http://rosoperator.ru/rameev100/
Всероссийская научно-техническая конференция, посвященная 100-летию со дня рождения одного из основоположников советской вычислительной техники Б.И. Рамеева, на тему: «Информационно-управляющие и телекоммуникационные системы специального назначения». Секция «Биометрическая поддержка криптовалют, блокчейн реестров, облачных сервисов», проводится в конференц-зале Технопарка высоких технологий «Рамеев», доклад 16 мая 2018 года с 13^{15} до 13^{30} (ул. Центральная, д. 1), г. Пенза.

УДК: 519.24; 53; 57.017

УСИЛЕНИЕ УРОВНЯ ЗАЩИЩЕННОСТИ БИОМЕТРИЧЕСКИХ ТЕХНОЛОГИЙ ЗА СЧЕТ ИСПОЛЬЗОВАНИЯ НЕЙРОНОВ КРАМЕРА-ФОН МИЗЕСА

Вятчанин С.Е., Малыгина Е.А.

Аннотация.

Рассматривается перспектива использования нейронов Крамера-фон Мизеса в биометрико-нейросетевых механизмах защиты в современных инфокоммуникационных системах обработки данных в облачных сервисах.

Показано, что мощность критерия Крамера – фон Мизеса на малых выборках примеров биометрических данных оказывается существенно выше, чем мощность аналогичного критерия хи-квадрат, что делает их крайне перспективными для применения в нейросетевых преобразователях биометрия-код.

*Ключевые слова:*критерий хи-квадрат, нейрон Крамера-фон Мизеса, нейросетевой преобразователь биометрия-код, биометрические данные, большая размерность данных.

Становление, развитие и защита российской цифровой экономики невозможна без активного использования отечественной криптографии всовременных инфокоммуникационных системах обработки данных в облачных сервисах.

К сожалению, пользователи современных инфокоммуникационных систем не в состоянии запоминать длинные пароли доступа к информационным ресурсам. Эту проблему предложено решить при помощи биометрических данных самого пользователя данных сервисов.

В США, Канаде, Евросоюзе развивается технология так называемых «нечетких» экстракторов [1, 2, 3]. Технология сводится к тому, что из биометрического образа извлекаются его параметры, далее их значения квантуются. Как отмечено в зарубежных научных публикациях полученный БиоКод может содержать от 20% до 30% ошибок. Для их устранения используют маскирование наиболее нестабильных бит БиоКода,не устраненные ошибки обнаруживаются и исправляются классическими самокорректирующимися кодами, обладающими примерно 20 кратной избыточностью. Таким образом длина БиоКода оказывается примерно в 20 меньше числа биометрических параметров, извлеченных из образа. Поэтому БиоКоды на выходе «нечетких» экстракторов оказываются недопустимо короткими для того, чтобы далее использовать полноценные криптографические механизмы.

В России создается, стандартизуется и поддерживается технология нейросетевого преобразования биометрических данных в код заранее полученного криптографического ключа [4, 5]. Основной проблемой применения искусственных нейронных сетей является то, что они очень медленно учатся и для их обучения [6] необходимы огромные обучающие выборки, содержащие порядка 100 000 примеров биометрических образов. Медленное обучение и большие выборки обучающих примеров для обучения «глубоких» нейронных сетей [6] необходимы из-за того, что итерационный алгоритм «обратного распространения ошибки» имеет экспоненциальную вычислительную сложность.

Выходом из тупика экспоненциальной вычислительной сложности обучения «глубоких» нейронных сетей стало решение использовать однослойные искусственные нейронные сети с большим числом выходов. Теоретически и практически получены результаты, показывающие, что для однослойных искусственных нейронных сетей вычислительная сложность обучения оказывается линейной [7]. Это позволило их обучение на выборке из 20 примеров образа «Свой». При этом время обучения оказывается незначительным (от 0.5 до 0.05 секунды) даже при использовании ПЭВМ с процессорами низкой производительности.

