REPORTS OF NATIONAL ACADEMY OF SCIENCES OF THE REPUBLIC OF KAZAKHSTAN
ISSN 2224-5227
Volume 5, Number 5 (2014), 49 – 60

UDC 681.32 2

ASSESSMENT OF COMPUTING COMPLEXITY OF THE ADDRESS OF MATRIXES OF NEURONETWORK FUNCTIONALITIES

B.S.Akhmetov[1], A.I. Ivanov[2], A.V. Bezyaev[3], S.V. Kachalin[4]
b_akhmetov@ntu.kz, ivan@pniei.penza.ru
[1]Kazakh national technical university named after K.I.Satpayev, Almaty
[2]Penza scientific-research electrotechnical institute, Russia
[3]FSUE «SRC «Atlas» Penza branch
[4]«RPE «Rubin» JSC, Penza.

Keywords: artificial neural networks, high dimension, address of matrixes of neuronetwork functionalities.

Abstract. It is shown that the problem of recognition of biometric images of the person within linear algebra is stubborn because of dimension "damnation". It is possible to bypass computing difficulties by training of big artificial neural networks which are described by nonlinear algebra of matrixes of neuronetwork functionalities. In considered algebra direct procedure of training of big neural networks has linear computing complexity, and the return procedure of the address of matrixes of neuronetwork functionalities has polynomial computing complexity. It is necessary to apply special measures for providing high level of protection of biometric data "blinding" the observer of high-dimensional entropy.

In work it is shown that the problem of recognition of biometric images of the person within linear algebra is stubborn because of dimension "damnation". It is possible to bypass computing difficulties by training of big artificial neural networks which are described by nonlinear algebra of matrixes of neuronetwork functionalities.

УДК 681.32 2

ОЦЕНКА ВЫЧИСЛИТЕЛЬНОЙ СЛОЖНОСТИ ОБРАЩЕНИЯ МАТРИЦ НЕЙРОСЕТЕВЫХ ФУНКЦИОНАЛОВ

Б.С.Ахметов[1], А.И. Иванов[2], А.В. Безяев[3], С.В. Качалин[4]

[1]Казахский национальный технический университет имени К.И. Сатпаева, г. Алматы
[2]Пензенский научно-исследовательский электротехнический институт, Россия
[3]Пензенский филиал ФГУП «НТЦ «Атлас»
[4]ОАО «НПП «Рубин» г. Пенза.

Ключевые слова: искусственные нейронные сети, высокая размерность, обращение матриц нейросетевых функционалов

Аннотация. Показано, что задача распознавания биометрических образов человека в рамках линейной алгебры трудноразрешима из-за «проклятия» размерности. Обойти вычислительные трудности удается путем обучения больших искусственных нейронных сетей, которые описываются нелинейной алгеброй матриц нейросетевых функционалов. В рассматриваемой алгебре прямая процедура обучения больших нейронных сетей имеет линейную вычислительную сложность, а обратная процедура обращения матриц нейросетевых функционалов имеет полиномиальную вычислительную сложность. Для обеспечения высокого уровня защиты биометрических данных необходимо применять специальные меры «ослепляющие» наблюдателя высокоразмерной энтропии.

[1] Статья подготовлена в рамках выполнения проекта «Исследование вариантов реализации и разработка действующего лабораторного образца ON-LINE системы биометрического обезличивания электронных историй болезней для медицинского учреждения» в соответствии с Приказом Председателя Комитета науки МОН РК №17-нж от 08.04.2013 г

Введение. В настоящее время во всем мире идут процессы создания средств биометрической защиты прав личности. Россия и Казахстан идут по пути использования больших искусственных нейронных сетей [1, 2, 3, 4] и стандартизации этой технологии. Англоязычные исследователи [5, ..., 19] идут по пути применения так называемых «нечетких экстракторов», являющихся частным случаем нейросетевых преобразователей.

Различие между «нечеткими экстракторами» и нейросетевыми преобразователями обусловлено их структурами, приведенными на рисунке 1.

«Нечеткие экстракторы» квантуют «сырые» биометрические данные, получая тем самым био-код с большим числом ошибок. Нейронные сети обогащают биометрические данные на сумматорах нейронов и уже после их обогащения квантуют эти данные, выдавая на выходах нейронной сети био-код почти не содержащий ошибок. Можно говорить о том, что «нечеткий экстрактор» - это вырожденная нейронная сеть, состоящая из нейронов с одним входом.

Так как «нечеткие экстракторы» дают «плохой» био-код с 30% нестабильных бит, в нем следует исправить ошибки. Для этого «нечеткие экстракторы» используют классические самокорректирующиеся коды с очень большой избыточностью. Так Даугман [11] преобразует рисунок радужной оболочки глаза в био-код длинной 2048 бит, для корректировки ошибок он использует самокорректирующийся код с 15-ти кратной избыточностью. В итоге Даугман [11] получает скорректированных выходной био-код длинной 128 бит, что вполне достаточно для применения последующих криптографических преобразований, защищающих личность.

