

ТЕСТИРОВАНИЕ КАЧЕСТВА НЕЙРОСЕТЕВОЙ КОРРЕКЦИИ ОШИБОК ВЫЧИСЛЕНИЯ СТАНДАРТНОГО ОТКЛОНЕНИЯ МАЛЫХ ВЫБОРОК БИОМЕТРИЧЕСКИХ ДАННЫХ

В. И. Волчихин¹, А. И. Иванов², Е. А. Малыгина³,
С. В. Качалин⁴, С. А. Полковникова⁵

^{1,2,3,5} Пензенский государственный университет, Пенза, Россия

² Пензенский научно-исследовательский электротехнический институт, Пенза, Россия

⁴ НПП «Рубин», Пенза, Россия

¹ vvi@pnzgu.ru, ² ivan@pniei.penza.ru, ³ tsib@pnzgu.ru, ⁴ s.kachalin@gmail.com, ⁵ 1996svetlanaserikova@gmail.com

Аннотация. *Актуальность и цели.* Рассматривается статистический анализ малых выборок в 16 опытов через оценку стандартного отклонения. Целью работы является нейросетевое предсказание ошибок вычисления стандартных отклонений на малых выборках биометрических данных. *Материалы и методы.* Для предсказания значений ошибок вычисления стандартных отклонений использованы многослойные сети искусственных нейронов. Алгоритмы обучения глубоких нейронных сетей хорошо известны. Главной проблемой для их реализации обычно является получение достаточно больших обучающих выборок. Новизна подхода заключается в том, что для рассматриваемой задачи использован автомат формирования обучающих выборок с разным значением ошибок оценки стандартного отклонения. *Результаты.* Созданный нейросетевой корректор ошибок позволяет снизить интервал ошибок вычисления стандартного отклонения на 22,7 % для выборок объемом в 16 опытов. При этом выявлена проблема, состоящая в необходимости выполнять длительное обучение многослойных нейронных сетей для каждого нового объема выборок. *Вывод.* Анализ полученных результатов в ходе проведенного исследования показал, что нейросетевые корректоры ошибок могут повысить достоверность статистических оценок и иных моментов. При этом могут быть созданы нейросетевые предсказатели ошибок вычисления математических ожиданий и коэффициентов корреляции. Предположительно, процесс улучшения достоверности будет монотонен и через один или два года удастся снизить интервал неопределенности вычислений дополнительно на 20 % за счет использования сетей из 15 или 20 слоев нейронов.

Ключевые слова: анализ малых выборок, многослойная сеть искусственных нейронов, предсказание ошибок вычисления стандартного отклонения

Для цитирования: Волчихин В. И., Иванов А. И., Малыгина Е. А., Качалин С. В., Полковникова С. А. Тестирование качества нейросетевой коррекции ошибок вычисления стандартного отклонения малых выборок биометрических данных // Измерения. Мониторинг. Управление. Контроль. 2021. № 4. С. 67–72. doi:10.21685/2307-5538-2021-4-8

TESTING THE QUALITY OF NEURAL NETWORK ERROR CORRECTION FOR CALCULATING THE STANDARD DEVIATION OF SMALL SAMPLES OF BIOMETRIC DATA

V.I. Volchikhin¹, A.I. Ivanov², E.A. Malygina³, S.V. Kachalin⁴, S.A. Polkovnikova⁵

^{1,2,3,5} Penza State University, Penza, Russia

² Penza Research Electrotechnical Institute, Penza, Russia

⁴ NPP "Rubin", Penza, Russia

¹ vvi@pnzgu.ru, ² ivan@pniei.penza.ru, ³ tsib@pnzgu.ru, ⁴ s.kachalin@gmail.com, ⁵ 1996svetlanaserikova@gmail.com

Abstract. *Background.* The statistical analysis of small samples in 16 experiments using the standard deviation estimation is considered. The aim of the work is the neural network prediction of errors in calculating standard deviations on small samples of biometric data. *Materials and methods.* Multi-layer networks of artificial neurons were used to predict the values of errors in calculating standard deviations. Deep neural network learning algorithms are well known.

