



Нейросетевой анализ нормальности малых выборок биометрических данных с использованием хи-квадрат критерия и критериев Андерсона – Дарлинга

В. И. Волчихин¹, А. И. Иванов^{2*}, А. В. Безяев¹,
Е. Н. Куприянов¹

¹ФГБОУ ВО «Пензенский государственный университет»
(г. Пенза, Россия)

²АО «Пензенский научно-исследовательский
электротехнический институт» (г. Пенза, Россия)

*ivan@pniei.penza.ru

Введение. Целью статьи является снижение требований к размерам тестовой выборки при проверке гипотезы нормальности.

Материалы и методы. Используется нейросетевое обобщение трех известных статистических критериев: хи-квадрат критерия, критерия Андерсона – Дарлинга в обычной форме и критерия Андерсона – Дарлинга в логарифмической форме.

Результаты исследования. Нейросетевое объединение хи-квадрат критерия и критерия Андерсона – Дарлинга позволяет снизить требования к объему выборки приблизительно на 40 %. Добавление третьего нейрона, воспроизводящего логарифмический вариант критерия Андерсона – Дарлинга, приводит к незначительному снижению вероятности ошибок (2 %). В статье рассматриваются однослойная и многослойная нейронные сети, обобщающие множество известных на данный момент статистических критериев.

Обсуждение и заключение. Высказано предположение о том, что каждому из известных статистических критериев может быть поставлен в соответствие искусственный нейрон. Необходимо изменить отношение к синтезу новых статистических критериев, господствовавшее в XX в. В настоящее время отсутствует необходимость стремиться к созданию статистических критериев высокой мощности. Гораздо более выгодными являются попытки обеспечить низкую коррелированность данных вновь синтезируемых статистических критериев с множеством уже созданных критериев.

Ключевые слова: критерий хи-квадрат, критерий Андерсона – Дарлинга, искусственная нейронная сеть, статистический критерий, нейросетевое воспроизведение статистических критериев, нейросетевой анализ, малая выборка

Для цитирования: Нейросетевой анализ нормальности малых выборок биометрических данных с использованием хи-квадрат критерия и критериев Андерсона – Дарлинга / В. И. Волчихин [и др.] // Инженерные технологии и системы. 2019. Т. 29, № 2. С. 205–217. DOI: <https://doi.org/10.15507/2658-4123.029.201902.205-217>



The Neural Network Analysis of Normality of Small Samples of Biometric Data through Using the Chi-Square Test and Anderson–Darling Criteria

V. I. Volchikhin¹, A. I. Ivanov^{2*}, A. V. Bezyaev¹,
E. N. Kupriyanov¹

¹*Penza State University (Penza, Russia)*

²*Penza Research Electrotechnical Institute (Penza, Russia)*

**ivan@pniei.penza.ru*

Introduction. The aim of the work is to reduce the requirements to test sample size when testing the hypothesis of normality.

Materials and Methods. A neural network generalization of three well-known statistical criteria is used: the chi-square criterion, the Anderson–Darling criterion in ordinary form, and the Anderson–Darling criterion in logarithmic form.

Results. The neural network combining of the chi-square criterion and the Anderson–Darling criterion reduces the sample size requirements by about 40 %. Adding a third neuron that reproduces the logarithmic version of the Anderson–Darling test leads to a small decrease in the probability of errors by 2 %. The article deals with single-layer and multi-layer neural networks, summarizing many currently known statistical criteria.

Discussion and Conclusion. An assumption has been made that an artificial neuron can be assigned to each of the known statistical criteria. It is necessary to change the attitude to the synthesis of new statistical criteria that previously prevailed in the 20th century. There is no current need for striving to create statistical criteria for high power. It is much more advantageous trying to ensure that the data of newly synthesized statistical criteria are low correlated with many of the criteria already created.

Keywords: chi-square test, Anderson–Darling criterion, artificial neural network, statistical criterion, neural network reproduction of statistical criteria, neural network analysis, small sample

For citation: Volchikhin V.I., Ivanov A.I., Bezyaev A.V., Kupriyanov E.N. The Neural Network Analysis of Normality of Small Samples of Biometric Data through Using the Chi-Square Test and Anderson–Darling Criteria. *Inzhenernyye tekhnologii i sistemy* = Engineering Technologies and Systems. 2019; 29(2):205-217. DOI: <https://doi.org/10.15507/2658-4123.029.201902.205-217>

Введение

Экспериментально полученных данных часто бывает недостаточно. Например, медику, чтобы получить данные 20 пациентов с одним и тем же заболеванием, может потребоваться лечебная практика длительностью 1 год и более. При этом проверить нормальность данных на столь малых выборках невозможно. Если пользоваться стандартными статистическими рекомендациями¹, то

для проверки гипотезы нормальности потребуется выборка из 200 и более опытов. Обычный врач сможет самостоятельно получить такое количество данных не менее чем за 10 лет практики.

Схожая ситуация возникает у биологов, а также при сборе биометрических данных. И медики, и биологи прежде были вынуждены ждать, когда у них появится достаточно большая выборка. Ситуация начала меняться с развитием

¹ Кобзарь А. И. Прикладная математическая статистика. Для инженеров и научных работников. М. : Физматлит, 2006. 816 с.; Р 50.1.033–2001. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Ч. 1. Критерии типа хи-квадрат. Введ. 2002.07.01. М. : Изд-во стандартов, 2001. 87 с. URL: <https://pdf.standartgost.ru/catalog/Data2/1/4293850/4293850693.pdf>

биометрии. Люди негативно относятся к тому, что при обучении нейронных сетей им придется 200 раз приложить палец к сканеру или 200 раз записать рукописный пароль. Однако пользователи согласны приложить к сканеру папиллярного рисунка свой палец 20 раз или 20 раз записать пароль. По этой причине в нейросетевой биометрии серьезное внимание уделяется разработке новых методов, позволяющих проверять статистические гипотезы на выборках малого объема.

