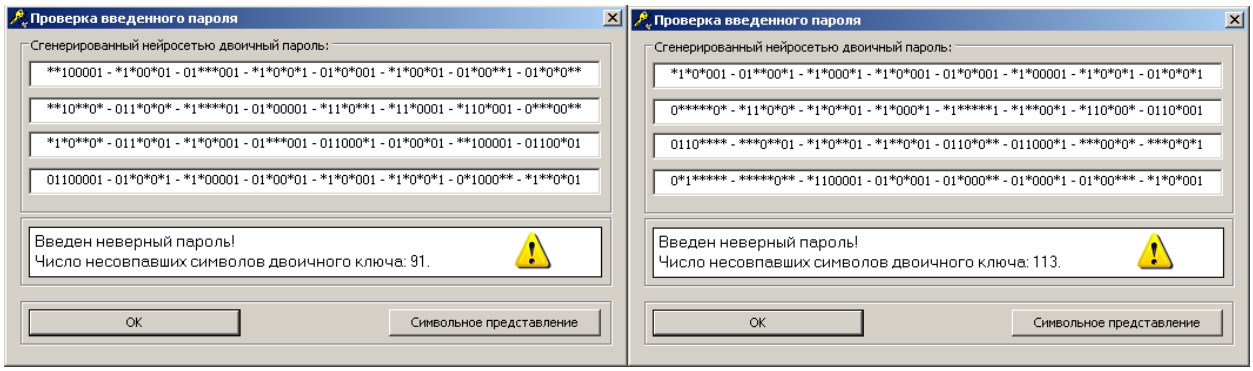


Рис. 4.7. Окно вывода полученного кода на примерах образа "Чужой"

Даже в том случае, когда воспроизводится один и тот же случайный рукописный образ "Чужой", выходные коды нейросети должны быть случайными (отличающиеся состояния должны располагаться в разных разрядах кодов).

Стойкость обученной нейронной сети тем выше, чем ближе число отличающихся разрядов кода к величине 128, так как для действительно случайных состояний выходного кода "Чужой" с наибольшей вероятностью угадывает примерно половину из 256 разрядов кода.

4.4. Как сохранить и загрузить биометрические образы

Для того, чтобы надежно тестировать качество работы обученного преобразователя биометрия-код, нужно создавать специальные базы тестовых образов "Свой" и "Чужой" по требованиям ГОСТ Р 52633.1–2009. Средство моделирования большой нейронной сети "БиоНейроАвтограф" имеет встроенные средства, позволяющие собирать базы биометрических образов.

Сохранение обучающей выборки примеров "Свой" осуществляется путем выбора пункта меню "Сохранение/Загрузка", подпункта "Сохранить образы на диск" либо одновременным нажатием комбинации клавиш "Ctrl+S" (рис. 4.8).

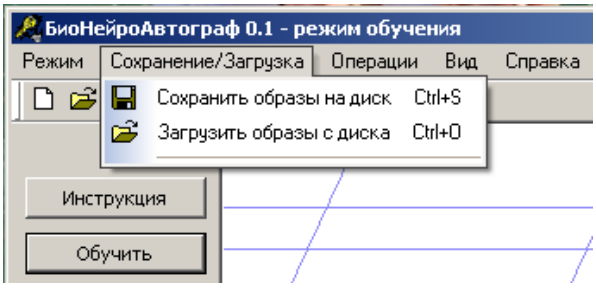


Рис. 4.8. Пункт меню сохранения и загрузки образов

Далее в открывшемся диалоговом окне укажите каталог, в котором будет сохранен файл, и задайте имя файла (по умолчанию задано имя "MyImages.dat"). Рекомендуется примеры, на которых производилось обучение, сохранять с именем "Обучение__.dat", а примеры для тестирования обозначать именами "Тестирование_СВОЙ_.dat" или "Тестирование_ЧУЖИЕ_.dat".

После успешного сохранения можно удалить все обучающие примеры.