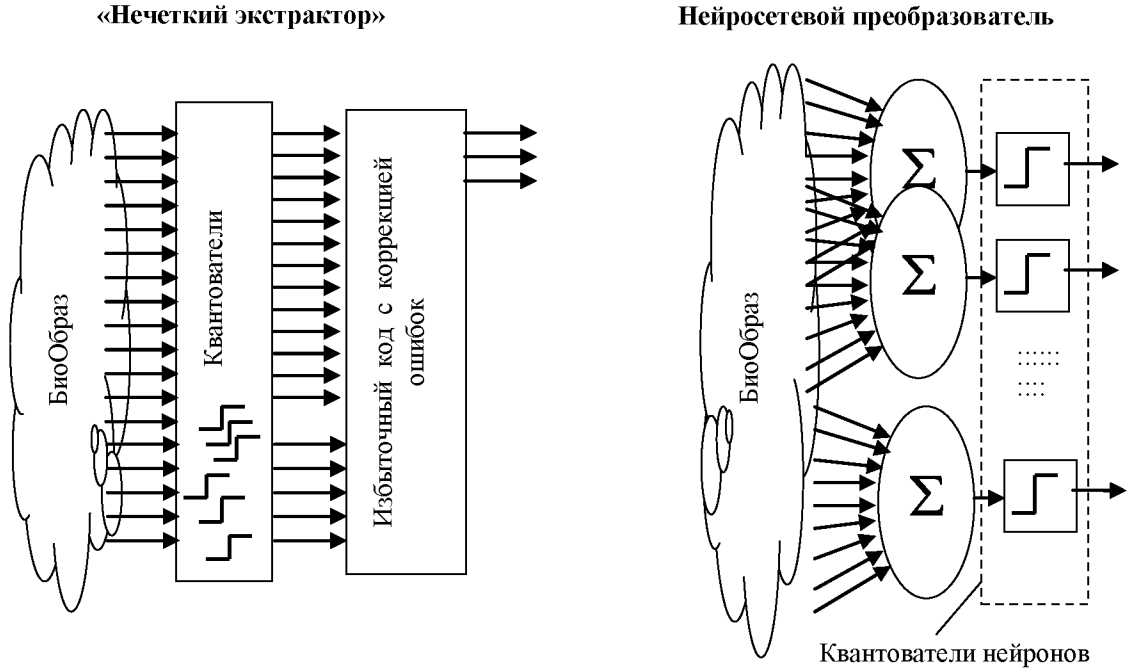

Рисунок 1 - Структурные схемы нечетких экстракторов и нейросетевых преобразователей

Совершенно иная ситуация возникает для рисунков отпечатков пальца [10, 13, 19]. Рисунки отпечатков пальца содержат от 20 до 40 особых точек, если каждую из особых точек (окончание папиллярных линий, ветвление папиллярных линий) описывать двумя координатами, то удается получить био-код длинной от 40 до 80 бит. Последующее применение самокорректирующегося кода с 15-ти кратной избыточностью приводит к хорошему конечному био-коду длинной от 2 до 4 бит. Столь короткий код нельзя применять для криптографической защиты личности человека. Перебор 128 бит неизвестного био-кода является сложной задачей, перебор 4 бит неизвестного био-кода является очень простой задачей.

В отличии от «нечетких экстракторов» нейросетевые преобразователи биометрия-код способны эффективно защищать личность человека при любом качестве биометрических данных.

Они так же обладают определенной избыточностью, например, преобразователи динамики рукописной подписи в код длинной 256 [20] имеют 416 входов. То есть число входов у нейросетевых преобразователей оказывается примерно в 2 раза больше чем выходов у нейронной сети. Это связано с тем, что энтропия каждого из биометрических параметров, как правило, оказывается меньше единицы. Однако суммарная энтропия всех 416 биометрических параметров оказывается достаточно велика и может обеспечить последующую криптографическую защиту персональных данных личности.

В связи с тем, что нейросетевые преобразователи биометрия-код оказались способны эффективно защищать личность человека идет процесс стандартизации нейросетевых технологий обработки биометрических данных. В таблице 1 приведены названия и номера уже созданных биометрических стандартов.

Таблица 1 – Биометрические стандарты

№	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 52633.7-20xx** «Защита информации. Техника защиты информации. Высоконадежная мультибиометрическая аутентификация»

Процесс стандартизации «нечетких экстракторов» не идет в силу их ограниченных возможностей и неспособности эффективно защищать биометрические данные человека.

Проблема плохой обусловленности матричных преобразований линейной алгебры

В ограниченных возможностях линейной алгебры может убедиться каждый самостоятельно, воспользовавшись средой моделирования «БиоНейроАвтограф» [20]. Эта среда моделирования преобразует рукописный пароль или рукописную букву в вектор из 416 биометрических параметров $\overline{\nu}$. Данные могут вводиться через графический планшет или манипулятором «мышь». Для обучения нейросетевого преобразователя обычно требуется предъявить 20 примеров образа «Свой». По этим данным среда моделирования вычисляет вектор математических ожиданий $E(\nu_i)$ по каждому из биометрических параметров и вектор стандартных отклонений $\sigma(\nu_i)$ по каждому i-тому биометрическому параметру. Далее по алгоритму ГОСТ Р 52633.5 вычисляются весовые коэффициенты нейронов, обучаемой искусственной нейронной сети. При этом все 416 биометрических параметра ν_i размещаются в файле weights.txt и могут быть использованы для организации матричной обработки. В частности для распознавания биометрического образа «Свой» может быть использована квадратичная форма:

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

где $[\rho]$ - матрица коэффициентов ковариации.

Проблема вычислений квадратичных форм (1) состоит в том, что необходимо осуществлять обращение матриц коэффициентов ковариации. Если в обучающей выборке содержится 20 примеров образа «Свой», то относительная ошибка при вычислении коэффициентов ковариации

может составлять од 30%. То есть число обусловленности обращаемой матрицы не должно быть более 3. На практике же для матриц 3 порядка число обусловленности достигает 200 и более. Это означает, что формирование сети квадратичных форм (1) обобщающих тройки биометрических параметров на обучающей выборке из 20 примеров невозможно. Для того, что бы ковариационные матрицы третьего порядка надежно обращались 20 примеров недостаточно, требуется привлекать от 200 2000 примеров образа «Свой».

Заставить обычного человека написать свой рукописный пароль 200 раз трудно. Люди не хотят прилагать столь значительные усилия для обучения сети квадратичных форм. Мы сталкиваемся с так называемым «проклятием» размерности, когда попытка незначительного увеличения размерности приводит к экспоненциальному росту объемов исходных данных.

Решение проблемы «проклятия» размерности использованием больших искусственных нейронных сетей

Следует отметить, что n-мерная квадратичная форма (1) описывает объем некоторого гиперэллипса n-мерного распределения данных «Свой». Сечение этого гипперэллипса по любым двум биометрическим параметрам дает обычный эллипс на плоскости. Эта ситуация отображена в правой части рисунка 2. Попадание внутрь объема гиперэллипса означает принадлежность данных образу «Свой». Попадание вне данных гиперэллипса соответствует ситуации, принадлежности биометрических данных образам «Чужие». .