Обзор литературы

В 1900 г. К. Пирсон создал хи-квадрат критерий, который на целое столетие определил вектор развития теории математической статистики². В середине XX в. появились первые исследования, посвященные синтезу искусственных нейронных сетей, ориентированных на эффективное распознавание образов. Вторая половина XX в. была посвящена тому, чтобы добиться эффективного обучения искусственных нейронных сетей. Только в начале XXI в. удалось достичь промышленного применения технологии обучения искусственных нейронных сетей. Каждый цифровой фотоаппарат сейчас имеет нейросетевое приложение, отыскивающее в кадре лица людей³. Автоматическое обучение искусственных нейронных сетей нейросетевых преобразователей биометрических образов в код криптографического ключа аутентификации регламентируется в России национальным стандартом⁴. Биометрия – особый сегмент рынка, подлежащий специальному регулированию. На данный момент международное сообщество разработало порядка 120 биометрических стандартов, из которых 44 адаптированы к русскому языку и введены в действие в России.

Хи-квадрат критерий и другие известные статистические критерии при проверке статистических гипотез ориентированы на сравнение экспериментальной функции вероятности $P(x)$ с теоретической функцией вероятности $P(x)$. Фактически все статистические критерии решают ту же задачу, что и обычные нейроны, то есть для каждого из известных статистических критериев может быть построен свой искусственный нейрон. В частности, для хи-квадрат критерия эквивалентный искусственный нейрон будет описываться следующими преобразованиями для выборки объемом в 21 опыт при гистограмме из 5 равных интервалов:

$$\left\{ \begin{array}{l} x \leftarrow \text{sort}(x) \\ \chi^2 \leftarrow \sum_{i=1}^5 \frac{\left\{ \frac{n_i}{21} - (P(x_{i+1}) - P(x_i)) \right\}^2}{(P(x_{i+1}) - P(x_i))}, \quad (1) \\ z \leftarrow "0" \text{ if } \chi^2 \leq 0.375 \\ z \leftarrow "1" \text{ if } \chi^2 > 0.375 \end{array} \right.$$

где n_i – число опытов, попавших в i -й интервал гистограммы.

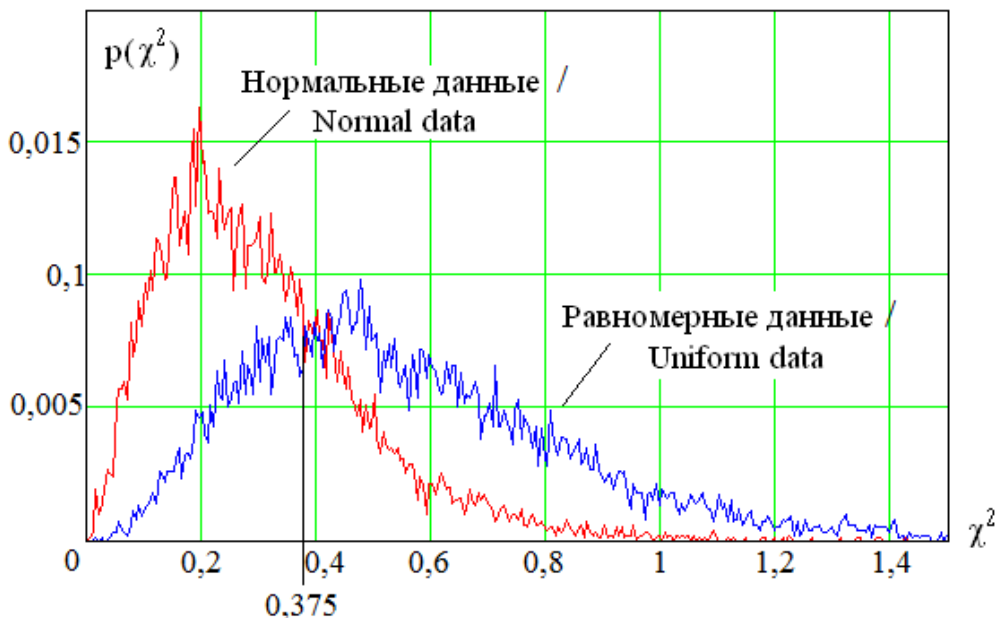
В работоспособности квадратичного нейрона (1) можно убедиться благодаря результатам имитационного моделирования, приведенным на рис. 1.

Результаты (рис. 1) показывают, что отклики нейрона на данные с нормальным законом распределения и отклики на данные с равномерным законом существенно отличаются. При квантовании выходных данных сумматора нейрона по порогу 0,375 вероятности ошибок первого и второго рода оказываются равными: $P_1 = P_2 = P_{EE} = 0,288$.

² Там же.

³ Гудфеллоу Я., Бенджио И., Курвиль А. Глубокое обучение. М.: ДМК Пресс, 2017. 652 с.

⁴ ГОСТ Р 52633.5–2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа. Введ. 2012.04.01. М.: СТАНДАРТИНФОРМ, 2012. 16 с. URL: <http://docs.cntd.ru/document/1200088764>



Р и с. 1. Отклик хи-квадрат нейрона при выборке в 21 опыт для данных с нормальным и равномерным распределениями

Fig. 1. The response to the chi-square of a neuron when sampling 21 experiments for data with normal and uniform distributions

Идеи К. Пирсона были активно поддержаны математическим сообществом; в XX в. были созданы десятки различных статистических критериев⁵. В частности, в 1952 г. был разработан критерий Андерсона – Дарлингга. Нейрон, эквивалентный данному критерию, описывается следующими преобразованиями:

