ИДЕНТИФИКАЦИЯ АВТОРСТВА РУКОПИСНЫХ ОБРАЗОВ С ИСПОЛЬЗОВАНИЕМ НЕЙРОСЕТЕВОГО ЭМУЛЯТОРА КВАДРАТИЧНЫХ ФОРМ ВЫСОКОЙ РАЗМЕРНОСТИ

***Качайкин Евгений Иванович**, советник отдела защиты информации ДОК Минюста России, г. Москва*
E-mail: kachjkin@gmail.com
***Иванов Александр Иванович**, доктор технических наук, доцент, начальник лаборатории биометрических и сетевых технологий ПНИЭИ, г. Пенза*
E-mail: ivan@pniei.penza.ru

Рассматривается задача идентификации авторства рукописного автографа или иного рукописного слова. Показано, что почерковедческая экспертиза, проводимая человеком–экспертом, всегда осуществляется в пространстве признаков низкой размерности (учитываются 16 признаков). Переход к проведению экспертизы средствами искусственного интеллекта позволяет поднять размерность решаемой задачи до 416 учитываемых параметров рукописного почерка. При этом воспользоваться квадратичными формами линейной алгебры технически невозможно из–за плохой обусловленности задачи обращения корреляционных матриц высокой размерности. Предложено использовать нейросетевой эмулятор квадратичных форм, отличающийся тем, что легко обучается и может иметь сколь угодно большую размерность.

***Ключевые слова**: биометрические данные, квадратичные формы высокой размерности, нейросетевой эмулятор квадратичных форм*

IDENTIFICATION OF AUTHORSHIP OF HANDWRITTEN IMAGES USING NEURAL NETWORK EMULATOR OF QUADRATIC FORMS HIGH DIMENSION

***Evgeny Kachaykin**, adviser for information Security Bureau of Management and Control Department of Russia Ministry of Justice, Moscow*
E-mail: kachjkin@gmail.com
***Alexander Ivanov**, Doctor of Science (Tech), Associate Professor, Head of Laboratory for Biometrics and Neural Network Technology of PREI, Penza*
E-mail: ivan@pniei.penza.ru

We consider the problem of identification of authorship of a handwritten signature or ano ther handwrit–ten word. It is shown, that handwriting examination conducted by a human expert is always exercised in a space of Low – dimensioned (there 16 signs). The transition to the conduction of examination by the means of artificial intelligence allows raising the dimension of the task to be solved to 416 of considered parameters of handwriting. So the usage quadratic forms in linear algebra are technically impossible due to the bad con–ditionality of the problem of matrices of high dimension. It's suggested to use a neural network emulator of quadratic forms, differed with that it is easily trained and can have any large dimension.

***Keywords**: biometrics, quadratic forms higher–dimensional neural network emulator of quadratic formsKeywords: attacks, the information system, the attacker, false negative, false positive, the window size.*

Введение

Развитие современного общества связано с его тотальной информатизацией, идут процессы перехода на электронный документооборот. Важным элементом электронных документов является внедренный в них образ подписи (автографа), что приближает электронный документ к традиционному документу на бумажном носителе. При необходимости, сам файл электронного документа с внедренным с в него образом автографа может быть охвачен криптографической электронной цифровой подписью [1].

Одной из проблем информатизации современного общества является использование длинных идентификационных логинов. Каждый значимый информационный ресурс (например, электронное правительство) создает для своих пользователей личные кабинеты. Для входа в личный кабинет необходимо набрать уникальный логин и ввести длинный пароль из случайных знаков. Для всех электронных кабинетов логины и пароли должны быть разными. Биометрическая идентификация и аутентификация личности при входе в личный кабинет может быть осуществлена с использованием автографа и рукописного пароля. Автограф следует преобразовывать в длинный логин, так как это открытый биометрический образ. Рукописный пароль может быть использован для его преобразования нейронной сетью в длинный случайный пароль доступа [2].