Основная идея быстрых стандартизованных алгоритмов не итерационного обучения сводится к тому, что обучение линейной части нейрона (сумматора) и нелинейной части нейрона (квантователя) разделены. Декомпозиция задачи на две простых подзадачи (линейную свертку в пространстве входных состояний и нелинейное преобразование выходных данных свертки [8]) всегда приводит к значительному упрощению вычислений. В такой декомпозиции задачи построен алгоритм обучения, регламентированный стандартом ГОСТ Р 52633.5 [8]. Прозрачная парадигма разделения задачи обучения нейронной сети на две составляющих (линейную и нелинейную) дает очень хорошие результаты по сокращению размеров обучающей выборки и сведению задачи до линейной вычислительной сложности. Тем не менее, этот конструктивный прием, не является универсальным, вполне могут существовать и другие варианты декомпозиции задачи, приводящие к близким или даже более качественным результатам.

В настоящее времясуществует множество статистических критериев для поверки гипотезы нормального распределения значений биометрических данных. Наиболее часто на практике используется хи-квадрат критерий [9, 10] в силу того, что для этого критерия Пирсон в 1900 году построил аналитическое описание хи-квадрат распределения.

К большому сожалению, хи-квадрат критерий оказывается наиболее точным при выборках в 200 опытов и больше. При обработке биометрических данных приходится иметь выборки в 20 примеров, что не даетиспользоватькритерийхи-квадрат.

Такая же ситуация возникает и случаи использования других не параметрических статистических критериев [11], все они так же дают надежные оценки при выборках существенно более 20 примеров.

Теоретически доказано, что для биометрии данные одного контролируемого параметра примеров образа «Свой» имеют распределение близкое к нормальному. Для данных случайно выбранных образов «Чужие» распределение оказывается близко к равномерному. В связи с этим есть возможность исследуемые оцениватькритерии по отношению к друг другу сравнивая между собой, даваемые ими равные ошибки первого и второго рода $P_1=P_2=P_{EE}$ при использовании малых выборок нормального и равномерного законов распределения. На рисунке 1 приведены функции снижения вероятностей ошибок по мере роста размеров тестовой выборки.

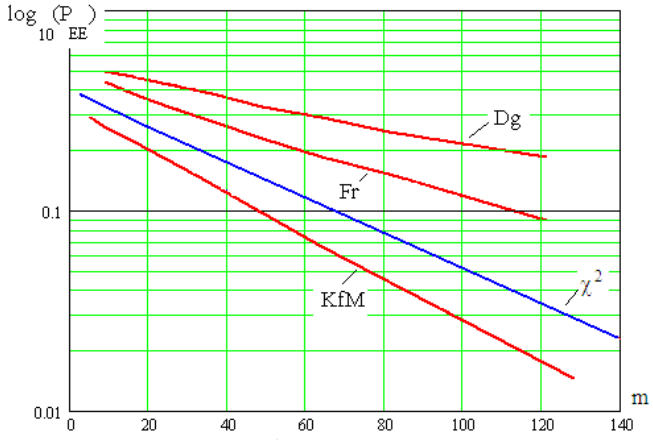

Рис.1 − График сопоставления мощностей интегральных статистических критериев Джини (Dg), Фроцини (Fr), хи-квадрат (χ^2) и Крамера-фон Мизеса (KfM)

Из данных рисунка 1 видно, что мощность критерия KfM при малых выборках оказывается выше, чем у хи-квадрат критерия. Это делает критерий KfM более перспективным для биометрии в сравнении с хи-квадрат критерием, критерием Фроцини и критерием Джини.

По всей вероятности, основной причиной высокой мощности «Крамера-фон Мизеса» (KfM) критерия является то, что у него оказываются минимальные шумы квантования континуальных данных.