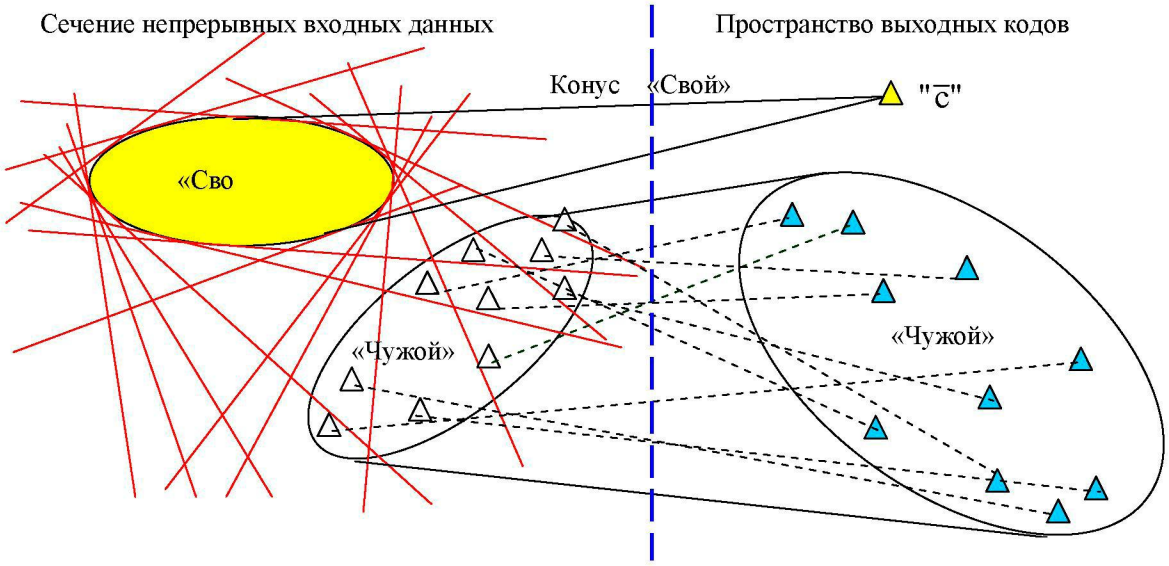

Рисунок 2 - Одно из двухмерных сечений гиперэллипсов «Свой» и «Чужой», эллипс «Свой» сворачивается в точку кода "$\bar{c}$"

При использовании большой искусственной нейронной сети возникает иная ситуация. Во время обучения нейронной сети многомерное пространство делится гиперплоскостями пополам. Например, эти гиперплоскости могут строиться как касательные к гиперэллипсу данных «Свой». В этом случае выходной код любого примера образа «Свой» будет постоянен (будет иметь нулевую энтропию). Нейросеть, обученная распознавать образ «Свой» будет описываться некоторым гиперконусом, опирающимся на многомерный гиперэллипс непрерывных данных «Свой» и имеющим единственный код "$\bar{c}$" в вершине выходных состояний.

Для образа «Чужой» возникает иная ситуация. Образы «Чужой», находящиеся вне объема гиперэллипса «Свой» оказываются многократно рассечены разделяющими многомерное пространство гиперплоскостями. В частности в правой части рисунка 2 отображены проекции 7 гиперплоскостей, разделяющих эллипс «Чужой» на 9 фрагментов. Каждому из этих 9 фрагментов будет соответствовать свое выходное состояние выходных кодов нейросети "$\bar{x}_i$". Практика

показывает, что каждый i-тый пример образа «Чужой» порождает свой выходной код "$\overline{x}_i$", значительно отличающийся от кода, порожденного другим примером того же биометрического образа «Чужой». Возникает эффект хеширования биометрических данных образа «Чужой». Нейросетевой преобразователь биометрия-код ведет себя совершенно по разному для биометрических данных «Свой» и «Чужой». Он полностью устраняет естественную энтропию данных «Свой» и усиливает неопределенность (естественную энтропию) данных «Чужой».

Матричная запись уравнений нейросетевых функционалов

Следует подчеркнуть, что автоматическое обучение нейросетевого преобразователя биометрия-код алгоритмом по ГОСТ Р 52633.5 всегда является устойчивым и имеет линейную вычислительную сложность. Для того, что бы воспользоваться этим алгоритмом нужно заранее задать требуемый код доступа "$\overline{c}$" и иметь 20 примеров образа «Свой» и 200 примеров образов «Чужие». Итогом обучения будут являться весовые коэффициенты сумматоров нейронов и значения порогов бинарных квантователей. Формально вектор биометрических данных образа «Свой» оказывается связан с состояниями выходного кода следующей системой нелинейных уравнений:

$$\overline{Z}\left\{\begin{bmatrix}\mu_{1,1} & \mu_{1,2} & :::: & \mu_{1,N}\\ \mu_{2,1} & \mu_{2,2} & :::: & \mu_{2,N}\\ :::: & :::: & :::: & ::::\\ \mu_{n,1} & \mu_{n,2} & :::: & \mu_{n,N}\end{bmatrix}\times\begin{bmatrix}v_{1,i}\\ v_{2,i}\\ v_{3,i}\\ ::::\\ v_{N,i}\end{bmatrix}+\begin{bmatrix}b_1\\ b_2\\ ::::\\ b_n\end{bmatrix}\right\}=\begin{bmatrix}"c_1"\\ "c_2"\\ ::::\\ "c_n"\end{bmatrix} \qquad (2),$$

где $\overline{Z}\{.\}$ - вектор операторов квантования, дающих на выходе состояние «1» для положительных входных воздействий и состояние «0» для отрицательных входных воздействий, $\mu_{i,j}$ - весовые коэффициенты сумматоров обученных нейронов, $\overline{b}$ - вектор настроек порогов бинарных квантователей нейронов, "$\overline{c}$" - бинарный код доступа «Свой».

