УДК 004.056.53

ПРОГНОЗИРОВАНИЕ ВЫХОДНЫХ ПАРАМЕТРОВ НЕЙРОСЕТЕВОГО ПРЕОБРАЗОВАТЕЛЯ «БИОМЕТРИЯ – КОД ДОСТУПА» НА ОСНОВЕ ЭЛЕКТРОЭНЦЕФАЛОГРАММЫ

С.М. Гончаров, А.Е. Боршевников

В статье рассматриваются выходные параметры модели нейросетевого преобразователя «Биометрия – код доступа» на основе электроэнцефалограммы. Приводится прогноз выходных значений вероятностей ошибок первого и второго рода модели нейросетевого преобразователя «Биометрия – код доступа» на основе электроэнцефалограммы.

Ключевые слова: нейросетевой преобразователь «Биометрия – код доступа», восстановление ключа, электроэнцефалограмма, P300, биометрическая аутентификация

Особое внимание в информационном обществе уделяют вопросам безопасности. В частности, одной из областей, в которой возникает необходимость проводить активные исследования, является область систем высоконадежной биометрической аутентификации. Под средствами высоконадежной биометрической аутентификацией понимается биометрическая аутентификация с приемлемой вероятностью ошибок первого рода и гарантированно малой вероятностью ошибок второго рода, сопоставимой по своему значению с вероятностью случайного подбора кода неизвестного криптографического ключа при малом числе попыток подбора [1].

В данной области одной из наиболее перспективных биометрических характеристик, обладающей высоким уровнем конфиденциальности, является электроэнцефалограмма (ЭЭГ). Несмотря на то, что данный вид биометрической характеристики является не совсем удобной для снятия данных пользователей, ЭЭГ целесообразно применять на различных объектах, требующих повышенных мер безопасности (например, критически важных объектов).

Реализация систем высоконадежной биометрической аутентификации заключается в разработке процедуры восстановления некоторой секретной информации (пароля, криптографического ключа). Отдельно стоит отметить исследования, в которых высоконадежные системы строились на основе больших и сверхбольших нейронных сетей с детерминированными алгоритмами обучения [2,3]. Подобные нейронные сети получили название нейросетевых преобразователей «Биометрия – код доступа».

В качестве выходных параметров систем биометрической аутентификации обычно используются вероятности ошибок первого и второго рода. Помимо указанных выходных параметров систем биометрической аутентификации в зарубежных исследованиях принято использовать дополнительные выходные параметры такие как точность аутентификации и уровень равенства ошибок (EER), т.е. значение вероятности при котором вероятность ошибки первого рода равна вероятности ошибки второго рода) [4,5].

Поскольку одним из выходных параметров систем аутентификации является точность, в зарубежных исследованиях принято использовать термин «высокоточная биометрическая аутентификация» [6]. Под системами высокоточной биометрической аутентификации понимают системы биометрической аутентификации, в которых точность аутентификации составляет 100%.

Точность аутентификации определяется как отношение правильно аутентифицированных пользователей к количеству всех опытов по аутентификации [7]:

Гончаров Сергей Михайлович – МГУ им. адм. Г.И. Невельского, канд. физ.-мат. наук, профессор, e-mail: sgprim143@gmail.com
Боршевников Алексей Егеньевич – ДВФУ, аспирант, e-mail: LAdG91@mail.ru

$$Acc = \frac{Tr}{All} \cdot 100\%, \quad (1)$$

где Tr – количество правильно аутентифицированных пользователей; All – общее количество попыток аутентификации.

Для проектирования биометрических систем важно производить прогнозирование выходных параметров для выбора оптимального количества входных параметров.

В рамках данной проведено прогнозирование выходных значений ошибки первого и второго рода, а также получены значение точности аутентификации и уровня равенства ошибок.

В основу проводимых исследований были положены данные и структура преобразователя из опытов, описанных в работах [2,3].