Исходный критерий «Крамера-фон Мизеса» вычисляется через интеграл квадрата разницы между гипотетической функцией вероятности – $P(v)$ эмпирической функции вероятности $\hat{P}(v)$:

$$KfM = \int_{-\infty}^{+\infty} (P(v) - \hat{P}(v))^2 \, dv. \qquad (1)$$

Формальный переход от интеграла (1) к операции суммирования нейроном с n входами не сложен:

$$\begin{cases} KfM(v) = \sum_{i=1}^{n} (P(v_i) - \hat{P}(v_i))^2 \\ KfM(\xi) = \sum_{i=1}^{n} (P(v_i) - \hat{P}(\xi_i))^2 \end{cases}, \qquad (2)$$

где v_i пример образа «Свой» по n-контролируемым нейроном биометрическим параметрам; ξ_i пример образа «Чужой» по n-контролируемым нейроном биометрическим параметрам.

Эталонная функция вероятности - $P(v)$ оказывается легко вычислима только в том случае, когда выполнено центрирование и нормирование, контролируемых биометрических параметров образа «Свой»:

$$\tilde{v}_i = \frac{v_i - E(v_i)}{\sigma(v_i)}. \qquad (3)$$

В новой центрированной и нормированной системе биометрических координат эталонная функция вероятности имеет нулевое математическое ожидание и единичное стандартное отклонение:

$$P_v(\tilde{v}) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\tilde{v}} \exp\left\{\frac{-u^2}{2}\right\} du. \qquad (4)$$

На рисунке 2 дан пример эталонной функции вероятности для нормального закона.

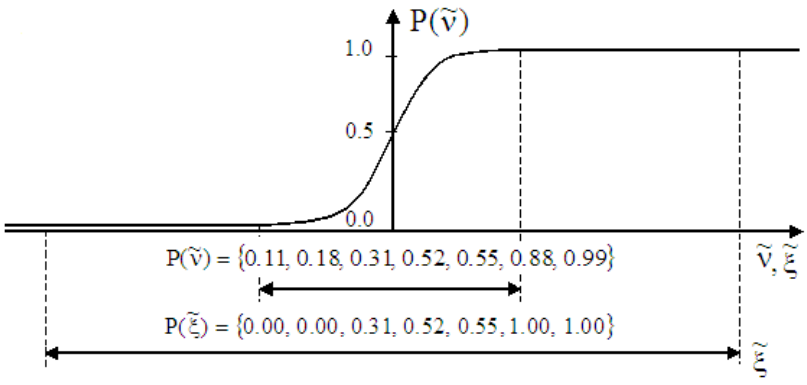

Рис.2 − Графический пример работы функции преобразования вектора значений переменных в вектор значений их вероятности

Из графика на рисунке2 видно, что данные примеров образа «Свой» после их преобразования в вероятность должны будут оказаться в интервале от 0.0 до 1.0. Причем появление предельных значений 0.0 и 1.0 маловероятно.

Совершенно иная ситуация возникает, если вероятностному нейрону «Крамера-фон Мизеса» будут предъявлены данные примера образа «Чужой». Как видно из рисунка 2 динамический диапазон данных - $\tilde{\xi}_i$ примерно в три раза больше, чем данных образа «Свой». Это означает, что на входы нейрона KfM будут поступать несколько предельных минимальных значений вероятности {0.0, 0.0,….} в начале упорядоченного списка и несколько предельных значений в конце упорядоченного списка {….., 1.0, 1,0}.

В целом вероятностный нейрон KfM работает за счет того, что выявляет неоправданно большое число предельных состояний вероятности в наблюдаемом биометрическом образе.

Нейрон KfM и структура нейронной сети, сформированной из нейронов KfM представлены на рисунке 3.