Для системы (2) естественная многомерная энтропия (неопределенность) входных биометрических данных всегда много больше почти нулевой выходной энтропии:

$$H(v_1, v_2, ..., v_N) \rangle\rangle H("c_1, c_2, ..., c_n") \approx 0 \qquad (3).$$

Нейросетевой преобразователь для данных образа «Чужой» $\xi_1, \xi_2, ..., \xi_N$, осуществляет функцию хеширования информации. Происходит увеличение естественной энтропии образа «Чужой»:

$$H(\xi_1, \xi_2, ..., \xi_N) \langle\langle H("x_1, x_2, ..., x_n") \approx \frac{n}{10} \qquad (4).$$

Приходится записывать систему нелинейных уравнений (1) в несколько иной форме, акцентируя внимание на то, что каждый пример образа «Чужой» дает свои состояния разрядов выходного кода:

$$\overline{Z}\left\{\begin{bmatrix}\mu_{1,1} & \mu_{1,2} & :::: & \mu_{1,N}\\ \mu_{2,1} & \mu_{2,2} & :::: & \mu_{2,N}\\ :::: & :::: & :::: & ::::\\ \mu_{n,1} & \mu_{n,2} & :::: & \mu_{n,N}\end{bmatrix}\times\begin{bmatrix}\xi_{1,i}\\ \xi_{2,i}\\ \xi_{3,i}\\ ::::\\ \xi_{N,i}\end{bmatrix}+\begin{bmatrix}b_1\\ b_2\\ ::::\\ b_n\end{bmatrix}\right\}=\begin{bmatrix}"x_{1,i}"\\ "x_{2,i}"\\ ::::\\ "x_{n,i}"\end{bmatrix} \qquad (5).$$

Если же мы каждый биометрический параметр будем квантовать «сырым», сделав N = n, то матрицы весовых коэффициентов в (2) и (5) вырождаются, а сами уравнения будут соответствовать «нечетким экстракторам» без последующей дополнительной обработки кодом с обнаружением и исправлением ошибок. Именно это и дает нам право рассматривать «нечеткие экстракторы» как частный случай нейросетевых преобразователей с вырожденными нейронами, имеющими всего по одному входу.

Обращение матриц нейросетевых функционалов (восстановление неизвестных входных биометрических данных)

Уметь осуществлять прямое нейросетевое преобразование биометрии в код, недостаточно. Полная нелинейная алгебра матриц нейросетевых функционалов возникает только тогда, когда мы умеем решать обратную задачу восстанавливая по коду "$\overline{c}$" вектор данных примеров образа «Свой». Утилитарный смысл процедуры обращения состоит в оценке стойкости нейросетевого преобразователя к атакам подбора. Процедура обращения матриц нейросетевых функционалов положена в основу стандарта ГОСТ Р 52633.3, регламентирующего то, как тестировать полученное при обучении нейросетевой решение.

Проблема тестирования качества найденных нейросетевых решений является далеко не тривиальной. Так, если производитель утверждает, что вероятность ошибок второго рода (ошибочное принятие чужого за своего) составляет величину 0.000000001, то для проверки этого утверждения потребуется тестовая база из 1000000000 случайных биометрических образов «Чужой». Собрать такую большую тестовую базу биометрических образов «Чужой» крайне сложно, если следовать рекомендациям ГОСТ Р 52633.1. В связи с этим тестирование по ГОСТ Р 52633.3 следует осуществлять переходя из обычного пространства выходных кодов в пространство расстояний Хэмминга между кодом «Свой» и кодами «Чужой»:

$$h = \sum_{i=1}^{n} ("x_i") \oplus ("c_i") \qquad (6),$$

где $\oplus$ - операция сложения разрядов кодов по модулю два.

Оказалось, что для хорошо обученных нейросетевых преобразователей биометрия-код, плотность распределения расстояний Хэмминга (6) хорошо описывается нормальным законом распределения значений. Эта ситуация отображена на рисунке 3 и многократно проверена численными экспериментами.

Поясним содержание рисунка 3 исходя из предположения, что у нас имеется тестовая база из 1000 образов «Чужой». (N_0=1000). И нам следует оценить стойкость к атакам подбора нейросетевого преобразователя с 256 выходами. Для этого мы подадим все образы «Чужой» на тестируемый преобразователь и рассчитаем распределение расстояний Хэмминга (5). При этом математическое ожидание расстояний Хэмминга должно оказаться близким к 128 битам (E(h)≈128), так как случайные образы «Чужой» для правильно обученного нейросетевого преобразователя биометрия-код, угадывают разряды био-кода с вероятностью 0.5 (это одно из основных требований базового стандарта ГОСТ Р 52633.0).

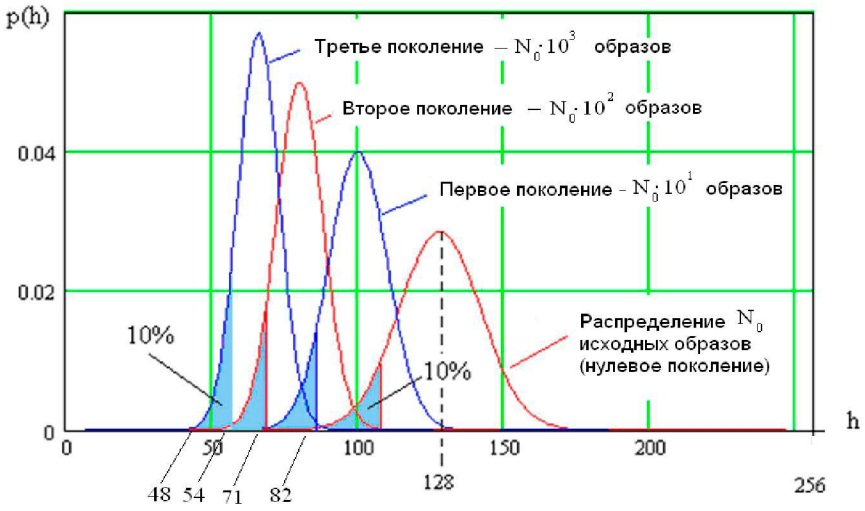

Рисунок 3 - Нормальные распределения расстояний Хэмминга для нулевого и последующих поколений тестовых биометрических образов «Чужой».