Для формирования потенциала P300 была использована зрительная стимуляция из поочередно меняющихся цифр от «0» до «9» с частотой 6 Гц. Данный выбор стимуляции обусловлен необходимостью удобного для пользователя распознавания стимуляции. Пользователь выбирал четыре символа, которые составляли мысленный пароль пользователя. Запись данных производилась в течение 20 секунд. Период записи делился на четыре этапа, каждый длительностью 5 секунд. В течение первого этапа пользователь мысленно концентрировался на первом символе мысленного пароля при его появлении на экране, в течение второго этапа пользователь концентрировался на втором символе мысленного пароля и так далее до окончания записи данных ЭЭГ.

В результате была сформирована база ЭЭГ, состоящая из 10 биометрических образов, для каждого из которых было отснято 20 примеров.

В качестве метода выделения параметров использовалось дискретное преобразование Фурье, применяемое к информативным участкам сигнала ЭЭГ, содержащим параметры потенциала P300.

При тестировании использовались следующие параметры преобразователя. Количество электродов – 14. Количество выбираемых коэффициентов Фурье для одного электрода – 70. Выбор такого количество параметров обусловлен ожидаемыми значениями вероятностей ошибок первого и второго рода. Количество нейронов первого слоя – 320. Размер восстанавливаемого ключа – 256 бит, что означает использование во втором слое нейронной сети 256 нейронов. Количество входов на нейрон было взято 4. Преобразователь обучался по стандарту ГОСТ Р 52633.5 [8].

Для проведения проверки была использована синтетическая база образов «Чужой», в которой содержится 10^6 биометрических образцов. Данная база была создана с помощью методов изложенных в стандарте ГОСТ Р 52633.2 [9].

Для исследования работы преобразователя проводилось обучение преобразователя с использованием дополненной синтетическими примерами базы образов «Чужой», а также без ее дополнения. Для эксперимента каждый из естественных образов фиксировались поочередно в качестве образа «Свой», остальные девять естественных образов формировали базу «Чужой». Для недополненной синтетическими образами базы «Чужой» среднее качество всех учитываемых преобразователем биометрических параметров $E(q(v)) = 0{,}14$. Для дополненной синтетическими примерами базы «Чужой» среднее качество параметров составило 0,18.

Был проведен опыт по возможности получения злоумышленником секретного ключа при условии знания весовых коэффициентов и мысленного пароля легитимного пользователя системы. Для каждого из пользователей было проведено 200 опытов. Суммарное количество проведенных опытов с использованием данных из базы «Чужой» составило 1800. В результате не один из злоумышленников не смог восстановить секретный ключ легитимного пользователя. Полученные средние результаты расстояний Хемминга приведены в таблице ниже.

Средние расстояния Хемминга до восстановленного ключа пользователя «Свой»

Номер пользователя	Среднее расстояние Хемминга (качество биометрических параметров $E(q(v)) = 0{,}14$)	Среднее расстояние Хемминга (качество биометрических параметров $E(q(v)) = 0{,}18$)
1	89	90
2	56	129
3	50	65
4	131	95
5	77	78
6	145	134
7	143	83
8	57	83
9	91	148
10	120	102

Также проводился опыт по восстановлению ключа пользователем «Свой». Количество проведенных опытов составило 200. Стоит отметить, что при проведении экспериментов по восстановлению ключа для образа «Свой», ключ безошибочно восстанавливался.

Для случаев, когда тестирующая выборка является небольшой, и ошибка первого рода не была выбрана, данную ошибку можно вычислить по формуле [10]:

$$P_1 \approx \int_1^{\infty} \frac{1}{2^{\frac{\Omega}{2}} \cdot \Gamma\left(\frac{\Omega}{2}\right)} \cdot x^{\frac{\Omega}{2}-1} \cdot e^{-\frac{x^2}{2}} \cdot dx, \qquad (2)$$

где Ω – количество степеней свободы в распределении χ^2.