Сохраненные ранее примеры рукописных образов всегда можно загрузить и повторно обучить нейронную сеть. Загрузка рукописных образов осуществляется путем выбора пункта меню "Сохранение/Загрузка", подпункта "Загрузить образы с диска" либо одновременным нажатием комбинации клавиш "Ctrl+O" (см. рис. 4.8).

В появившемся диалоговом окне выберите требуемый файл с образами и нажмите кнопку "Открыть": загруженные примеры автоматически добавляются в список обучающих примеров.

Режим сохранения и загрузки рукописных образов крайне важен для формирования больших баз биометрических образов. У людей существует порог "комфортности" требований к ним со стороны биометрических автоматов. Мы легко пишем подряд десяток рукописных слов, однако требование воспроизвести подряд 20 одинаковых слов уже воспринимается людьми как некоторое обременение. Требование воспроизвести своею рукой 200 одинаковых слов людьми воспринимается как существенное обременение (нужно написать страницу рукописного текста). В связи с этим сохранение образов и обмен базами "Все Чужие" – это мера, позволяющая существенно снизить трудоемкость лабораторных работ.

Следует иметь в виду, что режим "Сохранение/Загрузка" есть только у учебных средств моделирования искусственных нейронных сетей. Реальные средства биометрико-нейросетевой аутентификации должны уничтожать данные обучения и тестирования по требованиям ГОСТ Р 52633.0–2006.

4.5. Специальные режимы работы

4.5.1. Режим, воспроизводящий биометрическую аутентификацию

Среда моделирования "БиоНейроАвтограф" ориентирована на студентов, занимающихся изучением применения нейросетевого искусственного интеллекта к приложениям защиты информации. В связи с этим в этой среде реализован режим, имитирующий меню программного средства биометрической аутентификации. Попасть в этот режим можно через основное меню, нажав кнопку "Протестировать", либо выбрав пункт меню "Режим", подпункт "Тестирование системы", либо одновременным нажатием комбинации клавиш "Ctrl+T". Главное диалоговое окно тестирования работы нейронной сети в режиме аутентификации пользователей представлено на рис. 4.9.

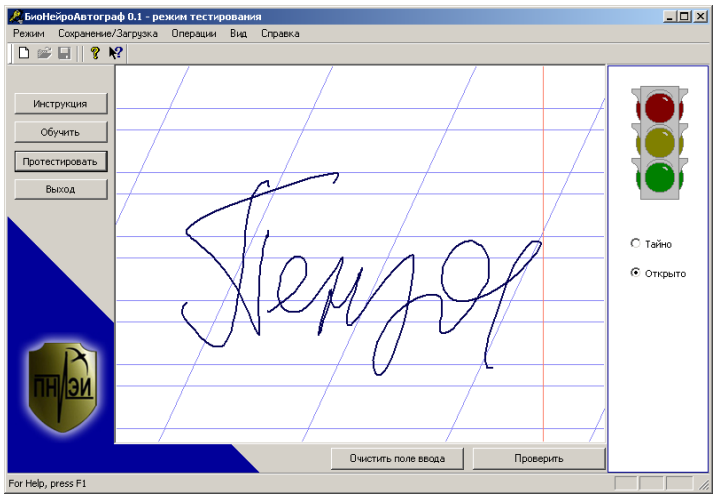


Рис. 4.9. Диалоговое окно режима тестирования

В правом верхнем углу диалогового окна расположен светофор с тремя состояниями: красный, желтый и зеленый. В случае положительного результата аутентификации загорается зеленый свет светофора. Если введенный биометрический образ близок к эталонному образу "Свой", загорается желтый свет светофора. Если введенный биометри-

ческий образ далек от эталонного образа "Свой", загорается красный свет. Светофор помогает пользователю "Свой" ориентироваться в текущем состоянии протокола биометрико-криптографической аутентификации. Индикатор состояния, выполненный в форме светофора, безопасен, так как выполнен в соответствии с требованиями ГОСТ Р 52633.6.