Следующим важным приложением искусственного интеллекта является осуществление почерковедческой экспертизы автографов в обычных и электронных документах. При преобразованиях рукописного пароля в открытый код доступа (логин) и при автоматизированной почерковедческой экспертизе анализируемые биометрические параметры являются открытыми. То есть эти задачи можно решать с использованием сетей Байеса, радиальных сетей Пирсона-Хэмминга, сетями квадратичных форм. Все эти инструменты не способны скрывать анализируемые ими статистические параметры, но работают лучше, чем стандартные нейронные сети из перспептронов [2], одновременно решающие и задачу защиты (сокрытия) биометрических данных и задачу их распознавания (преобразования в код пароля доступа).

Проведенные исследования показали [3, 4], что 36 мерная сеть Байеса-Хемминга дает при анализе автографов более высокие показатели в сравнении с 416 мерной сетью 256 персептронов, реализованных в среде моделирования «БиоНейроАвтограф» [5]. Получается, что при анализе биометрических параметров рукописных автографов сети Байеса имеют значительные преимущества в сравнении с сетями персептронов [5] и радиальными сетями Пирсона-Хэмминга.

Известно, что правило Байеса построено на учете вероятностей появления зависимых событий $P(v_m/v_k)$, что эквивалентно учету при статистических оценках корреляционной связи зависимых событий $r(v_m,v_k)$. Чем выше значение модуля корреляционных связей - $|r(v_m,v_k)|$ между биометрическими параметрами, тем эффективнее работает правило Байеса при сравнении биометрических параметров с эталонами. В связи с этим возникает актуальная задача разработки эффективных высоко размерных алгоритмов идентификации, учитывающих существующие у объекта идентификации корреляционные связи. Предположительно, что повышение размерности решаемой задачи при одновременном учете собственных корреляционных связей объекта идентификации дадут существенный рост качества принимаемых решений.

Проблема применения квадратичных форм

Если оставаться в рамках линейной алгебры, то при идентификации следует применять квадратичные формы:

$$y(\overline{v}) = (E(\overline{v}) - \overline{v})^T \cdot [R]^{-1} \cdot (E(\overline{v}) - \overline{v}) \qquad (1),$$

где $\overline{v}$ - вектор нормированных биометрических параметров с единичными стандартными отклонениями.

Проблема квадратичных форм состоит в том, что необходимо вычислять обратную корреляционную матрицу $[R]^{-1}$ для ее подстановки в (1). Для биометрических данных корреляционные матрицы 2 и 3 порядков могут быть обращены. Корреляционные матрицы более высоких порядков не обращаются. При попытках обращения 416-мерной корреляционной матрицы возникает эффект «проклятия» размерности.

Причина технических трудностей состоит в низкой точности оценок корреляционных функций, вычисляемых на малых выборках биометрических данных. Так в биометрических приложениях при обучении [2] и при тестировании [6, 7] искусственных нейронных сетей используется порядка 20 примеров рукописного образа. Аналогичная ситуация возникает и при почерковедческой экспертизе. При столь малых выборках данных ошибка вычисления коэффициента корреляции может составлять от 10% до 80%. Эта ситуация отражена на рисунке 1 при оценках нескольких значений коэффициентов корреляции. Наибольшая погрешность получается при оценке малых значений коэффициентов корреляции (центр рисунка 1). По мере того как модуль оцениваемого коэффициента корреляции увеличивается интервал ошибок оценки сжимается. Наиболее точными оказываются оценки единичных коэффициентов корреляции.