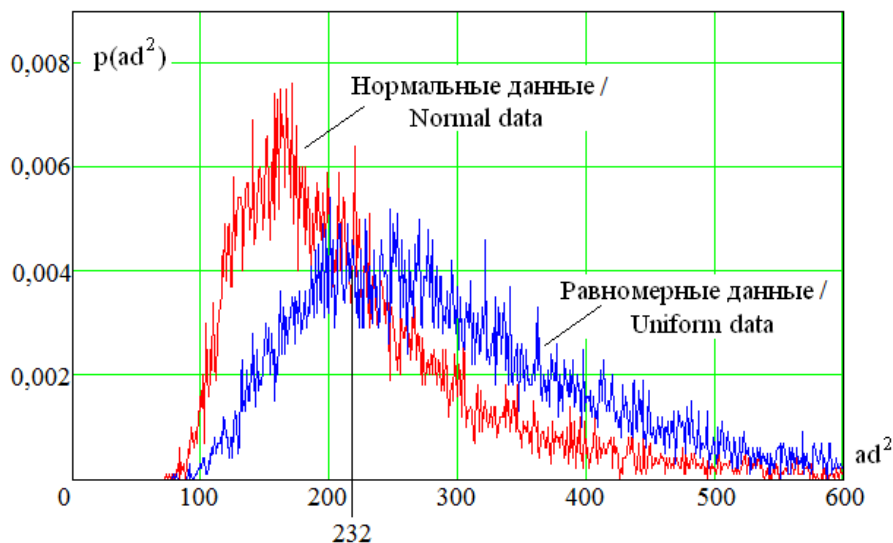
$$\left\{ \begin{array}{l} x \leftarrow \text{sort}(x) \\ ad^2 \leftarrow \sum_{i=1}^{21} \frac{\{i - 21 \cdot P(x_i)\}^2}{(1 - P(x_i)) \cdot P(x_i)}, \quad (2) \\ z \leftarrow "0" \text{ if } ad^2 \leq 232 \\ z \leftarrow "1" \text{ if } ad^2 > 232 \end{array} \right.$$

Об эффективности работы нейрона Андерсона – Дарлингга можно судить по данным имитационного моделирования, приведенным на рис. 2.

Видно, что отклики нейрона на данные с нормальным законом распределения и отклики на данные с равномерным законом существенно отличаются. При квантовании выходных данных сумматора нейрона по порогу 232 вероятности ошибок первого и второго рода оказываются равными: $P_1 = P_2 = P_{EE} = 0,349$.

Очевидно, что для малой выборки (21 опыт) мощность хи-квадрат критерия выше мощности критерия Андерсона – Дарлингга примерно на 21 %. Подобный эффект достигается выполнением следующего неравенства:

⁵ Кобзарь А. И. Прикладная математическая статистика. Для инженеров и научных работников. М. : Физматлит, 2006. 816 с.; Р 50.1.037-2002. Прикладная статистика. Правила проверки согласия опытного распределения с теоретическим. Ч. 2. Непараметрические критерии. Введ. 2012.04.01. М. : Изд-во стандартов, 2002. 62 с. URL: <http://www.gost.rf.com/normadata/1/4294846/4294846286.pdf>



Р и с. 2. Отклик нейрона Андерсона – Дарлинга при выборке в 21 опыт для данных с нормальным и равномерным распределениями

Fig. 2. The response of the Anderson–Darling neuron in a sample of 21 experiments for data with normal and uniform distributions

$P_{EE}(\chi^2) = 0,288 < P_{EE}(ad^2) = 0,349$. Если руководствоваться принципом бритвы Оккама, нужно использовать только хи-квадрат критерий, отбросив данные критерия Андерсона – Дарлинга. В настоящее время такой подход нерационален, так как мы можем совместно использовать два нейрона, как это принято в российской нейросетевой биометрии⁶ [1; 2]. Такая возможность обусловлена низким уровнем коррелированности выходных данных двух рассматриваемых критериев: $corr(\chi^2, ad^2) = 0,423$.

Если мы объединяем два нейрона, то они должны принимать состояния «00» и «01» для признания данных нормальными. Два других состояния («11» и «10») будем трактовать как признание предъявленных данных равномерными. В представленной ситуации вероятность ошибок снижается примерно на 20 %: с 0,288 до 0,239.

Соотношение (2) построено с опорой на аналитическую запись критерия Андерсона – Дарлинга, однако может быть выполнен переход к логарифмической форме записи данного критерия. Нейрон, соответствующий логарифмической форме записи, описывается системой следующих преобразований:

$$\left\{ \begin{array}{l} x \leftarrow \text{sort}(x) \\ adL \leftarrow \sum_{i=0}^{20} \left\{ \frac{i+1}{21} \cdot \ln(P(x_i)) + \right. \\ \left. + \left(1 - \frac{i+1}{21} \right) \cdot \ln(1 - P(x_i)) \right\} \cdot (3) \\ z \leftarrow "0" \text{ if } adL \leq 10.705 \\ z \leftarrow "1" \text{ if } adL > 10.705 \end{array} \right.$$

⁶ Иванов А. И. Нейросетевые алгоритмы биометрической идентификации личности. Серия «Нейрокомпьютеры и их применение». Кн. 15. М.: Радиотехника, 2004 г. 144 с.; Нейросетевая защита персональных биометрических данных // Ю. К. Язов [и др.] // М.: Радиотехника, 2012. 160 с.

На рис. 3 даны распределения откликов нейрона логарифмической формы Андерсона – Дарлингга.