Особенно важен светофор в режиме "Тайно", когда пользователь аутентифицируется по рукописному слову-паролю, которое не отображается на экране видеомонитора. Режим "Тайно" применяется пользователем в ситуации присутствия рядом с ним посторонних лиц. Режим "Открыто" применяется пользователем, когда он один и никто не может подсмотреть его тайный рукописный пароль.

Основное отличие режима аутентификации от режимов обучения и тестирования состоит в том, что в этом режиме ключ "Свой" неизвестен, следовательно, невозможно вычислить количество отличающихся бит ключа и показать их позиции. Так как в режиме аутентификации сравниваются не ключи/пароли, а хеш-код полученного ключа с хеш-кодом эталонного, то отображается только сигнал светофора и выводится сообщение с результатами проверки введенного пароля (рис. 4.10, 4.11).

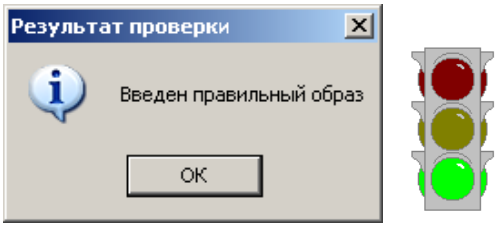


Рис. 4.10. Аутентификация пройдена

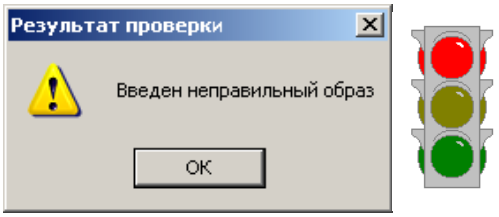


Рис. 4.11. Аутентификация не пройдена

Желтый сигнал светофора выдается вместе с сообщением "Введен неправильный образ".

4.5.2. Режим автоматического тестирования на базе тестовых образов

В случае, когда тестовая база создана заранее, можно воспользоваться специальным режимом тестирования на образах из базы. Для это-

го необходимо выбрать пункт меню "Операции", подпункт "Тестировать на тестовых образах" (рис. 4.12).

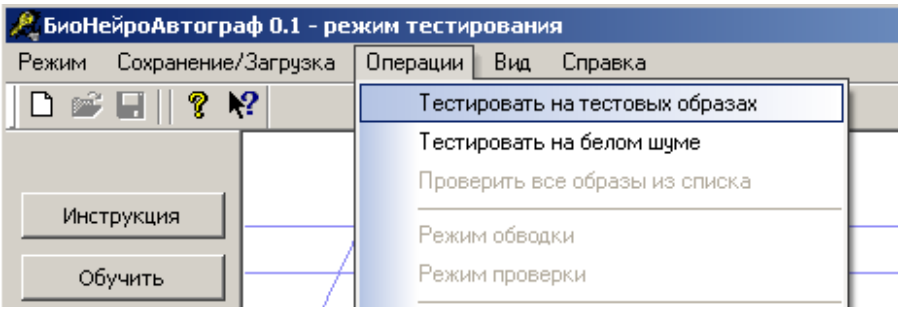


Рис. 4.12. Запуск тестирования на реальных образах

В результате появляется диалоговое окно выбора каталога, в котором хранятся файлы с тестовыми рукописными образами. После выбора нужного файла и нажатия кнопки "Открыть" запускается процесс тестирования обученной нейронной сети на выбранных образах. После завершения процесса тестирования выводится окно с результатами проверки (рис. 4.13).