В связи с этим возникает неприятная ситуация, когда ошибки оценок коэффициентов корреляции и обусловленность обращаемых корреляционных матриц препятствуют решению задачи. В ситуации, когда корреляционная матрица диагональна

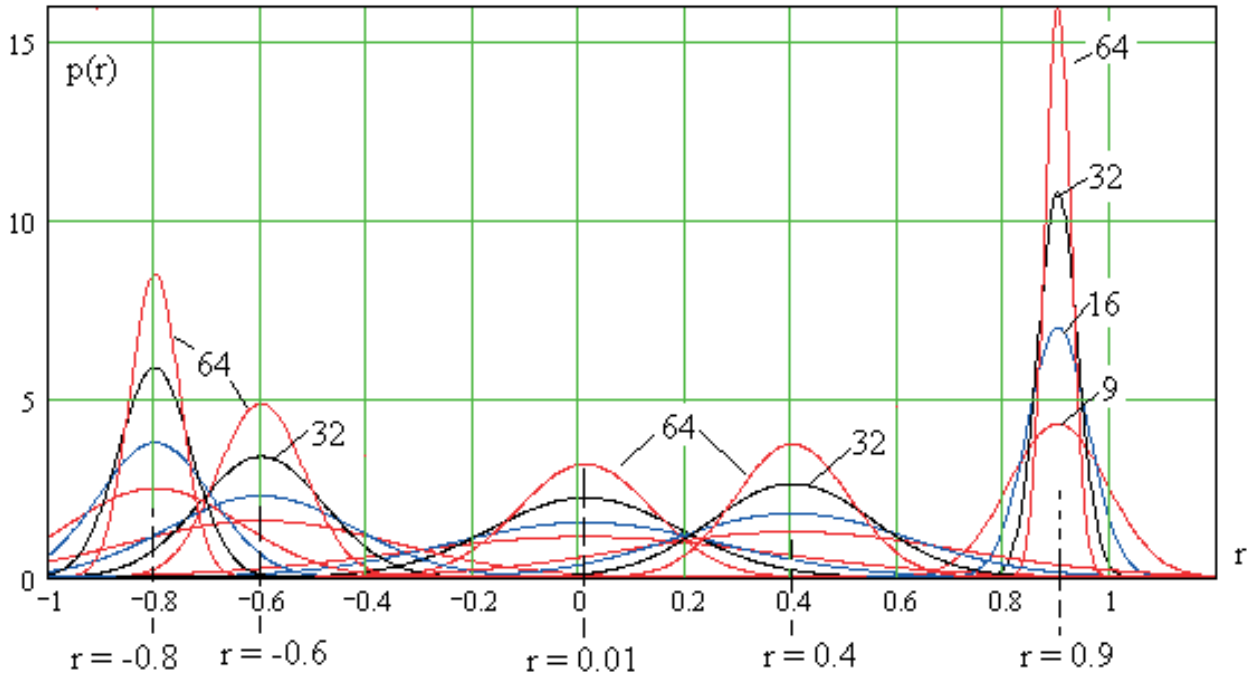

Рис. 1. *Распределения случайных значений вероятности коэффициентов корреляции, обусловленных конечной выборкой наблюдений*

(нет корреляционных связей) мы не можем надежно оценить коэффициенты корреляции (нулевая корреляция оценивается с наибольшей погрешностью). Если же все коэффициенты корреляции оказываются велики, то число обусловленности такой матрицы оказывается очень большим и не дает нам возможности решить задачу.

Описанное выше противоречие приводит к трудностям даже, если мы пытаемся использовать квадратичные формы второго и третьего порядка. Практика показывает, что даже для синтеза квадратичных форм второго и третьего порядка под задачи биометрической идентификации приходится использовать сотни примеров распознаваемого рукописного образа.

Понятно, что получить обратную корреляционную матрицу 416 порядка технически невозможно. Заметить 416-мерную квадратичную форму сетью квадратичных форм 2-го и 3-го порядка также оказывается невозможно при использовании выборок биометрического образа малых размеров.

Эмуляция квадратичных форм высокой размерности нейронной сетью с большим числом выходов

Так как полноценное обращение корреляционных матриц высокой размерности технически невыполнимо, будем использовать нейросетевой эмулятор квадратичных форм. Для этой цели следует использовать искусственные нейроны с четными функциями квантования [8, 9] или радиально-базисные функции возбуждения с двухсторонним ограничением. Структура нейронной сети с радиально-базисными нейронами и двухсторонним квантованием приведена на рисунке 2.