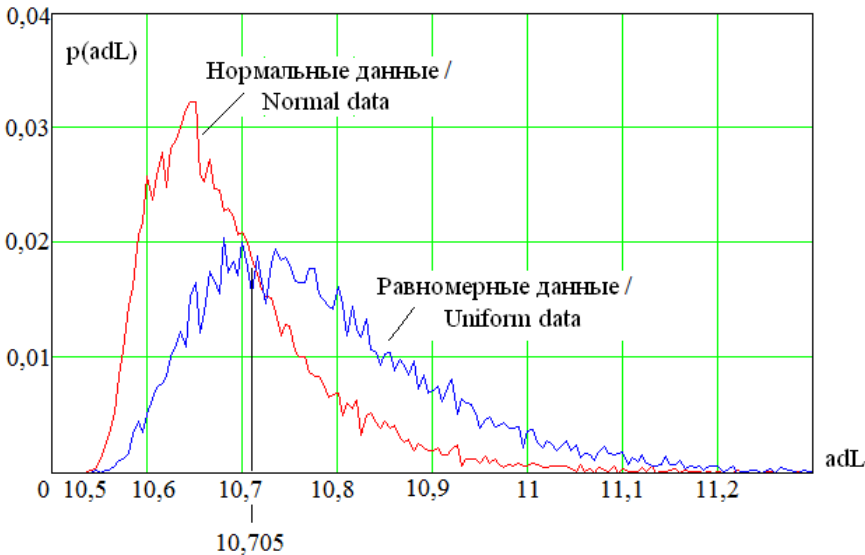
При значении порога квантования 10,705 совпадают значения вероятностей ошибок первого и второго рода: $P_1 = P_2 = P_{EE} = 0,32$. Значение данной величины меньше аналогичного значения нейросетевой реализации классической формы Андерсона – Дарлингга (2). Кроме того, данные две формы нейронов имеют в своих откликах независимые составляющие: $corr(ad^p, adL) = 0,644$. Фактически мы имеем два разных статистических критерия, дополняющих друг друга.

Объединенное использование хи-квадрат нейрона (1) и логарифмической формы нейрона (2) позволяет снизить вероятность ошибок до 0,236. Одновременное использование двух вариантов критерия Андерсона – Дарлингга и хи-квадрат критерия дает дополнительное, но незначительное снижение вероятности ошибок до 0,231.

Материалы и методы

Средневековый принцип бритвы Оккама является оправданием множества процедур обхода «проклятия» размерности задачи. Например, это может быть сделано методом главных компонент, созданным К. Пирсоном в 1901 г. Метод главных компонент является практической реализацией доказанной в 1947 г. теоремы Карунена – Лоэва⁷ [3].

Обоснование бритвы Оккама методом главных компонент кажется оптимальным, однако оно актуально лишь для линейных функционалов (только для линейной алгебры). Если мы строим нелинейную алгебру нейросетевых функционалов, все оказывается сложнее, что мы и постарались продемонстрировать. Бритва Оккама требует отбросить два варианта критериев Андерсона – Дарлингга и пользоваться только хи-квадрат критерием как самым информативным. Однако объединение всех трех крите-



Р и с. 3. Отклик логарифмической формы нейрона Андерсона – Дарлингга при выборке в 21 опыт для данных с нормальным и равномерным распределениями

Fig. 3. The response of the logarithmic form of the Anderson–Darling neuron in a sample of 21 experiments for data with normal and uniform distributions

⁷ Jolliffe I. T. Principal component analysis. 2nd ed. Series: Springer Series in Statistics. Springer-Verlag, 2002. 487 p. URL: [http://cda.psych.uiuc.edu/statistical_learning_course/Jolliffe%201.%20Principal%20Component%20Analysis%20\(2ed.,%20Springer,%202002\)\(518s\)_MVsa_.pdf](http://cda.psych.uiuc.edu/statistical_learning_course/Jolliffe%201.%20Principal%20Component%20Analysis%20(2ed.,%20Springer,%202002)(518s)_MVsa_.pdf)

риев одним слоем искусственных нейронов и корректировка выходного кода дает положительный результат: данные, которые необходимо было отбросить, удается объединить. Обобщенная структура этого направления синтеза новых процедур статистической обработки отображена на рис. 4.

Могут быть построены десятки нейронов, воспроизводящих работу известных и новых статистических критериев. При этом каждый из нейронов будет обладать своей относительной мощностью разделения нормальных и равномерных данных. Для статистических критериев, уже имеющих программные нейроны-аналоги, данные имитационного моделирования приведены в табл. 1.

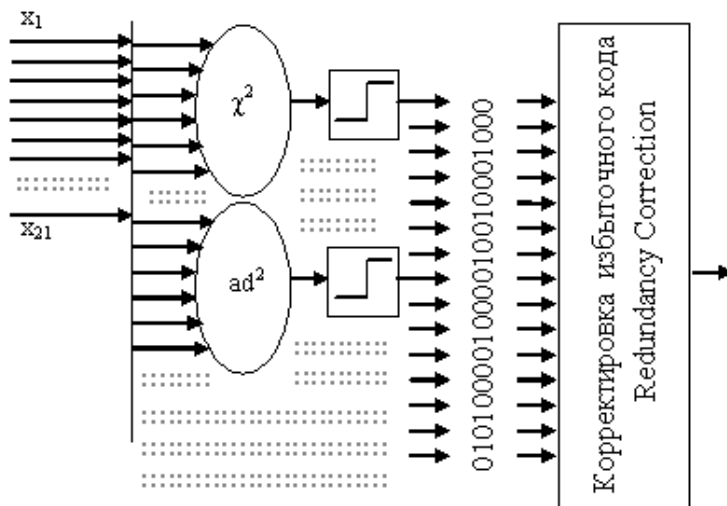
Данные в табл. 1 показывают, что наиболее информативным является

дифференциальный вариант критерия среднего геометрического⁸:

$$\begin{cases} x \leftarrow \text{sort}(x) \\ sg_d \leftarrow \sum_{i=1}^{21} \sqrt{\frac{i+1}{21} \cdot p(x_i)} \\ z \leftarrow "0" \text{ if } sg_d \leq 7.45 \\ z \leftarrow "1" \text{ if } sg_d > 7.45 \end{cases}, \quad (4)$$

где $p(x_i)$ – значение плотности распределения или производная $\frac{\partial P(x_i)}{\partial x}$ функции вероятности.