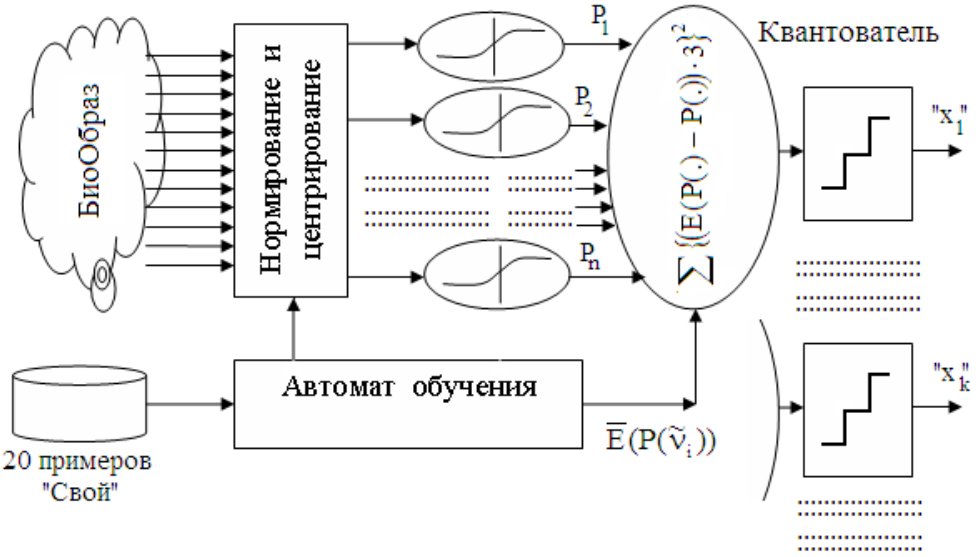

Рис. 3 − Обобщенная структура нейросетевого преобразователя биометрических данных в код, построенная на использовании вероятностных нейронов «Крамера-фон Мизеса»

Полностью автоматическое обучение сети нейронов «Крамера-фон Мизеса»

Нейронных сетей может быть много, так же, как и типов нейронов из которых они собраны. Для нейросетевой биометрии принципиально важным является то, чтобы нейронная сеть преобразователя биометрия-код была способна быстро обучаться на малой выборке примеров «Свой» в полностью автоматическом режиме. В частности, этими свойствами обладает сеть персептронов, обученных алгоритмом ГОСТ Р 52633.5 [8].

В соответствии с требованиями алгоритма ГОСТ Р 52633.5 [8] нейроны подключены ко входам нейросети случайным образом. При синтезе нейронной сети, связи каждого нейрона задаются таблицей, заполняемой от генератора псевдослучайных чисел. Для сети нейронов KfM связи каждого нейрона должны формироваться также.

Важнейшим аспектом автоматического обучения сетей KfM является вычисление вектора математических ожиданий - $\overline{E}(\nu_i)$ и вектора стандартных отклонений - $\overline{\sigma}(\nu_i)$ параметров 20-ти примеров образа «Свой». Эти данные запоминаются и используются для центрирования и нормирования (3) всех входных параметров нейронной сети.

Вторым важнейшим аспектом является вычисление средних значений вероятности для упорядоченной выборки биометрических параметров образа «Свой» $E(P(\tilde{\nu}_1)),\ \ E(P(\tilde{\nu}_2)),.....,E(P(\tilde{\nu}_n))$. Эти параметры далее используются как эталоны для сравнения с предъявленными нейрону данными:

$$\begin{cases} \mathrm{KfM}(\nu) = \sum_{i=1}^{n} \left\{3\left(\mathrm{E}(\mathrm{P}(\tilde{\nu}_i) - \mathrm{P}(\tilde{\nu}_i))\right)\right\}^2 \\ \mathrm{KfM}(\xi) = \sum_{i=1}^{n} \left\{3\left(\mathrm{E}(\mathrm{P}(\tilde{\nu}_i) - \mathrm{P}(\tilde{\xi}_i))\right)\right\}^2 \end{cases}. \qquad (5)$$

Система (5) содержит два уравнения, так как для данных образа «Свой» и для данных образа «Чужой» нейрон работает совершенно по-разному. Для данных образа «Свой», каждое из которых относительно мало происходит их квадратичное сжатие и накопление сумматором.