Из рисунка 3 видно, что в выборке из 1000 расстояний Хэмминга минимальным оказывается расстояние в 82 бита, то есть $3 \cdot \sigma(h) \approx 128 - 82 = 46$ или стандартное отклонение составит $\sigma(h) \approx 15.3$ бита. Так как мы знаем, что закон распределения нормален, а его параметры составляют E(h)≈128, $\sigma(h) \approx 15.3$, становится возможна оценка вероятности появления ошибок второго рода (ошибочной коллизии, когда «Чужой» угадывает код «Свой»). Подобная коллизия наступает когда псевдо непрерывное расстояние Хэмминга попадает в интервал от 0 до 1:

$$P_2 \approx \frac{1}{\sigma(h)\sqrt{2\pi}} \int_0^1 \exp\left\{ -\frac{(E(h)-u)^2}{2 \cdot (\sigma(h))^2} \right\} \cdot du \qquad (7).$$

В свою очередь знание вероятности ошибок второго рода (6) позволяет нам оценить многомерную энтропию зависимых био-кодов «Чужие»:

$$H("x_1, x_2, \ldots, x_{256}") \approx -\log_2(P_2) \qquad (8).$$

Проверить приближенные оценки (7) и (8) удается в том случае, когда мы начинаем применять генетические алгоритмы. Для этой цели нужно выбрать 10% образов «Чужой», которые наиболее близки образам «Свой» в метрике расстояний Хэмминга. На рисунке 2 положение выбранных биометрических образов отмечено заливкой. Мы поступаем совершенно так же, как обычные селекционеры [21], задавшись направлением создаваемой нами искусственной эволюции, планомерно сдвигающей последующие поколения в сторону увеличения похожести образов-потомков на образ «Свой» в метрике Хэмминга.

Для того, чтобы восстановить численность следующего поколения необходимо воспользоваться скрещиванием выбранных образов-родителей из нулевого поколения. Процедура скрещивания образов-родителей регламентируется ГОСТ Р 52633.2. Для получения образов-потомков прежде всего необходимо задать число потомков, получаемых от каждой пары образов-родителей, а так же необходимо указать степень похожести потомков на родителей. Выберем получение одного потомка от пары родителей и потребуем равной похожести образа-потомка на обоих родителей. Тогда каждый биометрический параметр образов-потомков получается простым усреднением параметров образов-родителей:

$$\psi_{i,(j,k)} = \frac{\xi_{i,j} + \xi_{i,k}}{2} \qquad (9),$$

где $\psi_{i,(j,k)}$ - значения i-го биометрического параметра образов-потомков, полученные усреднением i-тых биометрических параметров образа-родителя-j и образа-родителя-k.

Для восстановления численности популяции после ее генетической селекции потребуется

скрещивание всех 100 выбранных образов-родителей с 10 случайно выбранными другими образами из той же сотни. Повторять процедуру селекции и скрещивания можно многократно. При этом происходит смещение распределений расстояний Хэмминга в сторону точки h=0, как это показано на рисунке 3. Каждое следующее поколение будет иметь многомерное распределение биометрических данных, все ближе и ближе подходящее к распределению данных образа «Свой».

Для рассматриваемого нами случая генерация третьего поколения образов-потомков, не имеющих в своем составе образа, соответствующего точке h=0, эквивалентна доказательству того факта, что стойкость тестируемого нейросетевого преобразователя биометрия-код выше 1 000 000 попыток. Обнаружение еще трех поколений даст оценку стойкости к атакам подбора на уровне одного миллиарда попыток. При этом способе тестирования нет необходимости заранее создавать тестовую базу, состоящую из миллиарда биометрических образов. Экономия состоит в том, что при тестировании создаются не все возможные синтетические биометрические образы «Чужой», а только те которые похожи на образ «Свой» (эволюционируют в сторону похожести на образ «Свой»).

Для нас так же важно то, что описанная выше процедура тестирования фактически является процедурой обращения матриц нейросетевых функционалов (2), (5), приводящая к компрометации биометрических данных образа «Свой» до 96% на пятом, шестом поколениях генетической селекции. Фактически речь идет о создании алгоритма решения обратной задачи нейросетевой биометрии, имеющем полиномиальную вычислительную сложность. Фактически мы получили еще один вариант алгоритма для решения задач линейного программирования помимо уже известных вариантов алгоритмов Хачияна и Кармаркара [22], имеющих доказанную полиномиальную вычислительную сложность. В нашем случае полиномиальная вычислительная сложность алгоритма (тестирования, компрометации, обращения матриц, восстановления входных данных) обусловлена тем, что в пространстве расстояний Хэмминга мы видим куда следует направлять эволюцию.

Интересно отметить, что задача линейного программирования и задача обращения матриц нейросетевых функционалов однослойной сети полностью совпадают. Если же мы будем решать задачу обращения нейросетевых функционалов, порождаемых двухслойной нейронной сетью, то мы получаем инструмент для решения задач нелинейного программирования. В этом состоит значительное преимущество рассматриваемого в данной статье подхода в сравнении с процедурами, предложенными ранее Хачияном и Кармаркаром [22].

Ослепление наблюдателей высокоразмерной энтропии

Нейронные сети требуют их обучения. Во время обучения [1, 2, 3] устанавливаются связи между нейронами подбираются весовые коэффициенты сумматоров нейронов. В конечном итоге после обучения по ГОСТ Р 52633.5 формируются таблица 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 био-кода нейросетевого преобразователя используется для шифрования части еще не использованных данных нейросетевого контейнера. Блок-схема защиты нейросетевого контейнера самошифрованием приведена на рисунке 4.