Входы у нейронов выбираются случайно в интервале от 1 до 416, число входов должно быть равно примерно трети от общего числа учитываемых биометрических параметров 416/3≈139. Каждый выходной сумматор радиально-базисных нейронов описывается следующим соотношением:

$$y_k = \sum_{i=1}^{139} \mu_i \cdot \nu_i \qquad (2).$$

Пороговая функция настраивается таким образом, что бы примеры автографа «подлинник» давали состояние «0» для интервала $E(y_k)\pm 3\sigma(y_k)$. При настройке нейросетевого эмулятора на требуется знать математическое ожидание - $E(y_k)$ и стандартное отклонение - $\sigma(y_k)$ имеющихся в нашем распоряжении примеров «подлинник».

Следует подчеркнуть, что выполнить подбор весовых коэффициентов μ_i в уравнении (2) процедурами обычного итерационного обучения [8] нельзя из-за конечности обучающей выборки.

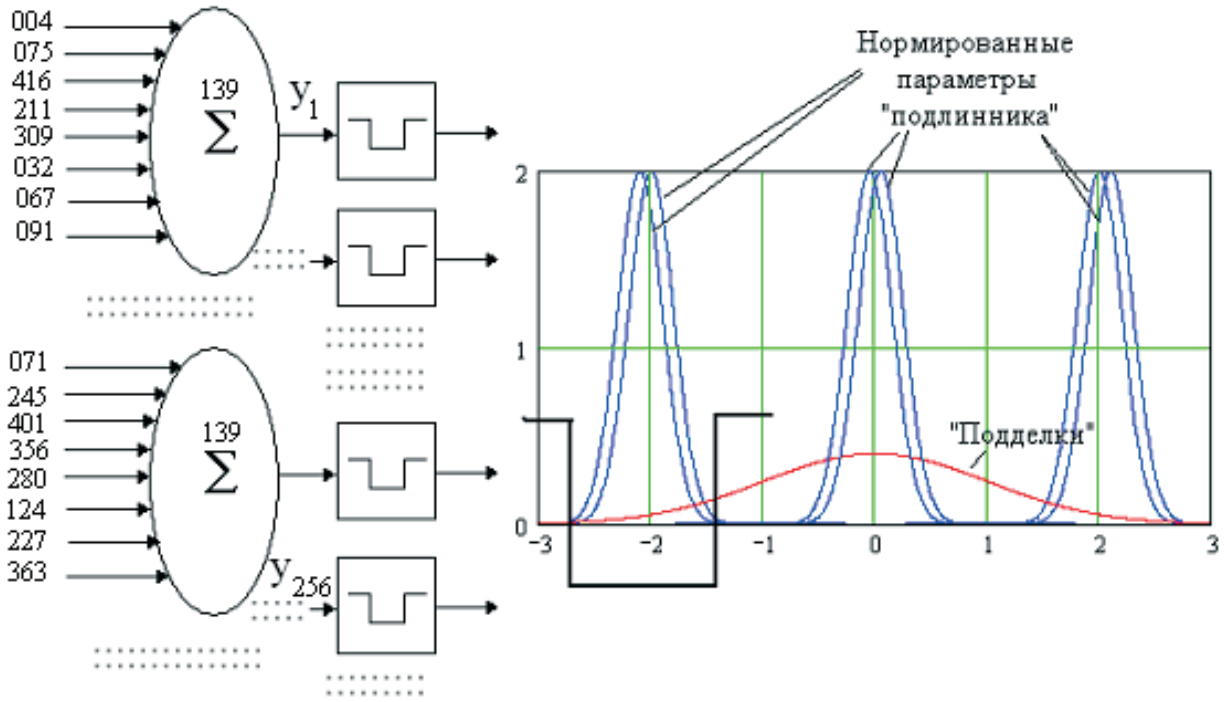

Рис. 2. *Структура нейросетевого эмулятора квадратичных форм*

Если выборка состоит из 20 примеров «подлинник», то обычные итерационные алгоритмы обучения будут работоспособны только для нейронов с числом входов до 20. В нашем случае число входов в 5 раз больше, приходится принимать специальные меры регуляризации вычислений.