Для данных образа «Чужой» Происходит не только линейное обогащение данных за счет суммирования, но и его нелинейное усиление за счет того, что нормировка биометрических данных «Чужой» выполняется на базе статистических данных «Свой»:

$$\tilde{\xi}_i = \frac{\xi_i - \mathrm{E}(\nu_i)}{\sigma(\nu_i)}. \qquad (6)$$

Именно по этой причине с вероятностью выше 0.5 данные образа «Чужой» оказываются больше 1.0 и усиливаются квадратичной функцией. В конечном итоге выполняется условие:

$$\mathrm{KfM}(\xi) > \mathrm{KfM}(\nu). \qquad (7)$$

Так как нейроны «Крамера-фон Мизеса» работают с вероятностями, на каждом входе сумматора таких нейронов должен стоять функциональный преобразователь континуумов входных состояний в вероятности их появления. На рисунке 4 приведена схема процедуры нейросетевой аутентификации БиоОбразов, построенная на использовании множества вероятностных нейронов «Крамера-фон Мизеса». По такой схема каждый нейрон отвечает за один или два выходных разрядов кода аутентификации.

Принципиально важным свойством вероятностных нейронов KfM является очень высокая устойчивость их обучения на малых выборках. Для того, чтобы выполнить нормирование и центрирование биометрических данных под БиоОбраз «Свой» достаточно базы из 20 примеров. Как показала практика на базе обучающей выборки из 20 примеров образа «Свой» можно с достаточной точностью вычислить вектор математических ожиданий - $\overline{\mathrm{E}}(\nu_i)$ и вектор стандартных отклонений $\overline{\sigma}(\nu_i)$ всех контролируемых биометрических параметров.

Единственной не тривиальной функцией автомата обучения является вычисление по 20 примерам вектора математических ожиданий вероятности появления сортировки по возрастанию значений контролируемых биометрических параметров. При обучении нет переборов и запоминания промежуточных данных. Это означает, что обучение больших сетей искусственных нейронов «Крамера-фон Мизеса» имеет линейную вычислительную сложность, так же, как и алгоритм ГОСТ Р 52633.5 [8].

Усиление хэширующих свойств данных образов «Чужой» сетью нейронов «Крамера-фон Мизеса»

Основным свойством всех преобразователей биометрия-код является устранение почти до нуля естественной энтропии биометрических данных образа «Свой»:

$$\mathrm{H}(\overline{\nu}) >> \mathrm{H}(\text{"}\overline{\mathrm{c}}\text{"}) \cong 0.03 \text{ бита}. \qquad (8)$$

Для образов «Чужой» все нейросетевые преобразователи должны выполнять обратную функцию, усиливая естественную энтропию биометрических образов «Чужие»:

$$H(\bar{\xi}) < H("\bar{x}") \cong 45.03 \text{ бита.} \qquad (9)$$

Проведенные исследования показали, что сети нейронов «Крамера-фон Мизеса» при двухуровневых квантователях значительно уступают по их хэширующим свойствам, стандартизованным нейросетям, обученным по ГОСТ Р 52633.5. Это происходит из-за того, что все нейросети, обученные по ГОСТ Р 52633.5 сбалансированы по вероятности появления выходных состояний «0» и «1» в разрядах выходного кода для образов «Чужие». В сети нейронов KfM нет сбалансированности по вероятности появления в разрядах разных состояний. На рисунке 5 приведены распределения вероятностей образов «Чужой» и наиболее вероятные положения порогов срабатывания квантователей.