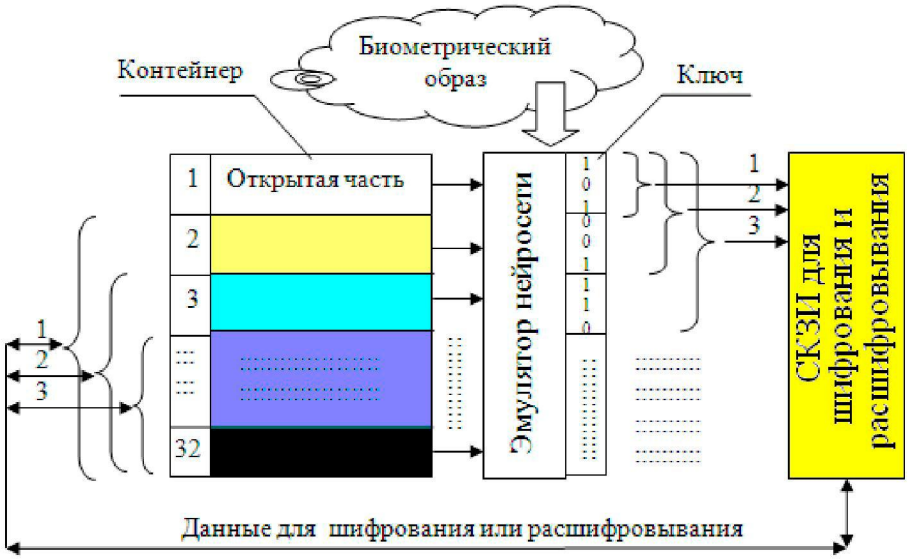

Рисунок 4 - Защита нейросетевого контейнера самошифрованием

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 так же показатели стабильности и коррелированности разрядов био-кодов. 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 хеширования 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.

Хеширование данных нейросети после их неверного расшифровывания приводит к тому, что дисперсия распределения расстояний Хэмминга сжимается, как это показано на рисунке 5.

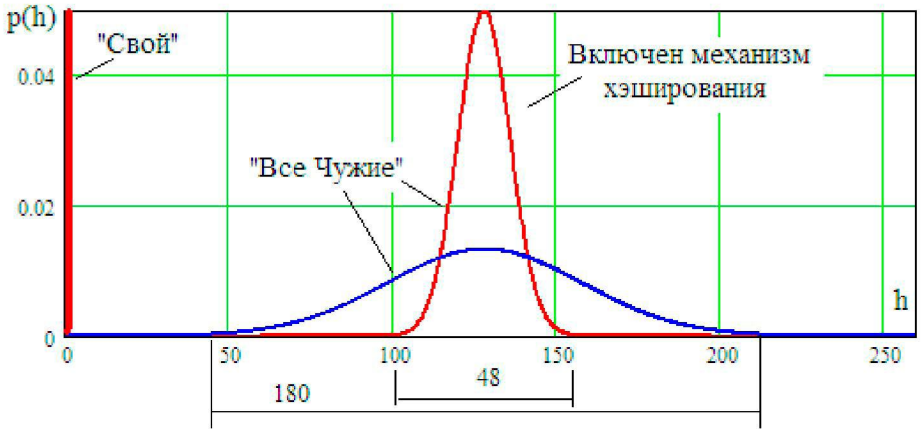

Рисунок 5 - Эффект сжатия распределения расстояний Хэмминга между кодом «Свой» и кодами «Все Чужие» после включения механизма размножения ошибок

Из рисунка 5 видно, что при включении механизма размножения ошибок (хеширования данных) происходит примерно трехкратное сжатие среднеквадратического отклонения распределений расстояний Хэмминга. Нейросеть без хеширования имеет σ(h)=30 битам, нейросеть с включенным механизмом хеширования имеет σ(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).

Подстановка данных распределений рисунка 5 в формулу 7 дает для не защищенного нейросетевого контейнера $P_2=10^{-5}$, что соответствует энтропии био-кода 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 длинной 256 разрядов. То есть включение механизма размножения ошибок препятствует наблюдение реальных статистик распределений расстояний Хэмминга кодов «Все Чужие». В место реального распределение расстояний Хэмминга мы видим идеальное распределение, соответствующее идеальной криптографической защите. Механизм размножения ошибок образов «Чужой» (механизм перемешивания данных) фактически ослепляет наблюдателя высокоразмерной энтропии. Наблюдатель высокоразмерной энтропии (8) не может видеть куда следует двигаться. Задача обращения матриц нейросетевых функционалов в этом случае имеет экспоненциальную вычислительную сложность.

Заключение

В данной статье мы попытались показать, что некоторые привычные ограничения исчезают если переходить от обычной линейной алгебры к алгебре нейросетевых функционалов. Кажется, что переходя из непрерывных пространств линейной алгебры в континуально-квантовые пространства алгебры нейросетевых функционалов мы теряем точность и должны получить многократные усложнения вычислений. На самом деле это не так алгебра нейросетевых функционалов обладает полнотой (существуют прямое преобразование – обучение и обратное преобразование – обращение матриц). Более того, то что называется «проклятием» размерности в линейной алгебре превращается в его инверсию - «благодать» высоких и сверхвысоких размерностей нейросетевых преобразований. Чем больше данных учитывает искусственная нейронная сеть, тем эффективнее она работает и тем легче выполнять обращение матриц ее нейросетевых функционалов. В биометрии приходится применять специальные меры ослепляющие высокоразмерных наблюдателей энтропии возвращающие экспоненциальную вычислительную сложность процедурам обращения.

Видимо «благодать» высокоразмерных нейросетевых преобразований является принципиально важным свойством естественного и искусственного интеллекта. Далеко не любое нелинейное искажение многомерных пространств полезно и дает положительный эффект. В данной работе мы попытались показать, что многомерные нейросетевые преобразования как раз

относятся к очевидно полезным нелинейным преобразованиям многомерных пространств. Чем выше размерность нейросетевых преобразований тем они полезнее.

ЛИТЕРАТУРА

[1] *Волчихин В.И., Иванов А.И., Фунтиков В.А.* Быстрые алгоритмы обучения нейросетевых механизмов биометрико-криптографической защиты информации. Монография. Пенза: Издательство Пензенского государственного университета, 2005г., 273 с.