Сумматор с 139 входами следует представить в виде пирамиды из $N=139/1+139/2+139/4\ldots+139/2^7=276$ двухвходовых сумматоров. Каждый следующий слой пирамиды имеет в два раза меньше двухвходовых сумматоров, чем предыдущий слой. Настройка двухвходовых сумматоров оказывается очень простой, если все биометрические данные поделить на три группы:

- данные с большими отрицательными значениями параметров от -3 до -0.7;
- данные с малыми значениями параметров от -0.7 до +0.7;
- данные с большими положительными значениями параметров от 0.7 до 3.

При такой декомпозиции задача настройки каждого из двухвходовых сумматоров сводится к выбору нормированного биометрического параметра из своей группы дающего минимум стандартного отклонения на выходе. При настройке каждого их двухвходовых сумматоров (на всех уровнях пирамиды) приходится иметь дело с сопоставимыми по точности данными. Возникающие при расчетах погрешности оказываются разных знаков и на конечный результат влияют слабо. Представление большого сумматора пирамидой двухвходовых сумматоров является приемом регуляризации процедуры настройки весовых коэффициентов нейрона. После настройки всех сумматоров результирующий весовой коэффициент получается взвешенным суммированием последовательности всех весовых коэффициентов сумматоров, образующих один из путей движения к вершине пирамиды.

Эквивалентность нейросетевого эмулятора технически не реализуемым высокоразмерным квадратичным формам

Каждый из квантователей радиально-базисных нейронов в 139 – мерном пространстве входных данных выделяет 139-мерный бесконечный многогранник. Если мы будем рассматривать любое из двухмерных сечений 139-мерного пространства, то будем наблюдать в этом пространстве эллипс распределения данных «подлинник» окруженный касательными проекций разделяющих многомерное пространство гиперплоскостей. Эта ситуация отображена на рисунке 3.

Все выходные состояния нейронов, попадающие внутрь эллипса «подлинник», на выходе ней-