Вторым по уровню информативности оказывается хи-квадрат нейрон (1), третьим – нейрон Смирнова – Крамера – фон Мизеса:



Р и с. 4. Обобщение большого числа статистических критериев однослойной сетью искусственных нейронов (аналогия с нейросетевыми преобразователями биометрия-код, выполненными в соответствии с российскими стандартами пакета ГОСТ Р 52633.xx)

F i g. 4. Generalizing of a large number of statistical criteria with a single-layer network of artificial neurons (analogy with neural network biometrics-code converters made in accordance with the Russian standards of the GOST R 52633.xx package)

⁸ **Иванов А. И., Перфилов К. А.** Оценка соотношения мощностей семейства статистических критериев «среднего геометрического» на малых выборках биометрических данных // Современные охраняемые технологии и средства обеспечения комплексной безопасности объектов : мат-лы XI Всерос. науч.-практ. конф. Пенза, 2016. С. 223–229.

Данные о значениях ошибок первого и второго рода для различных статистических критериев с оценкой порядка их следования по уровню их информативности

Data on the values of the errors of the first and second kinds for various statistical criteria with an assessment of the order of their sequence in terms of their informativeness

Критерий / Criterion	P_{EE}	Информативность / Informativeness
χ^2	0,288	2
$ad^2(2)$	0,349	6
$adL(3)$	0,320	4
sg	0,320	4
sg_d	0,278	1
ω^2	0,351	7
ω_c^2	0,311	3
su^2	0,322	5

$$\left\{ \begin{array}{l} x \leftarrow \text{sort}(x) \\ \omega_c \leftarrow \sum_{i=1}^{21} \left\{ \frac{i+1}{21} - (P(x_i)) \right\}^2 \\ z \leftarrow "0" \text{ if } sg_d \leq 31.6 \\ z \leftarrow "1" \text{ if } sg_d > 31.6 \end{array} \right. \quad (5)$$

$$\left\{ \begin{array}{l} x \leftarrow \text{sort}(x) \\ su^2 \leftarrow \frac{1}{\sigma(x)} \left[\sum_{i=0}^9 a_i \cdot (x_{21-i} - x_i) \right]^2 \\ z \leftarrow "0" \text{ if } su^2 \leq 19.88 \\ z \leftarrow "1" \text{ if } su^2 > 19.88 \end{array} \right. \quad (7)$$

Четвертое место по информативности принадлежит логарифмической форме нейрона Андерсона – Дарлингга (3) и интегральному варианту критерия среднего геометрического⁹ [4]:

$$\left\{ \begin{array}{l} x \leftarrow \text{sort}(x) \\ sg \leftarrow \sum_{i=1}^{21} \sqrt{\frac{i+1}{21}} \cdot (1 - P(x_i)) \\ z \leftarrow "0" \text{ if } sg \leq 8.52 \\ z \leftarrow "1" \text{ if } sg > 8.52 \end{array} \right. \quad (6)$$

Пятое место по информативности независимого применения занимают нейроны Шапиро – Уилка¹⁰:

Наиболее низкой информативностью из рассмотренных обладает нейрон, воспроизводящий наиболее раннюю версию критерия Крамера – фон Мизеса, созданную в 1928 г.¹¹:

$$\left\{ \begin{array}{l} x \leftarrow \text{sort}(x) \\ \omega^2 \leftarrow \sum_{i=0}^{20} \left[\left\{ P(x_i) - \frac{i+1}{21} \right\}^2 \cdot \frac{x_{i+1} - x_i}{x_{20} - x_0} \right] \\ z \leftarrow "0" \text{ if } \omega^2 \leq 0.0043 \\ z \leftarrow "1" \text{ if } \omega^2 > 0.0043 \end{array} \right. \quad (8)$$

Принципиально важным является то, что на данный момент созданы десятки различных статистических

⁹ Использование среднего геометрического, ожидаемой и наблюдаемой функций вероятности как статистического критерия оценки качества биометрических данных / Б. С. Ахметов [и др.] // Надежность и качество : тр. XX междунар. симпозиума : в 2 т. Пенза : Изд-во Пенз. гос. ун-та, 2015. Т. 2. С. 283–285.

¹⁰ Кобзарь А. И. Прикладная математическая статистика. Для инженеров и научных работников. М. : Физматлит, 2006. 816 с.

¹¹ Там же.

критериев, то есть длина выходного избыточного кода нейронной сети (рис. 4) достаточно велика. Если подавать нормальные данные на нейронную сеть, она будет откликаться кодами с большинством нулевых разрядов, то есть корректоры редких ошибок избыточных кодов оказываются достаточно примитивными. При их создании может быть использован опыт разработки подобных корректоров кода для биометрических приложений [5–8].

Во времена Пирсона не существовало вычислительных машин и искусственных нейронов. В настоящее время ситуация изменилась коренным образом: могут быть созданы программные продукты, комплексно учитывающие десятки и даже сотни статистических критериев. Все это позволяет ожидать появления в ближайшем будущем статистических программ, позволяющих с очень высокой точностью выполнять прогнозы на малых выборках. Если сейчас для достоверного принятия решений о нормальности закона распределения требуются выборки в 200 опытов, то в скором времени появятся прецизионные средства статистических оценок для выборок в 20 опытов. Все это является следствием перехода от попыток обобщений в линейных пространствах к многомерным нейросетевым обобщениям.

Результаты исследования

Следует отметить, что удовлетворительной теории кодов, корректирующих ошибки искусственных нейронных сетей, пока не существует, а классические самокорректирующиеся коды дают низкие результаты. Вместо создания новой теории корректирующих свойств гибрида нейронной сети и избыточного самокорректирующегося кода необходимо переходить от однослойных нейронных сетей к многослойным.