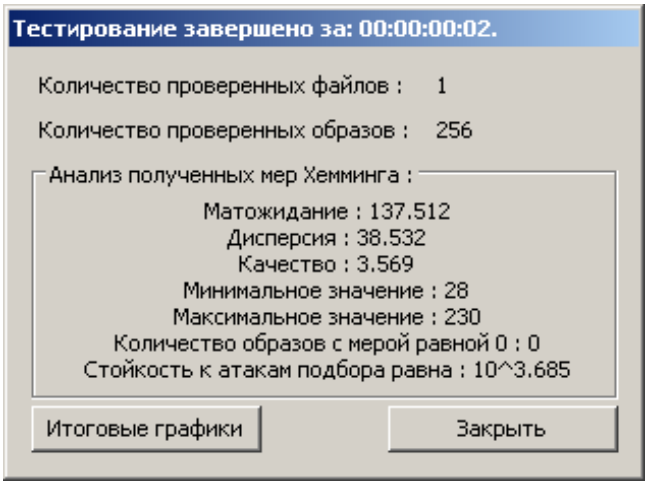


Рис. 4.13. Окно с результатами тестирования

Окно результатов содержит информацию о количестве использовавшихся во время тестирования примеров реальных рукописных образов, математическое ожидание, среднеквадратическое отклонение, качество, минимальное и максимальное значения полученных мер Хэмминга. Также отображаются количество "взломов" системы (количество образов с нулевой мерой) и вычисленная на тестовой базе стойкость к атакам подбора. Все вычисляемые во время тестирования меры Хэмминга записываются в файл Data/<Имя_пользователя>/mera.txt. Каталог Data находится рядом с запускаемым файлом БиоНейроАвто-

граф.exe. После каждого тестирования происходит перезаписывание данных в файле mera.txt.

Примечание. В каталоге Data также хранятся весовые коэффициенты обученной нейронной сети в файле weights.txt, а в файле coefs.txt хранятся коэффициенты двухмерного преобразования Фурье последнего поданного на нейронную сеть примера биометрического образа.

4.5.3. Режим автоматического тестирования на «белом шуме»

Если нет корректно собранной базы биометрических образов, то тестирование может быть выполнено на данных, получаемых от генератора случайного шума, т.е. тестирование на искусственно синтезированных образах. Для запуска тестирования на «белом шуме» необходимо выбрать пункт меню "Операции", подпункт "Тестировать на белом шуме" (рис. 4.14).

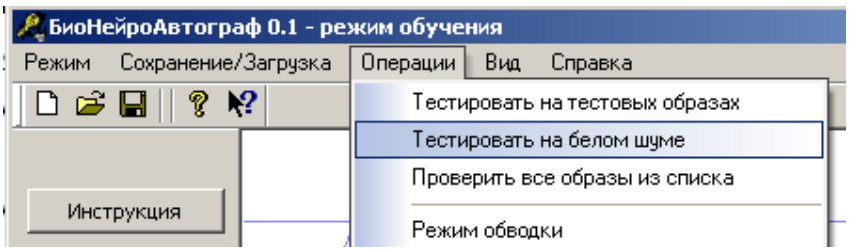


Рис. 4.14. Запуск тестирования на синтезированных образах

Тест запускается автоматически. Тестирование обученной нейронной сети осуществляется на 1 000 сгенерированных примерах рукописных образов. После завершения процесса тестирования выводится окно с результатами проверки (рис. 4.15).

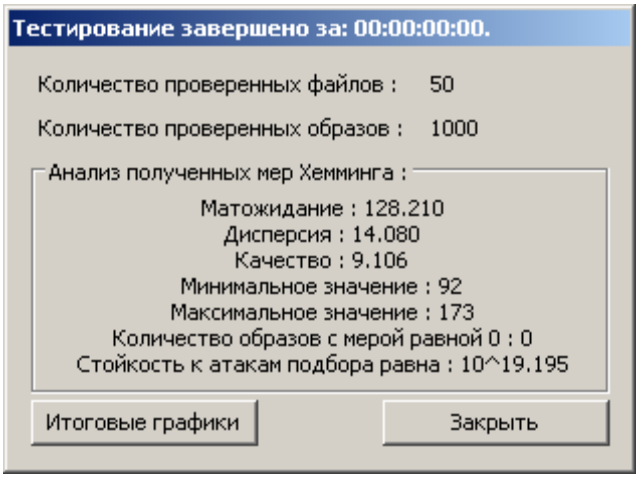


Рис. 4.15. Окно с результатами тестирования

Окно результатов содержит информацию о количестве использовавшихся во время тестирования примеров реальных рукописных образов, математическое ожидание, среднеквадратическое отклонение, качество, минимальное и максимальное значения полученных мер Хэмминга. Также отображаются количество "взломов" системы (количество образов с нулевой мерой) и вычисленная на тестовой базе стойкость к атакам подбора. Все вычисляемые во время тестирования меры Хэмминга записываются в файл Data/<Имя_пользователя>/mera.txt. Каталог Data находится рядом с запускаемым файлом БиоНейроАвтограф.exe. После каждого тестирования происходит перезаписывание данных в файле mera.txt.