The main problem for their implementation is usually to obtain sufficiently large training samples. The novelty of the approach lies in the fact that for the problem under consideration, an automatic machine for forming training samples with different values of errors in estimating the standard deviation is used. *Results.* The created neural network error corrector reduces the error interval for calculating the standard deviation by 22.7 % for samples of 16 experiments. At the same time, the problem is revealed, which consists in the need to perform long-term training of multilayer neural networks for each new volume of samples. *Conclusion.* The analysis of the results obtained in the course of the study showed that neural network error correctors can increase the reliability of statistical estimates and other points. At the same time, neural network predictors of errors in calculating mathematical expectations and correlation coefficients can be created. It is assumed that the process of improving confidence will be monotonous and in one or two years it will be possible to reduce the uncertainty interval of calculations by an additional 20 % by using networks of 15 or 20 layers of neurons.

Keywords: analysis of small samples, multilayer network of artificial neurons, prediction of errors in calculating the standard deviation

For citation: Volchikhin V.I., Ivanov A.I., Malygina E.A., Kachalin S.V., Polkovnikova S.A. Testing the quality of neural network error correction for calculating the standard deviation of small samples of biometric data. *Izmereniya. Monitoring. Upravlenie. Kontrol' = Measurements. Monitoring. Management. Control.* 2021;(4):67–72. (In Russ.). doi:10.21685/2307-5538-2021-4-8

Введение

Отечественная нейросетевая биометрия опирается на ряд национальных стандартов, в том числе на стандарт ГОСТ Р 52633.5-2011, регламентирующих процедуры автоматического обучения нейронной сети на 16 примерах образа «Свой»¹. Эти процедуры построены на использовании математических ожиданий и стандартных отклонений малых выборок обучения. К сожалению, при вычислении младших статистических моментов на малых выборках возникают значительные ошибки. При вычислении стандартного отклонения по классической формуле относительная ошибка $\Delta\sigma$ может достигать десятков процентов:

$$\sigma(16) = \sqrt{\frac{1}{n-1} \sum_{i=0}^{15} (E(x) - x_i)^2}, \quad (1),$$

где $E(x)$ – математическое ожидание малой выборки в 16 опытов.

Пример распределения значений σ для малых выборок в 16 примеров с эталонными данными с единичным стандартным отклонением приведено на рис. 1.

Из рис. 1 видно, что представленное на нем распределение близко к нормальному со стандартным отклонением $\sigma(\sigma, 16) = 0,177$.

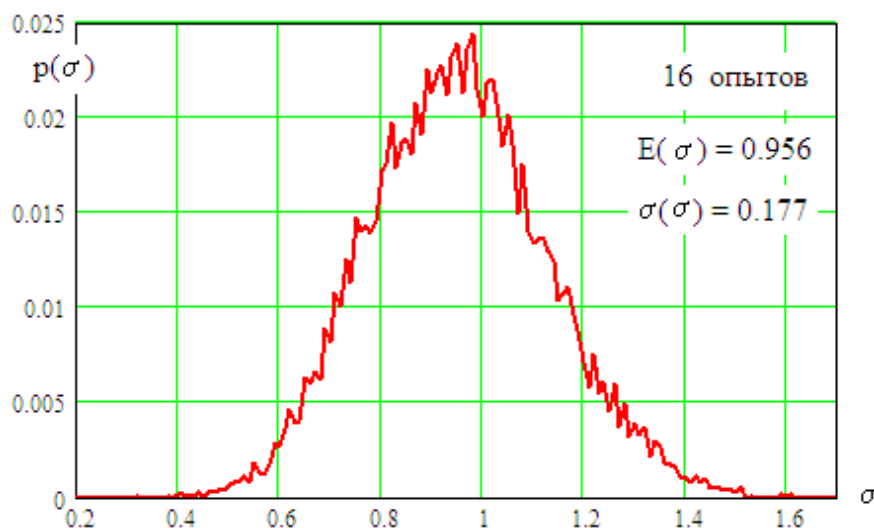


Рис. 1. Распределение значений ошибок при вычислении стандартного отклонения по классической формуле на малых выборках

¹ ГОСТ Р 52633.5-2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа.

Если обучающая выборка будет монотонно расти, то стандартное распределение значений, вычисленных по классической формуле (1), будет уменьшаться:

$$\sigma(\sigma, N) \approx \frac{0,177}{\sqrt{(N-16)/16}}, \tag{2}$$

где N – число примеров в выборке большей, чем 16 примеров.