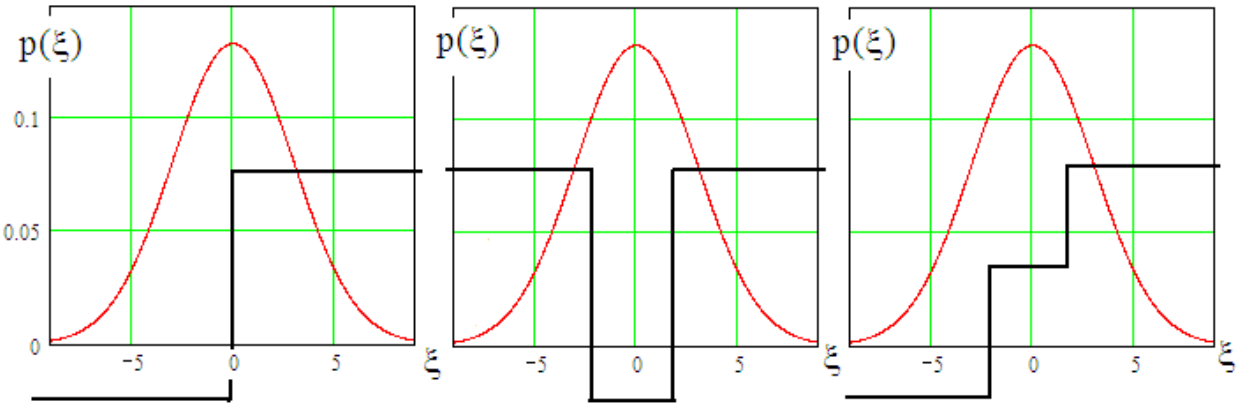

Рис. 5 - Графики примеров использования разного типа квантователей для линейных нейронов

При обучении линейных нейронов по ГОСТ Р 52633.5 пороги всех квантователей настроены на срабатывание в центе распределения параметров образов «Все «Чужие» как это показано на в левой части рисунка 5.

В центральной части рисунка 5 отображена ситуация, когда в место нейронов с линейным накоплением используется одна из возможных квадратичных форм. Тогда превышение порога сравнения дает двустороннее ограничение данных, как это показано в центральной части рисунка 5. В этом случае утрачивается баланс состояний «0» и «1» на выходах квадратичных нейронов. Этот эффект наблюдается и для нейронов KfM, как следствие наблюдается эффект снижения энтропии (9) квадратичных нейронов в сравнении с линейными нейронами, обученными по ГОСТ Р 52633.5 [8].

Для устранения этого нежелательного эффекта в нейронах KfM предложено использовать квантователи с тремя выходными состояниями[12 – 15], как это показано в правой части рисунка 5. Для этой цели производится контроль значений математических ожиданий - $E(P(\tilde{\xi}))$ на входе нейрона KfM:

$$\begin{cases} z(Kfm) = "01" \quad if \ \ KfM > \gamma \wedge E(P(\tilde{\xi})) > 0.5, \\ z(Kfm) = "00" \quad if \ \ KfM < \gamma, \\ z(Kfm) = "10" \quad if \ \ KfM > \gamma \wedge E(P(\tilde{\xi})) < 0.5 \end{cases}, \qquad (10)$$

где γ - порог срабатывания компаратора, находящегося на выходе кадратичного нейрона KfM, узнающего примеры образа «Свой» с высокой вероятностью в форме выходного состояния «00».

При использовании трехуровнего квантования (10) вероятности состояний «0» и «1» на выходе разрядов сети нейронов KfM выравниваются. Наблюдается рост энтропии выходных кодов «Чужой» до величины порядка 45 бит, что в полтора раза больше, чем тот же показатель для гостовских линейных нейронов.За счет использованиямеханизмов

трехуровнего квантования устраняется проведение «Атаки Маршалко» по извлечению данных из обученной нейронной сети[10, 11].

Заключение

Действующий в России ГОСТ Р 52633.5 распространяется только на автоматическое обучение сетей персептронов. Как было показано в этой статье сети KfM также имеют полностью автоматическое обучение с линейной вычислительной сложностью. Фактически действующий стандарт ГОСТ Р 52633.5 может быть дополнен таким же стандартом для сетей нейронов «Крамера-фон Мизеса».

Ранее было известно множество квадратичных нейросетевых функционалов [6, 7], однако они не использовались в нейросетевых преобразователях биометрия-код. Это было обусловлено низким уровнем энтропии их выходных кодов. Переход к троичным нейронам решает эту задачу, увеличивает энтропию выходных кодов по сравнению с сетями ГОСТ Р 52633.5 [5] и устраняет проведение «Атаки Маршалко» по извлечению данных из обученной нейронной сети[10, 11].