4.5.4. Режим проверки примеров из списка образов

Быстрая проверка всех примеров из списка обучающих примеров осуществляется выбором подпункта "Проверить все образы из списка" пункта меню "Операции". При этом все примеры из списка последовательно подаются на обученную нейронную сеть, вычисляется выходной код и сравнивается с эталонным. На экран выводится отчет об общем количестве проверенных примеров и количестве примеров, распознанных как "Свой" и "Чужой". Данный режим подходит для быстрой оценки качества обучения нейронной сети, позволяет увидеть, все ли обучающие примеры правильно распознаются как "Свой". Если есть тестовые примеры образа "Свой", то можно увидеть, какова ошибка первого рода, т.е. какой процент примеров правильно распознан как "Свой" и неправильно отнесен в группу "Чужой".

Также в режиме обучения можно активировать режим быстрой проверки примеров из списка обучающих примеров. Для активации режима проверки необходимо выбрать пункт меню "Операции", подпункт "Режим проверки" (рис. 4.16). Деактивация осуществляется аналогично.

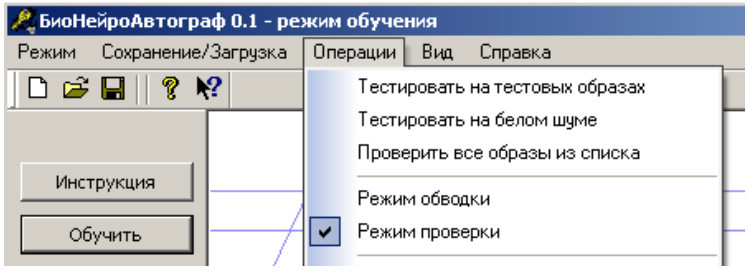


Рис. 4.16. Активация режима проверки

После активации в левом нижнем углу поля ввода рукописных образов появится красная надпись "Включен режим проверки". Теперь, ес-

ли выбрать пример из списка, он будет автоматически подаваться на обученную нейронную сеть, вычисляться выходной код и сравниваться с эталонным, полученная мера Хэмминга будет выводиться в верхнем левом углу поля ввода (рис. 4.17).

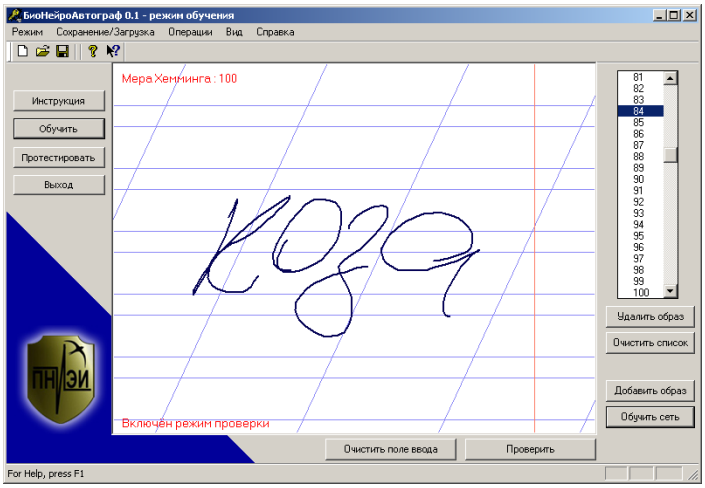


Рис. 4.17. Вычисление меры Хэмминга образов в режиме проверки

Данный режим позволяет быстро вычислить меры Хэмминга на всех примерах из списка, увидеть близкие и далекие к обучающему примеру и оценить качество обучения нейронной сети.