Нейросетевое предсказание значений ошибок при вычислениях по классической формуле

В прошлом веке наш соотечественник А. И. Галушкин [1–3] предложил использовать глубокие сети искусственных нейронов и метод обратного распространения ошибки для их обучения. К сожалению, ЭВМ в 1974 г. не имели достаточных вычислительных ресурсов, и, соответственно, идеи А. И. Галушкина в СССР не могли быть реализованы по объективным причинам.

Ситуация изменилась только в 2014 г., усилиями ряда зарубежных исследователей была создана промышленная технология для обработки биометрических данных по распознаванию лиц и голосов пользователей [4, 5].

Очевидным является то, что многослойные нейронные сети универсальны, т.е. многослойные нейронные сети могут быть использованы для предсказания ошибок вычисления стандартных отклонений на малых выборках. Блок-схема такого преобразования приведена на рис. 2.

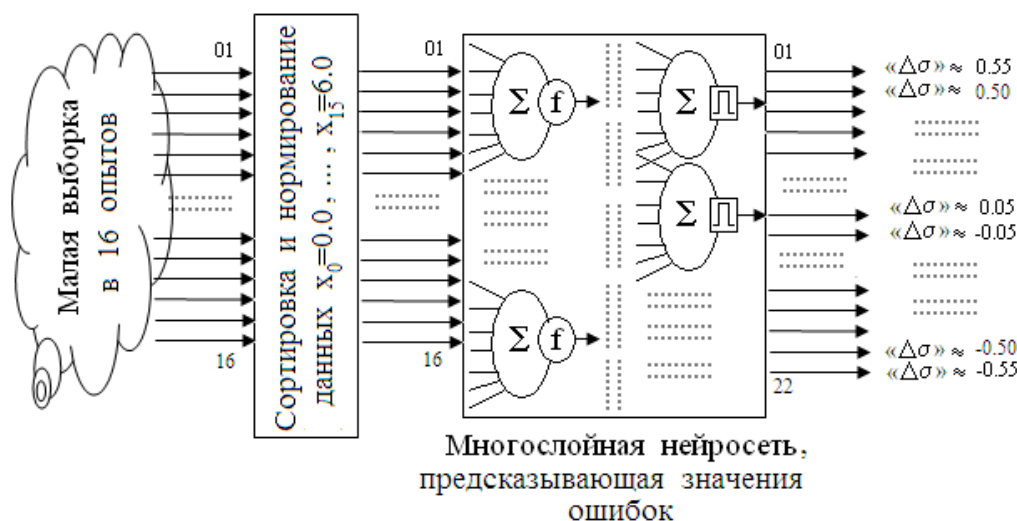


Рис. 2. Блок-схема предсказания 22 значений дискретных состояний ошибок – «Δσ» обученной многослойной сетью искусственных нейронов

Так как биометрические данные изменяются в достаточно широком динамическом диапазоне, перед их подключением к входам нейросети необходимо выполнить их нормировку:

$$\begin{cases} x \leftarrow \text{sort}(x), \\ \text{for } i \in 0 \dots 15, \\ x_i \leftarrow x_i - x_0, \\ x_i \leftarrow \frac{6x_i}{x_{15}}. \end{cases} \tag{3}$$

Нормировка сводится к сортировке данных по их возрастанию и приведению минимального значения параметров к нулевому состоянию $x_0 = 0,0$. Наибольшее значение в выборке после выполнения преобразований (3) составляет $x_{15} = 6,0$.

Важным является также то, что обучение многослойных нейронных сетей требует больших объемов примеров в обучающих выборках. В нашем случае задача является простой. Данные могут быть получены от программного генератора псевдослучайных чисел с нор-

мальным распределением с произвольным стандартным распределением. Проблема предварительной ручной разметки биометрических данных, существующая при обучении многослойных нейронных сетей распознаванию лиц пользователей и голосов пользователей, не существует.