Литература:

1. Dodis Y., Reyzin L., Smith A. Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy // Proc. EUROCRYPT, April 13, pages 523-540, 2004.
2. Monrose F., Reiter M., Li Q., Wetzel S. Cryptographic key generation from voice. //Proc. IEEE Symp. on Security and Privacy, pp. 202-213, 2001.
3. Hao F., Anderson R., Daugman J. Crypto with Biometrics Effectively //IEEE TRANSACTIONS ON COMPUTERS, VOL. 55, NO. 9, SEPTEMBER, Page(s):1073 – 1074, 2006.
4. Язов Ю.К. и др. Нейросетевая защита персональных биометрических данных. //Ю.К. Язов, В.И. Волчихин, А.И. Иванов, В.А. Фунтиков, И.Г. Назаров // М.: Радиотехника. - 2012. - 157 с. IBSN 978-5-88070-044-8.
5. Иванов А.И. Биометрическая идентификация личности по динамике подсознательных движений. /А.И. Иванов. Монография. – Пенза: ПГУ.- 2000.-178 с.
6. Гудфеллоу Я., Бенджио И., Курвиль А. Глубокое обучение. М.: ДМК Пресс. 2017. - 652 с. IBSN 978-597060-554-7
7. Иванов А.И. Нейросетевые технологии биометрической аутентификации пользователей открытых систем. Автореферат диссертации на соискание ученой степени доктора технических наук по специальности 05.13.01 «Системный анализ, управление и обработка информации». - Пенза. - 2002. 34 с.
8. ГОСТ Р 52633.5-2011 «Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа».
9. Кобзарь А. И. Прикладная математическая статистика для инженеров и научных работников. М.: ФИЗМАТЛИТ, 2006. 816 с.
10. ГОСТ Р 50.1.037-2002. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Ч. 1. Критерии типа хи-квадрат. Госстандарт России. М., 2001.140 с.
11. ГОСТ Р 50.1.037-2002 Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Ч. 2. Непараметрические критерии. Госстандарт России. М., 2002.123 с.
12. Подавление шумов квантования биометрических данных при использовании многомерного критерия Крамера-фон Мизеса / А.И. Иванов, Газин А.И., Вятчанин С.Е. и др. // Проблемы информационной безопасности. Компьютерные системы. Санкт Петербург: ПТУ 2016. № 2. С. 21-28.
13. Иванов А.И., Малыгина Е.А., Перфилов П.А., Вятчанин С.Е. Сравнение мощности критерия среднего геометрического и Крамера-фон Мизеса на малых выборках

биометрических данных. // Модели, системы, сети в экономике, технике, природе и обществе. №2 2016, с 155-158.
14. Волчихин В.И. Перспективы использования искусственных нейронных сетей с многоуровневыми квантователями в технологии биометрико-нейросетевой аутентификации/В.И.Волчихин, А.И.Иванов, В.А.Фунтиков, Е.А. Малыгина //Известия высших учебных заведений. Поволжский регион. Технические науки. – Пенза: Изд-во ПГУ, 2013. №4(28) С.88 – 99.
15. Ахметов Б.С. Алгоритмы обучения и тестирования нейросетевых преобразователей биометрия-код /монография/ Б.С. Ахметов, В.И. Волчихин, А.И. Иванов, В.А. Фунтиков, Е.А. Малыгина//Республика Казахстан, Алматы: Изд-во КазНТУ, 2014 – 136 с.

Сведения об авторах:

Вятчанин Сергей Евгеньевич, доцент, начальник кафедры института Военного обучения ФБГОУ ВО «Пензенский государственный университет», Российская Федерация, 440026, г. Пенза, ул. Красная д. 40. E-mail: vyt5@list.ru

Малыгина Елена Александровна – кандидат технических наук, научный сотрудник межотраслевой лаборатории тестирования биометрических устройств и технологий, ФГБОУ ВО «Пензенский государственный университет», Российская Федерация, 440026, г. Пенза, ул. Красная д. 40. E-mail: mal890@yandex.ru