4.6. Завершение работы

Окончание работы среды моделирования "БиоНейроАвтограф" осуществляется нажатием крестика в верхнем правом углу главного диалогового окна. При этом появляется диалоговое окно с предложением завершить или продолжить работу (рис. 4.18).

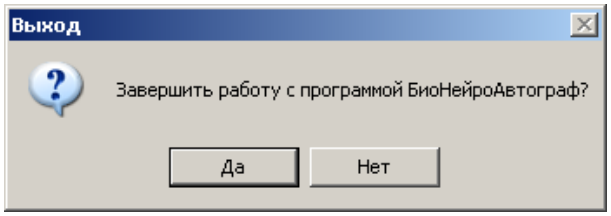


Рис. 4.18. Выход из программы

Для окончания работы необходимо нажать кнопку "Да".

Перед завершением работы необходимо сохранить обучающие примеры (если это необходимо). Все несохраненные данные после завершения работы приложения автоматически удаляются.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

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

2. Feng, Hao. Crypto with Biometrics Effectively / Hao Feng, Ross Anderson, John Daugman // IEEE TRANSACTIONS ON COMPUTERS. – 2006. – Vol. 55, № 9 (September).

3. Dodis, Y. Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy / Y. Dodis, L. Reyzin, A. Smith // EUROCRYPT. – 2004. – April 13. – P. 523–540.

4. Monrose, F. Cryptographic key generation from voice / F. Monrose, M. Reiter, Q. Li, S. Wetzel // IEEE Symp. on Security and Privacy. – 2001. – URL: dblp.uni–trier.de>db/conf/sp/index.html

5. Juels, A. A Fuzzy Commitment Scheme / A. Juels, M. Wattenberg // Proc. ACM Conf. Computer and Communications Security. – 1999. – P. 28–36. – URL: dl.acm.org

6. Ушмаев, О. В. Алгоритмы защищенной верификации на основе бинарного представления топологии отпечатка пальцев / О. В. Ушмаев, В. В. Кузнецов // Информатика и ее применения. – 2012. – № 6 (1). – С. 132–140.

7. Чморра, А. Л. Маскировка ключа с помощью биометрии / А. Л. Чморра // Проблемы передачи информации. – 2011. – № 2 (47). – С. 128–143.

8. Морелос-Сарагоса, Р. Искусство помехоустойчивого кодирования / Р. Морелос-Сарагоса. – Москва : Техносфера, 2007. – 320 с.

9. Питерсон, У. Коды, исправляющие ошибки : монография / У. Питерсон, Э. Уэлдон ; под ред. Р. Л. Добрушиной, С. И. Самойленко. – Москва : МИР, 1976. – 364 с.

10. Самойленко, С. И. Помехоустойчивое кодирование / С. И. Самойленко. – Москва : Радио и связь, 1968. – 240 с.

11. Боос, В. Лекции по математике. Перебор и эффективные алгоритмы / В. Боос. – Москва : Изд-во ЛКИ, 2012. – Т. 10. – 216 с.

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

13. Рутковская, Д. Нейронные сети, генетические алгоритмы и нечеткие системы / Д. Рутковская, М. Пилиньский, Л. Рутковский ; пер. с польск. И. Д. Рудинского. – Москва : Горячая линия – Телеком, 2004. – 452 с.

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

Учебное издание

Иванов Александр Иванович

Среда моделирования
«БиоНейроАвтограф»

Редактор *Т. Н. Судовчихина*
Технический редактор *Ю. В. Анурова*
Компьютерная верстка *Ю. В. Ануровой*

Подписано в печать 22.12.2020.
Формат 60×84 1/16. Усл. печ. л. 3,49.
Заказ № 509. Тираж 50.

Издательство ПГУ
440026, Пенза, Красная, 40
Тел.: (8412) 66-60-49, 66-67-77; e-mail: iic@pnzgu.ru


ПензГУ