[2] *Язов Ю.К. и др.* Нейросетевая защита персональных биометрических данных. //Ю.К. Язов (редактор и автор), соавторы В.И. Волчихин, А.И. Иванов, В.А. Фунтиков, И.Г. Назаров // М.: Радиотехника, 2012 г. 157 с.

[3] *Ахметов Б.С., Иванов А.И., Фунтиков В.А., Безяев А.В., Малыгина Е.А.* Технология использования больших нейронных сетей для преобразования нечетких биометрических данных в код ключа доступа. Монография, Казахстан, г. Алматы, ТОО «Издательство LEM», 2014 г. -144 с., свободный доступ http://portal.kaznu.kz/files/publicate/2014-06-27-11940.pdf

[4] *Ахметов Б.С., Волчихин В.И., Иванов А.И., Малыгин А.Ю.* Алгоритмы тестирования биометрико-нейросетевых механизмов защиты информации Казахстан, Алматы, КазНТУ им. Сатпаева, 2013 г.- 152 с. ISBN 978-101-228-586-4, http://portal.kaznu.kz/files/publicate/2014-01-04-11940.pdf

[5] *Juels A., Wattenberg M. A.* Fuzzy Commitment Scheme // Proc. ACM Conf. Computer and Communications Security, 1999, p. 28–36

[6] *F. Monrose, M. Reiter, Q. Li, S. Wetzel.* Cryptographic key generation from voice. In Proc. IEEE Symp. on Security and Privacy, 2001

[7] *Juels A., Sudan M. A Fuzzy Vault Scheme* // IEEE International Symposium on Information Theory, 2002

[8] *Y. Dodis, L. Reyzin, A. Smith*. Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy, Data April 13, In EUROCRYPT, pages 523-540, 2004.

[9] *Yang S., Verbauwhede I.* Automatic Secure Fingerprint Verification System Based on Fuzzy Vault Scheme // Proc. IEEE ICASSP 2005, p.609-612

[10] *Ramirez-Ruiz J., Pfeiffer C., Nolazco-Flores J.* Cryptographic Keys Generation Using FingerCodes. //Advances in Artificial Intelligence - IBERAMIA-SBIA 2006 (LNCS 4140), p. 178-187, 2006

[11] *Feng Hao, Ross Anderson, and John Daugman.* Crypto with Biometrics Effectively, IEEE TRANSACTIONS ON COMPUTERS, VOL. 55, NO. 9, SEPTEMBER 2006.

[12] *Cauchie S., Brouard T., Cardot H.* From features extraction to strong security in mobile environment: A new hybrid system. //On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops, Springer, pp. 489-498, 2006

[13] *Arakala A., Jeffers J., Horadam K.J.* Fuzzy Extractors for Minutiae-Based Fingerprint Authentication. //Advances in Biometrics (LNCS 4642), Springer, pp. 760-769, 2007

[14] *Lee Y.J., Bae K., Lee S.J., Park K.R., Kim J.* Biometric Key Binding: Fuzzy Vault Based on Iris Images. // Proceedings of 2nd International Conference on Biometrics, p. 800–808, Seoul, South Korea, August 2007

[15] *Nandakumar K., Jain A.K., Pankanti S.* Fingerprint-Based Fuzzy Vault: Implementation and Performance. //IEEE Transactions on Information Forensics and Security 2(4), pp. 744–757, 2007

[16] *Balakirsky V.B., Ghazaryan A.R., Han Vinck A.J.* Constructing Passwords from Biometrical Data. //Advances in Biometrics (LNCS 5558), Springer, pp. 889-898, 2009

[17] *Kanade S., Petrovska-Delacretaz D., Dorizzi B.* Multi-Biometrics Based Cryptographic Key Regeneration Scheme. //Proceedings of the 3rd IEEE international conference on Biometrics: Theory, applications and systems, p. 333-339, 2009

[18] *Чморра А.Л.* Маскировка ключа с помощью биометрии «Проблемы передачи информации» 2011 № 2(47) с. 128-143.

[19] *Ушмаев О.В., Кузнецов В.В.* Алгоритмы защищенной верификации на основе бинарного представления топологии отпечатка пальцев. //Информатика и ее применения. 2012 г. №6(1), с. 132-140.

[20] Среда моделирования «БиоНейроАвтограф» размещена на сайте ОАО «Пензенский научно-исследовательский электротехнический институт» http://пниэи.рф/activity/science/noc.htm. Продукт создан лабораторией биометрических и нейросетевых технологий ОАО «ПНИЭИ» для свободного распространения среди университетов России, Белоруссии, Казахстана.

[21] *Инге-Вечтомов С. Г.* Генетика с основами селекции. М., Высшая школа, 1989, 592 с.

[22] *Бооc В.* Лекции по математике. Том 10. Перебор и эффективные алгоритмы: Учебное пособие. – М.: Издательство ЛКИ, 2012 г. – 216 с.

REFERENCES

[1] *Volchikhin V.I., Ivanov A.I., Funtikov V.A.* Bystrye algoritmy obucheniya nejrosetevykh mekhanizmov biometriko-kriptograficheskoj zashhity informatsii. Monografiya. Penza: Izdatel'stvo Penzenskogo gosudarstvennogo universiteta, 2005g., 273 s.

[2] *Yazov Yu.K. i dr.* Nejrosetevaya zashhita personal'nykh biometricheskikh dannykh. //Yu.K. Yazov (redaktor i avtor), soavtory V.I. Volchikhin, A.I. Ivanov, V.A. Funtikov, I.G. Nazarov // M.: Radiotekhnika, 2012 g. 157 s.