В случае, когда в проведенной серии испытаний по предъявлению биометрической характеристики образа «Свой», состоящей из m опытов, не обнаружен факт отказа в доступе, число степеней свободы в распределении χ^2 вычисляется по формуле:

$$\Omega = \frac{1}{m+1}. \qquad (3)$$

Для $m = 20$ получим вероятность ошибки первого рода $P_1 \approx 7 \cdot 10^{-3}$ [3]. Для $m = 400$ вероятность ошибки первого рода $P_1 \approx 7 \cdot 10^{-4}$. На рис. 1 приведен график зависимости вероятности ошибки первого рода от количества испытаний, в которых ключ легитимного пользователя безошибочно восстанавливался. График был построен для 20, 100, 500, 1000, 2000, 3000 и 10000 испытаний. Стоит отдельно выделить, что для 10000 успешных опытов по восстановлению ключа $P_1 \approx 10^{-5}$. В рамках данной работы не проводилось исследований того, какие значения приобретает вероятность ошибки первого рода при неточном восстановлении легитимным пользователем собственного ключа, так как для получения значений необходимо знать величину расстояний Хемминга, полученных при неуспешных опытах по восстановлению ключа [10].

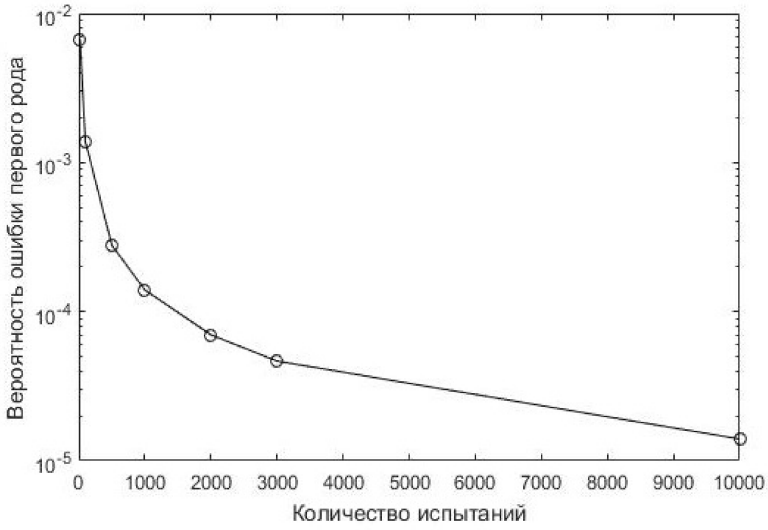

Рис. 1. График зависимости вероятности ошибки первого рода от количества успешных испытаний

Опишем схему приблизительной оценки вероятности ошибки второго рода. Вероятность ошибки второго рода P_2 можно вычислить приближенно, исходя из гипотезы нормального закона распределения значений вероятности ошибок, по формуле [11]:

$$P_2 \approx \frac{1}{\sqrt{2\pi}} \int_{\sqrt{n}E(q(v))}^{\infty} e^{-\frac{x^2}{2}} dx, \qquad (4)$$

где n – число учитываемых преобразователем биометрических параметров; $E(q(v))$ – среднее качество всех учитываемых преобразователем биометрических параметров.

Для описанного эксперимента количество параметров $n = 980$. При качестве биометрических параметров $E(q(v)) = 0{,}14$ $P_2 \approx 3*10^{-5}$. При качестве биометрических параметров $E(q(v)) = 0{,}18$ $P_2 \approx 10^{-9}$. На рис. 2 и рис. 3 показаны графики зависимости вероятности ошибки второго рода от количества используемых входных параметров при $E(q(v)) = 0{,}14$ и $E(q(v)) = 0{,}18$ соответственно.