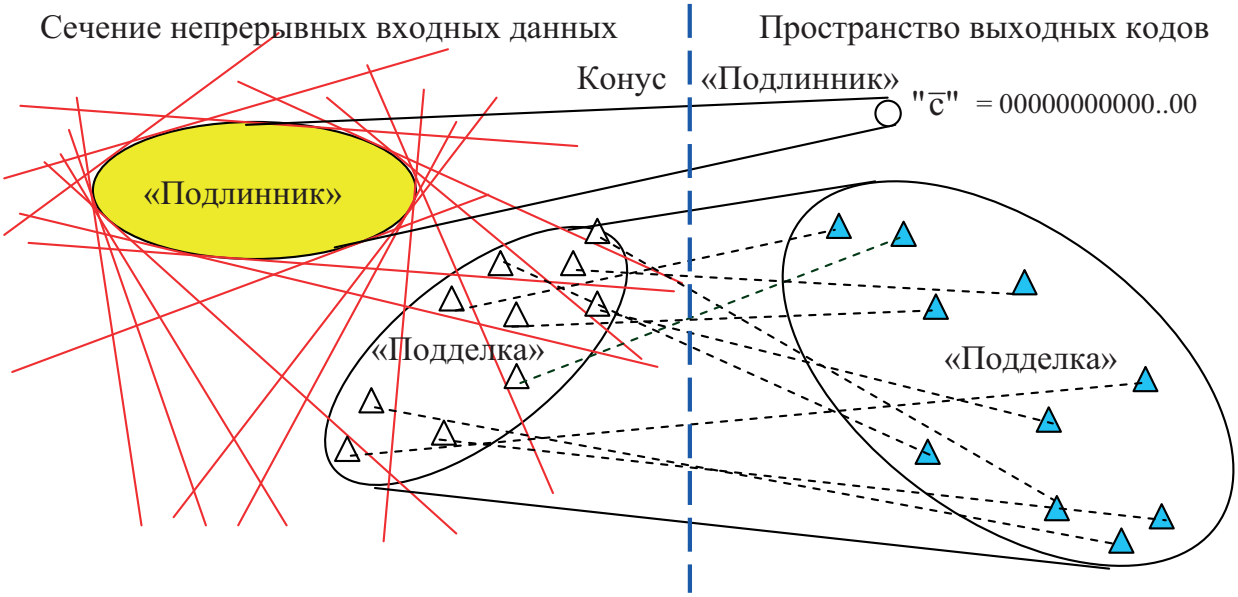

Рис. 3. *Иллюстрация работы нейросетевого эмулятора высоко размерных квадратичных форм*

росети будут давать один и тот же 256 разрядный код «0000000…00». Как только какой-то их выходных разрядов получает состояние «1», можно утверждать, что соответствующий 139-мерный наблюдатель увидел выпадение данных за пределы 139- мерного гиперэллипсоида «подлинник». Число обнаруженных единиц в выходном коде нейросетевого эмулятора соответствует числу 139 мерных наблюдателей, принявших решение «подделка». Близкие к «подлиннику» «подделки» будут иметь малое число состояний «1». Чем больше единиц в выходном коде, тем сильнее «подделка» отличается от «подлинника».

Для нас принципиально важным является то, что синтез и обучение нейросетевого эмулятора 416-мерных квадратичных форм является простой в вычислительном отношении процедурой. Классические квадратичные формы (1) столь высокой размерности технически реализовать невозможно, оставаясь в рамках линейной алгебры. Однако если перейти к нейросетевым эмуляторам квадратичных форм они реализуются без особых проблем, при этом мы оказываемся в рамках некоторой нейросетевой нелинейной алгебры с рекордно устойчивыми в вычислительном отношении преобразованиями. Заметим, что квадратичные формы – это только одно из применений наглядно демонстрирующих возможности нелинейной алгебры нейросетевых преобразований. Обращение матриц высокоразмерных нейросетевых преобразований [10, 11], так же оказывается очень устойчивой процедурой.

Литература:

1. Патент № RU2543928 Способ формирования электронного документа и его копий. Заявка: 2013151257/08 от 18.11.2013, H04L 9/32, G06F 21/64. Опубликовано: 10.03.2015. Бюл. № 7, авторы: Ложников П.С., Иванов А.И.
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. 157 с.
3. Сулавко А.Е. Идентификация пользователей компьютерных систем по динамике подсознательных движений на основе статистической теории принятия решений. Автореферат диссертации к.т.н. по специальности 05.13.19, 2014, совет Д-212.288.07, ФГБОУ ВПО «Уфимский государственный авиационный технический университет».
4. Епифанцев Б.Н., Ложников П.С., Сулавко А.Е. Алгоритм идентификации гипотез в пространстве малоинформативных признаков на основе последовательного применения формулы Байеса // Межотраслевая информационная служба. – 2013. №2 (163). – С.57-62.
5. Среда моделирования «БиоНейроАвтограф» размещена на сайте ОАО «ПНИЭИ» http://пниэи.рф/activity/science/noc.htm. Продукт создан лабораторией биометрических и нейросетевых технологий ОАО «ПНИЭИ» в период 2009-2014 г.г. для свободного использования университетами России, Белоруссии, Казахстана.
6. Ахметов Б.С., Надеев Д.Н., Фунтиков В.А., Иванов А.И., Малыгин А.Ю. Оценка рисков высоконадежной биометрии. Монография. Алматы: Из-во КазНТУ им. К.И. Сатпаева, 2014 - 108 с.