Следует отметить, что впервые многослойные нейронные сети начал обу-

чать А. И. Галушкин¹², создав в 1974 г. метод обратного распространения ошибок. На тот момент реализовать данный метод для многослойных нейронных сетей было технически невозможно. Ситуация изменилась только через 10 лет, когда Дж. Хинтон предложил обучать первые слои нейронов машинами Больцмана [9]. В конечном итоге это и позволило получить в XXI в. промышленно применимый алгоритм обучения глубоких нейронных сетей [10]. В таком контексте промышленно применяемые алгоритмы обучения можно называть алгоритмом Галушкина – Хинтона. Их применение позволяет обучать любые «глубокие» нейронные сети¹³ (например, сеть со структурой, приведенной на рис. 5).

Сеть на рис. 5 имеет три слоя нейронов, при этом нейроны первого и второго слоя должны быть гладкими. Только один выходной нейрон всей «глубокой» сети имеет полностью дискретный квантователь.

Потенциальное качество работы нейронов второго слоя «глубокой» сети зависит от числа входов нейрона, от качества его входных данных и от их взаимной коррелированности.

Значения взаимной корреляции нейронов первого слоя приведены в табл. 2.

При оптимизации выбора связей нейронов необходимо стремиться к минимизации среднего геометрического вероятностей ошибок первого и второго рода $G(P_{EE})$ (данные из табл. 1). Одновременно необходимо минимизировать среднее значение модулей корреляционных входных связей $\bar{E}(|corr|)$ (данные из табл. 2). Балансировка двух данных факторов для нейронов с двумя входами выполняется следующим образом:

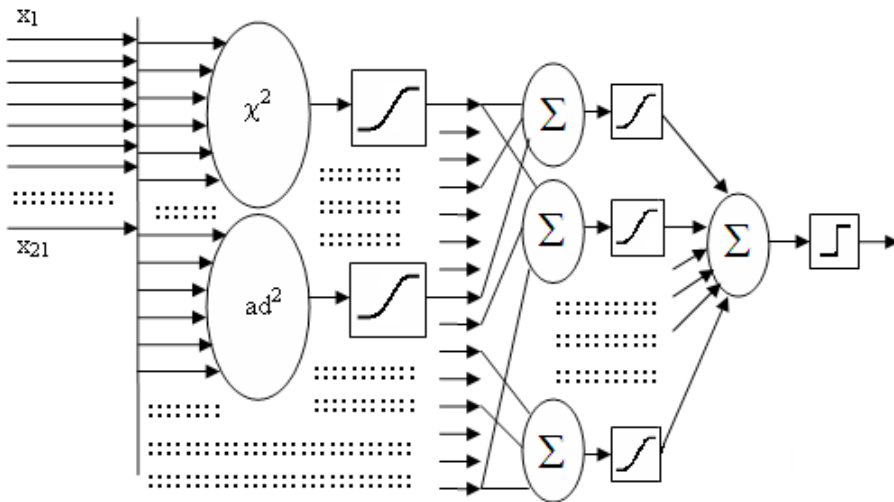
$$\min \left\{ \frac{0.91 \cdot G(P_{EE})^{2-E(|corr|)} + 1.09 \cdot G(P_{EE})^{2-(E|corr|)^2}}{2} \right\}. \quad (9)$$

¹² Галушкин А. И. Синтез многослойных систем распознавания образов. М.: Энергия, 1974. 368 с.

¹³ Гудфеллоу Я., Бенджио И., Курвиль А. Глубокое обучение. М.: ДМК Пресс, 2017. 652 с.

Коэффициенты корреляции между парами рассматриваемых статистических критериев
The coefficients of correlation between pairs of statistical criteria under consideration

	χ^2	ad^2	adL	sg	sg_d	ω^2	ω_c^2	su^2
χ^2	1	0,423	0,672	0,037	-0,042	0,559	0,401	-0,726
ad^2	0,423	1	0,644	0,018	-0,145	0,226	0,393	-0,113
adL	0,672	0,644	1	0,056	0,209	0,827	0,832	-0,917
sg	0,037	0,018	0,056	1	0,132	0,414	0,402	-0,212
sg_d	-0,042	-0,145	0,209	0,132	1	-0,242	-0,142	-0,041
ω^2	0,559	0,226	0,827	0,414	-0,242	1	0,885	-0,667
ω_c^2	0,401	0,393	0,832	0,402	-0,142	0,885	1	-0,764
su^2	-0,726	-0,113	-0,917	-0,212	-0,041	-0,667	-0,764	1



Р и с. 5. Обобщение большого числа статистических критериев многослойной сетью искусственных нейронов (аналогия с использованием глубоких нейронных сетей для распознавания лиц)

Fig. 5. Generalizing a large number of statistical criteria by a multilayered network of artificial neurons (analogy with the use of deep neural networks for face recognition)

Для нейронов с тремя входами связывание двух факторов описывается близким соотношением:

$$\min \left\{ \frac{1.75 \cdot G(P_{EE})^{3-1.5E(korr)} + 0.25 \cdot G(P_{EE})^{3-1.5(E(korr))^2}}{2} \right\}. \quad (10)$$

При необходимости могут быть построены аналогичные соотношения для нейронов с любым числом входов.

Обсуждение и заключение

Следует отметить, что усилия сотен исследователей, развивавших математическую статистику в XX в., несопоставимы по размерам ресурсов, затраченных США, Канадой, Евросоюзом, Китаем, Россией на развитие технологий биометрической идентификации личности. С начала XXI в. совместными усилиями технологических лидеров в рамках международного комитета

по стандартизации ISO/IEC JTC1 sc 37 (Биометрия) было создано и введено в действие порядка 120 международных стандартов; в стадии разработки находится около 30 стандартов. В России 44 из 120 международных стандартов гармонизованы и введены в действие. В настоящее время развиваются несколько технологий сканирования и обработки уникальных биометрических образов. Данные по биометрическим технологиям отражены в табл. 1.