Корректировка данных классической формулы вычисления стандартного отклонения и тестирование результатов

Обучение многослойных нейросетевых предсказателей значений ошибок может занимать достаточно много времени. Однако после обучения с достаточно высокой доверительной вероятностью на уровне 0,71 мы обладаем информацией о том или ином дискретном значении ошибок вычисления – « $\Delta\sigma$ », т.е. мы можем выполнить аддитивную корректировку данных, полученных по классической формуле

$$\tilde{\sigma}(16) = \sigma(16) - \langle \Delta\sigma(16) \rangle. \quad (4)$$

Чем выше доверительная вероятность, с которой нейросетевой предсказатель указывает одно из 22 дискретных значений ошибки « $\Delta\sigma(16)$ », тем точнее выполняется аддитивная корректировка (4). В будущем, когда возможно будет угадывать дискретное значение ошибки вычисления с высокой доверительной вероятностью 0,99, интервал неопределенности может снизиться до 22 раз:

$$\frac{\sigma(16)}{\tilde{\sigma}(16)} \approx 22. \quad (5)$$

Дальнейшему сужению интервала ошибок станет мешать принятая структура нейросетевого преобразователя (рис. 2). Для продолжения повышения точности оценок придется увеличивать число нейронов в последнем слое нейросетевого предсказателя.

Результаты тестирования рассматриваемого аддитивного нейросетевого корректора приведены на рис. 3.

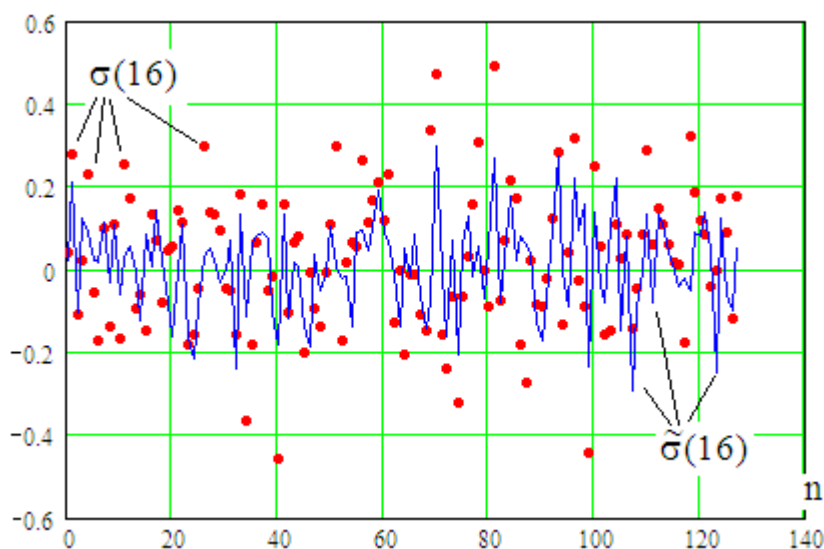


Рис. 3. Результаты тестирования работы нейросетевого корректора $\tilde{\sigma}(16)$ ошибок вычисления по классической формуле $\sigma(16)$ на 128 опытах

Проведение тестирования показало, что нейросетевой корректор позволяет снизить исходный интервал неопределенности (см. рис. 1) на 22,7 %. Стандартное отклонение данных, вычисленных по классической формуле (1) $\sigma(\sigma) \approx 0,177$, после корректировки снижается до величины 0,145. Это эквивалентно увеличению обучающей выборки с 16 опытов до существенно большей эквивалентной выборки в 23 опыта (оценка выполнена по формуле (2)). Получается, что привлечение дополнительной нейросетевой обработки биометрических данных

позволяет фактически на 7 примеров увеличить объем эквивалентной обучающей выборки с объема в 16 примеров до 23 примеров.

Заключение

Рассматриваемая в статье нейросеть имеет 16 входов, 22 выхода и 5 слоев нейронов. Предположительно, увеличение числа слоев искусственных нейронов в ближайшем будущем позволит и дальше снижать интервал неопределенности вычислений.

На данный момент границы повышения достоверности принимаемых решений за счет усложнения нейросетевой предобработки биометрических данных не определены. Предположительно, процесс улучшения достоверности будет монотонен, и через один или два года удастся снизить интервал неопределенности вычислений дополнительно на 20 % за счет использования сетей из 15 или 20 слоев нейронов.