[3] *Akhmetov B.S., Ivanov A.I., Funtikov V.A., Bezyaev A.V., Malygina E.A.* Tekhnologiya ispol'zovaniya bol'shikh nejronnykh setej dlya preobrazovaniya nechetkikh biometricheskikh dannykh v kod klyucha dostupa. Monografiya, Kazakhstan, g. Almaty, TOO «Izdatel'stvo LEM», 2014 g. -144 c., svobodnyj dostup http://portal.kaznu.kz/files/publicate/2014-06-27-

11940.pdf
[4] *Akhmetov B.S., Volchikhin V.I., Ivanov A.I., Malygin A.Yu.* Algoritmy testirovaniya biometriko-nejrosetevykh mekhanizmov zashhity informatsii Kazakhstan, Almaty, KazNTU im. Satpaeva, 2013 g.- 152 s. ISBN 978-101-228-586-4, http://portal.kaznu.kz/files/publicate/2014-01-04-11940.pdf
[5] *Juels A., Wattenberg M. A.* Fuzzy Commitment Scheme // Proc. ACM Conf. Computer and Communications Security, 1999, p. 28–36
[6] *F. Monrose, M. Reiter, Q. Li, S. Wetzel.* Cryptographic key generation from voice. In Proc. IEEE Symp. on Security and Privacy, 2001
[7] *Juels A., Sudan M. A Fuzzy Vault Scheme* // IEEE International Symposium on Information Theory, 2002
[8] *Y. Dodis, L. Reyzin, A. Smith*. Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy, Data April 13, In EUROCRYPT, pages 523-540, 2004.
[9] *Yang S., Verbauwhede I.* Automatic Secure Fingerprint Verification System Based on Fuzzy Vault Scheme // Proc. IEEE ICASSP 2005, p.609-612
[10] *Ramírez-Ruiz J., Pfeiffer C., Nolazco-Flores J.* Cryptographic Keys Generation Using FingerCodes. //Advances in Artificial Intelligence - IBERAMIA-SBIA 2006 (LNCS 4140), p. 178-187, 2006
[11] *Feng Hao, Ross Anderson, and John Daugman.* Crypto with Biometrics Effectively, IEEE TRANSACTIONS ON COMPUTERS, VOL. 55, NO. 9, SEPTEMBER 2006.
[12] *Cauchie S., Brouard T., Cardot H.* From features extraction to strong security in mobile environment: A new hybrid system. //On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops, Springer, pp. 489-498, 2006
[13] *Arakala A., Jeffers J., Horadam K.J.* Fuzzy Extractors for Minutiae-Based Fingerprint Authentication. //Advances in Biometrics (LNCS 4642), Springer, pp. 760-769, 2007
[14] *Lee Y.J., Bae K., Lee S.J., Park K.R., Kim J.* Biometric Key Binding: Fuzzy Vault Based on Iris Images. // Proceedings of 2nd International Conference on Biometrics, p. 800–808, Seoul, South Korea, August 2007
[15] *Nandakumar K., Jain A.K., Pankanti S.* Fingerprint-Based Fuzzy Vault: Implementation and Performance. //IEEE Transactions on Information Forensics and Security 2(4), pp. 744–757, 2007
[16] *Balakirsky V.B., Ghazaryan A.R., Han Vinck A.J.* Constructing Passwords from Biometrical Data. //Advances in Biometrics (LNCS 5558), Springer, pp. 889-898, 2009
[17] *Kanade S., Petrovska-Delacretaz D., Dorizzi B.* Multi-Biometrics Based Cryptographic Key Regeneration Scheme. //Proceedings of the 3rd IEEE international conference on Biometrics: Theory, applications and systems, p. 333-339, 2009
[18] *Chmorra A.L.* Maskirovka klyucha s pomoshh'yu biometrii «Problemy peredachi informatsii» 2011 № 2(47) s. 128-143.
[19] *Ushmaev O.V., Kuznetsov V.V.* Algoritmy zashhishhennoj verifikatsii na osnove binarnogo predstavleniya topologii otpechatka pal'tsev. //Informatika i ee primeneniya. 2012 g. №6(1), s. 132-140.
[20] Sreda modelirovaniya «BioNejroAvtograf» razmeshhena na sajte OAO «Penzenskij nauchno-issledovatel'skij ehlektrotekhnicheskij institut» http://pniehi.rf/activity/science/noc.htm. Produkt sozdan laboratoriej biometricheskikh i nejrosetevykh tekhnologij OAO «PNIEHI» dlya svobodnogo rasprostraneniya sredi universitetov Rossii, Belorussii, Kazakhstana.
[21] *Inge-Vechtomov S. G.* Genetika s osnovami selektsii. M., Vysshaya shkola, 1989, 592 s.
[22] *Boos V.* Lektsii po matematike. Tom 10. Perebor i ehffektivnye algoritmy: Uchebnoe posobie. – M.: Izdatel'stvo LKI, 2012 g. – 216 s.

Нейрожелілік функционалдың матрицаларының айналымының есептеуші күрделілігін бағалау

Б.С.Ахметов[1], А.И. Иванов[2], А.В. Безяев[3], С.В. Качалин[4]

Негізгі сөздер: жасанды нейронды желілер, жоғары өлшемдік, нейрожелілік функционалдың матрицаларының айналымы

Аннотация. Жұмыста адамның биометриалық бейнесін тану есебі сызықты алгебра ішінде «лағынет» өлшемдігіне байланысты өте қиын шешілетін есепке жататыны көрсетілген. Есептеуші қиындықтарды үлкен жасанды нейронды желілерді үйрету арқылы айналып өтуге болады. Олар нейрожелілік функционалдар сызықты емес алгебра матрицасымен сипатталады.

Сведения об авторах

Ахметов Бахытжан Сражатдинович, д.т.н., профессор, г. Алматы, ул. Сатпаева, 22, директор Института информационных и телекоммуникационных технологий, Казахский Национальный Технический Университет имени К.И. Сатпаева, 8(727) 257-70-34, E-mail: b_akhmetov@ntu.kz

Иванов Александр Иванович, д.т.н., доцент, 440000, г. Пенза, ул. Советская, 9, начальник лаборатории биометрических и нейросетевых технологий ОАО «Пензенский научно-исследовательский электротехнический институт», Тел. (8412) 36-80-92, E-mail: ivan@pniei.penza.ru

Безяев Александр Викторович – к.т.н., ведущий специалист Пензенского филиала ФГУП «НТЦ «Атлас», 440000, г. Пенза

Качалин Сергей Викторович - ведущий специалист ОАО «НПП «Рубин», 440000, г. Пенза.