После создания хи-квадрат критерия К. Пирсоном в 1900 г. его последователями было разработано несколько десятков подобных статистических критериев. Можно рассматривать весь XX в. как время синтеза новых статистических критериев. Видимо, в XXI в. тенденция синтеза новых статистических критериев продолжится, однако акценты развития статистической теории существенно изменятся. В данной статье мы показали, что эффективным является нейросетевое

обобщение уже созданных критериев и критериев, которые еще будут разработаны. Существующих технологий обучения искусственных нейронных сетей вполне достаточно. В данном контексте крайне важным является создание новых статистических критериев, слабо коррелирующих с уже известными. Становится не столь важной относительная мощность нового статистического критерия; важнее то, как он дополняет другие критерии в нейросетевых обобщениях.

В рамках данной статьи мы показали, что нейросетевое объединение всего трех статистических критериев позволяет снизить ошибки первого и второго рода на 22 %. Предположительно, обобщение порядка 100 статистических критериев глубокой нейронной сетью позволит многократно снизить требования к размерам тестовых выборок. Вместо 200 опытов, необходимых сейчас, в ближайшем будущем будет достаточно выборки в 20 опытов.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. **Иванов А. И.** Биометрическая идентификация личности по динамике подсознательных движений : монография. Пенза : Изд-во Пенз. гос. ун-та, 2000. 188 с.
2. **Волчихин В. И., Иванов А. И., Фунтиков В. А.** Быстрые алгоритмы обучения нейросетевых механизмов биометрико-криптографической защиты информации : монография. Пенза : Изд-во Пенз. гос. ун-та, 2005. 273 с.
3. Principal manifolds for data visualization and dimension reduction / Eds. A. N. Gorban [et al.]. Lecture Notes in Computational Science and Engineering. Springer, 2007. Vol. 58. 340 p. DOI: <https://doi.org/10.1007/978-3-540-73750-6>
4. **Иванов А. И., Перфилов К. А., Малыгина Е. А.** Многомерный статистический анализ качества биометрических данных на предельно малых выборках с использованием критериев среднего геометрического, вычисленного для анализируемых функций вероятности // Измерение. Мониторинг. Управление. Контроль. 2016. № 2 (16). С. 64–72. URL: <https://imuk.pnzgu.ru/files/imuk.pnzgu.ru/9216.pdf>
5. **Безяев А. В.** Нейросетевой преобразователь биометрии в самокорректирующийся код, совершенно не обладающий избыточностью // Нейрокомпьютеры: разработка, применение. 2012. № 3. С. 52–56. URL: <http://www.radiotec.ru/article/10421>
6. **Безяев А. В., Иванов А. И., Фунтикова Ю. В.** Оптимизация структуры самокорректирующегося био-кода, хранящего синдромы ошибок в виде фрагментов хеш-функций // Вестник УРФО. Безопасность в информационной сфере. 2014. № 3 (13). С. 4–13. URL: http://www.info-sec.ru/is_13/bezyaev.htm
7. Оценка эффекта ускорения вычислений, обусловленного поддержкой квантовой суперпозиции при корректировке выходных состояний нейросетевого преобразователя биометрии в код /

В. И. Волчихин [и др.] // Известия высших учебных заведений. Поволжский регион. Технические науки. Информатика, вычислительная техника. 2017. № 1 (41). С. 43–55. URL: https://izvuz_tn.pnzgu.ru/files/izvuz_tn.pnzgu.ru/4117.pdf

8. **Иванов А. И.** Сопоставительный анализ показателей конкурирующих технологий биометрико-криптографической аутентификации личности // Защита информации. ИНСАЙД. 2014. № 3. С. 32–39. URL: http://www.inside-zi.ru/pages/3_2014/32.html

9. **Ackley D. H., Hinton G. E., Sejnowski T. J.** A learning algorithm for boltzmann machines // Cognitive Science. 1985. Vol. 9, issue 1. P. 147–169. DOI: [https://doi.org/10.1016/S0364-0213\(85\)80012-4](https://doi.org/10.1016/S0364-0213(85)80012-4)

10. **Hinton G. E.** Training products of experts by minimizing contrastive divergence // Neural Computation. 2002. Vol. 14, issue 8. P. 1771–1800. DOI: <https://doi.org/10.1162/089976602760128018>

Поступила 15.02.2019; принята к публикации 25.04.2019; опубликована онлайн 28.06.2019

Об авторах:

Волчихин Владимир Иванович, президент, ФБГОУ ВО «Пензенский государственный университет» (440026, Россия, г. Пенза, ул. Красная, 40), доктор технических наук, профессор, ResearcherID: O-9718-2015, ORCID: <https://orcid.org/0000-0002-9986-521X>, vvi@pnzgu.ru

Иванов Александр Иванович, начальник, лаборатория биометрических и нейросетевых технологий, АО «Пензенский научно-исследовательский электротехнический институт» (440000, Россия, г. Пенза, ул. Советская, д. 9), доктор технических наук, доцент, ResearcherID: R-4514-2019, ORCID: <https://orcid.org/0000-0003-3475-2182>, ivan@pniei.penza.ru

Безяев Александр Викторович, докторант, кафедра информационной безопасности систем и технологий, ФБГОУ ВО «Пензенский государственный университет» (440026, Россия, г. Пенза, ул. Красная, д. 40), кандидат технических наук, ResearcherID: Q-9589-2019, ORCID: <https://orcid.org/0000-0003-0703-3270>, ibst@pnzgu.ru

Куприянов Евгений Николаевич, аспирант, кафедра информационной безопасности систем и технологий, ФБГОУ ВО «Пензенский государственный университет» (440026, Россия, г. Пенза, ул. Красная, д. 40), Publons: <https://publons.com/researcher/2956834/evgenyi-kupriyanov>, ORCID: <https://orcid.org/0000-0003-0806-1476>, ibst@pnzgu.ru

Заявленный вклад соавторов:

В. И. Волчихин – концепция нейросетевого объединения статистических критериев; А. И. Иванов – формализация нейросетевого описания рассматриваемых статистических критериев; А. В. Безяев – корректировка выходных кодов нейронной сети, обобщающей статистические критерии; Е. Н. Куприянов – программная реализация вычислений, формирование таблиц с данными.