Очевидным также является то, что нейросетевые корректоры ошибок могут повышать достоверность статистических оценок и иных моментов, в том числе могут быть созданы нейросетевые предсказатели ошибок вычисления математических ожиданий и коэффициентов корреляции.

Список литературы

1. Галушкин А. И. Синтез многослойных систем распознавания образов. М. : Энергия, 1974.
2. Хайкин С. Нейронные сети. Полный курс. 2-е изд., испр. М. : Вильямс, 2006. 1104 с.
3. Горбаченко В. И. Нейроинформатика : конспект лекций. Пенза : Пенз. гос. пед. ун-т им. В. Г. Белинского, 2011. 81 с.
4. Гудфеллоу Я., Бенджио И., Курвилл А. Глубокое обучение. М. : ДМК Пресс, 2017. 652 с.
5. Николенко С., Кадурин А., Архангельская Е. Глубокое обучение. СПб. : Питер, 2018. 480 с.

References

1. Galushkin A.I. *Sintez mnogoslonykh sistem raspoznavaniya obrazov = Synthesis of multilayer pattern recognition systems*. Moscow: Energiya, 1974. (In Russ.)
2. Khaykin S. *Neyronnye seti. Polnyy kurs. 2-e izd., ispr. = Neural networks. Full course. 2nd ed., corr.* Moscow: Vil'yams, 2006:1104. (In Russ.)
3. Gorbachenko V.I. *Neyroinformatika: konspekt lektiy = Neuroinformatics : lecture notes*. Penza: Penz. gos. ped. un-t im. V. G. Belin-skogo, 2011:81. (In Russ.)
4. Gudfellou Ya., Bendzhio I., Kurvil' A. *Glubokoe obuchenie = Deep learning*. Moscow: DMK Press, 2017:652. (In Russ.)
5. Nikolenko S., Kadurin A., Arkhangel'skaya E. *Glubokoe obuchenie = Deep learning*. Saint Petersburg: Pi-ter, 2018:480. (In Russ.)

Информация об авторах / Information about the authors

Владимир Иванович Волчихин

доктор технических наук, профессор,
президент
Пензенского государственного университета
(Россия, г. Пенза, ул. Красная, 40)
E-mail: vvi@pnzgu.ru

Vladimir I. Volchikhin

Doctor of technical sciences, professor,
president of the Penza State University
(40 Krasnaya street, Penza, Russia)

Александр Иванович Иванов

доктор технических наук, доцент,
профессор кафедры информационной
безопасности и технологий,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40);
научный консультант,
Пензенский научно-исследовательский
электротехнический институт
(Россия, г. Пенза, ул. Советская, 9)
E-mail: ivan@pniei.penza.ru

Alexander I. Ivanov

Doctor of technical sciences, associate professor,
professor of sub-department of technical means
of information security,
Penza State University
(40 Krasnaya street, Penza, Russia);
scientific consultant,
Penza Research Electrotechnical Institute
(9 Sovetskaya street, Penza, Russia)

Елена Александровна Малыгина

кандидат технических наук, докторант
кафедры технических средств
информационной безопасности,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: tsib@pnzgu.ru

Elena A. Malygina

Candidate of technical sciences, doctoral student
of the sub-department of technical means
of information security,
Penza State University
(40 Krasnaya street, Penza, Russia)

Сергей Викторович Качалин

кандидат технических наук,
заместитель начальника отделения,
НПП «Рубин»
(Россия, г. Пенза, ул. Байдукова, 2)
E-mail: s.kachalin@gmail.com

Sergey V. Kachalin

Candidate of technical sciences,
deputy head of branch,
SPE "Rubin"
(2 Baydukova street, Penza, Russia)

Светлана Андреевна Полковникова

аспирант,
Пензенский государственный университет
(Россия, г. Пенза, ул. Красная, 40)
E-mail: 1996svetlanaserikova@gmail.com

Svetlana A. Polkovnikova

Postgraduate student,
Penza State University
(40 Krasnaya street, Penza, Russia)

**Авторы заявляют об отсутствии конфликта интересов /
The authors declare no conflicts of interests.**

Поступила в редакцию/Received 17.06.2021

Поступила после рецензирования/Revised 24.06.2021

Принята к публикации/Accepted 29.09.2021