7. 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 - 152 с. ISBN 978-101-228-586-4, http://portal.kaznu.kz/files/publicate/2014-01-04-11940.pdf
8. Саймон Хайкин. Нейронные сети: полный курс. М.: «Вильямс», 2006. — С. 1104.
9. Ахметов Б.С., Иванов А.И., Фунтиков В.А., Малыгин А.Ю. Основы биометрической идентификации личности. Учебное пособие. КазНТУ им. К.И. Сатпаева, Алматы, 2014, 151 с. ISBN 978-601-228-689-2
10. Иванов А.И., Малыгина Е.А. Биометрическая аутентификация личности: обращение матриц нейросетевых функционалов в пространстве метрики Хемминга // Вопросы защиты информации. №1, 2015, с.23-29.
11. Ахметов Б.С., Иванов А.И., Безяев А.В., Качалин С.В. Оценка вычислительной сложности обращения матриц нейросетевых функционалов. Доклады национальной академии наук республики Казахстан. 2014, №5, с. 49-61.

References:

1. Patent № RU2543928 Sposob formirovaniya ehlektronnogo dokumenta i ego kopij. Zayavka: 2013151257/08 ot 18.11.2013, H04L 9/32, G06F 21/64. Opublikovano: 10.03.2015. Byul. № 7, avtory: Lozhnikov P.S., Ivanov A.I.
2. YAzov YU.K. i dr. Nejrosetevaya zashchita personal'nyh biometricheskih dannyh. // YU.K.YAzov (redaktor i avtor), soavtory V.I. Volchihin, A.I. Ivanov, V.A. Funtikov, I.G. Nazarov // M.: Radiotekhnika, 2012 g. 157 s.
3. Sulavko A.E. Identifikaciya pol'zovatelej komp'yuternyh sistem po dinamike podsoznatel'nyh dvizhenij na osnove statisticheskoj teorii prinyatiya reshenij. Avtoreferat dissertacii k.t.n. po special'nosti 05.13.19, sentyabr' 2014 g., sovet D-212.288.07, FGBOU VPO «Ufimskij gosudarstvennyj aviacionnyj tekhnicheskij universitet».
4. Epifancev B.N., Lozhnikov P.S., Sulavko A.E. Algoritm identifikacii gipotez v prostranstve maloinformativnyh priznakov na osnove posledovatel'nogo primeneniya formuly Bajesa // Mezhotraslevaya informacionnaya sluzhba. – 2013. №2 (163). – S.57-62.
5. Sreda modelirovaniya «BioNejroAvtograf» razmeshchena na sajte OAO «PNIEHI» http://pniehi.rf/activity/science/noc.htm. Produkt sozdan laboratoriej biometricheskih i nejrosetevyh tekhnologij OAO «PNIEHI» v period 2009-2014 g.g. dlya svobodnogo ispol'zovaniya universitetami Rossii, Belorussii, Kazahstana.
6. Ahmetov B.S., Nadeev D.N., Funtikov V.A., Ivanov A.I., Malygin A.YU. Ocenka riskov vysokonadezhnoj biometrii. Monografiya. Almaty: Iz-vo KazNTU im. K.I. Satpaeva, 2014 g.- 108 s.
7. Ahmetov B.S., Volchihin V.I., Ivanov A.I., Malygin A.YU. Algoritmy testirovaniya biometriko-nejrosetevyh mekhanizmov zashchity informacii Kazahstan, 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
8. Sajmon Hajkin. Nejronnye seti: polnyj kurs. M.: «Vil'yams», 2006. — S. 1104.
9. Ahmetov B.S., Ivanov A.I., Funtikov V.A., Malygin A.YU. Osnovy biometricheskoj identifikacii lichnosti. Uchebnoe posobie. KazNTU im. K.I. Satpaeva, Almaty, 2014 g., 151 s. ISBN 978-601-228-689-2
10. Ivanov A.I., Malygina E.A. Biometricheskaya autentifikaciya lichnosti: obrashchenie matric nejrosetevyh funkcionalov v prostranstve metriki Hemminga // Voprosy zashchity informacii. №1, 2015 g. s.23-29.
11. Ahmetov B.S., Ivanov A.I., Bezyaev A.V., Kachalin S.V. Ocenka vychislitel'noj slozhnosti obrashcheniya matric nejrosetevyh funkcionalov. Doklady nacional'noj akademii nauk respubliki Kazahstan. 2014, №5, s. 49-61.