Все авторы прочитали и одобрили окончательный вариант рукописи.

REFERENCES

1. Ivanov A.I. Biometric identification based on the dynamics of subconscious movements: a monograph. Penza: Publishing House of PSU; 2000. (In Russ.)

2. Volchikhin V.I., Ivanov A.I., Funtikov V.A. Fast learning algorithms for neural network mechanisms of biometric-cryptographic information protection: a monograph. Penza: Publishing House of PSU; 2005. (In Russ.)

3. Gorban A.N., Kégl B., Wunsch D.C., Zinovyev A.Y. Principal manifolds for data visualisation and dimension reduction. Lecture Notes in Computational Science and Engineering. Springer; 2007. Vol. 58. DOI: <https://doi.org/10.1007/978-3-540-73750-6>

4. Ivanov A.I., Perfilov K.A., Malygina E.A. Multivariate statistical analysis of the quality of biometric data on extremely small samples using the criteria of the geometric mean calculated for the analyzed

probability functions. *Izmereniye. Monitoring. Upravleniye. Kontrol* = Measuring. Monitoring. Management. Control. 2016; 2:64-72. Available at: <https://imuk.pnzgu.ru/files/imuk.pnzgu.ru/9216.pdf> (In Russ.)

5. Bezyaev A.V. Neural network biometrics – to-self-correcting code converter without redundancy. *Neyrokomp'yutery: razrabotka, primeneniye* = Neurocomputers. 2012; 3:52-56. Available at: <http://www.radiotec.ru/article/10421> (In Russ.)

6. Bezyaev A.V., Ivanov A.I., Funtikova Yu.V. Optimization of the structure self-correcting bio-code, storing syndromes error as fragments hash-functions. *Vestnik URFO. Bezopasnost v informatsionnoy sfere* = UrFR Newsletter. Information Security. 2014; 3:4-13. Available at: http://www.info-secur.ru/is_13/bezyaev.htm (In Russ.)

7. Volchikhin V.I., Ivanov A.I., Bezyaev A.V., Elfimov A.V., Yunin A.P. Evaluation of the calculation acceleration effect, caused by the support of quantum superposition states during adjustment of output conditions of a “biometrics – code” neural network converter. *Izvestiya vysshikh uchebnykh zavedeniy. Povolzhskiy region. Tekhnicheskie nauki. Informatika, vychislitel'naya tekhnika* = University Proceedings. Volga Region. Engineering Sciences. Computer Science, Computer Engineering and Control. 2017; 1:43-55. Available at: https://izvuz_tn.pnzgu.ru/files/izvuz_tn.pnzgu.ru/4117.pdf (In Russ.)

8. Ivanov A.I. Comparative analysis of the indicators of competing technologies of biometric-cryptographic authentication of the person. *Zashchita informatsii. INSAYD* = Protection of Information. INSIDE. 2014; 3:32-39. Available at: http://www.inside-zi.ru/pages/3_2014/32.html (In Russ.)

9. Ackley D.H., Hinton G.E., Sejnowski T.J. A learning algorithm for Boltzmann machines. *Cognitive Science*. 1985; 9(1):147-169. DOI: [https://doi.org/10.1016/S0364-0213\(85\)80012-4](https://doi.org/10.1016/S0364-0213(85)80012-4)

10. Hinton G.E. Training products of experts by minimizing contrastive divergence. *Neural Computation*. 2002; 14(8):1771-1800. DOI: <https://doi.org/10.1162/089976602760128018>

Received 15.02.2019; revised 25.04.2019; published online 28.06.2019

About authors:

Vladimir I. Volchikhin, President, Penza State University (40 Krasnaya St., Penza 440026, Russia), D.Sc. (Engineering), Professor, ResearcherID: O-9718-2015, ORCID: <https://orcid.org/0000-0002-9986-521X>, vvi@pnzgu.ru

Aleksandr I. Ivanov, Head, Laboratory of Biometric and Neural Network Technologies, Penza Research Electrotechnical Institute (9 Sovetskaya St., Penza 440000, Russia), D.Sc. (Engineering), Associate Professor, ResearcherID: R-4514-2019, ORCID: <https://orcid.org/0000-0003-3475-2182>, ivan@pniei.penza.ru

Alexander V. Bezyaev, Doctoral Candidate, Chair of Information Security of Systems and Technologies, Penza State University (40 Krasnaya St., Penza 440026, Russia), Ph.D. (Engineering), ResearcherID: Q-9589-2019, ORCID: <https://orcid.org/0000-0003-0703-3270>, ibst@pnzgu.ru

Evgeniy N. Kupriyanov, Graduate Student, Chair of Information Security of Systems and Technologies, Penza State University (40 Krasnaya St., Penza 440026, Russia), Publons: <https://publons.com/researcher/2956834/evgeniy-kupriyanov>, ORCID: <https://orcid.org/0000-0003-0806-1476>, ibst@pnzgu.ru

Contribution of the authors:

V. I. Volchikhin – the development of the concept of a neural network association of statistical criteria; A. I. Ivanov – the formalization of the neural network description of the statistical criteria under consideration; A. V. Bezyaev – the adjustment of the output codes of a neural network that generalizes statistical criteria; E. N. Kupriyanov – the software implementation of calculations and formation of tables with data.

All authors have read and approved the final version of the paper.