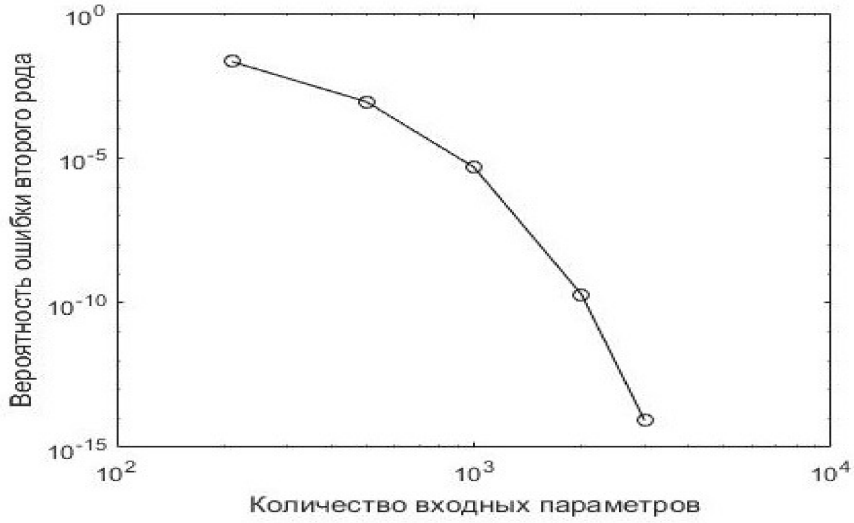

Рис. 2. График зависимости вероятности ошибки второго рода от количества используемых входных параметров ($E(q(v)) = 0{,}14$)

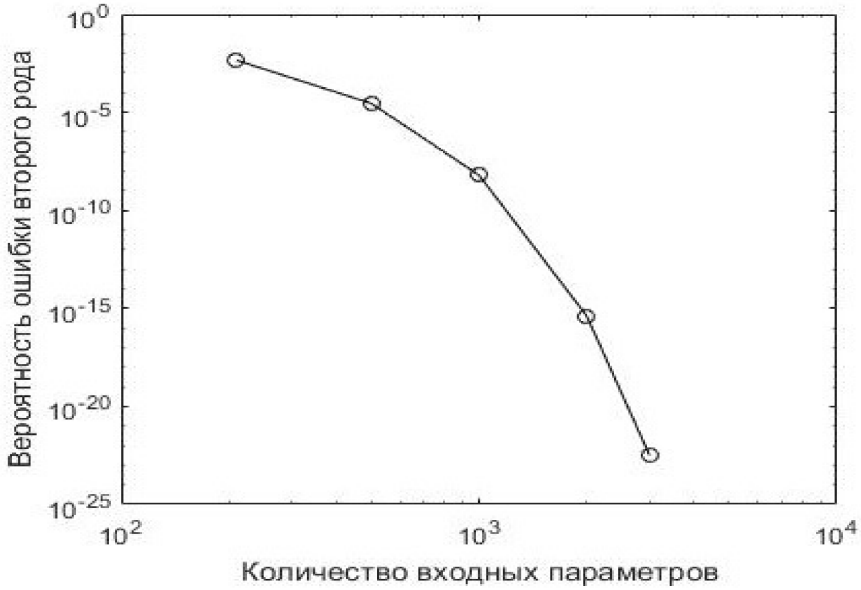

Рис. 3. График зависимости вероятности ошибки второго рода от количества используемых входных параметров ($E(q(v)) = 0{,}18$)

В соответствии с формулой (1) при $Tr = 2000$ и $All = 2000$ точность аутентификации $Acc = 100\%$.

Определим уровень равенства ошибок. Вероятность ошибки первого рода имеет постоянный характер из-за того, что количество примеров образа пользователя «Свой» не изменяется со временем (за исключением возможности добавления новых образов), а база образов «Чужой» может свободно увеличиваться за счет добавления синтетических образов, то вероятность ошибки первого рода носит постоянный характер, а вероятность ошибки второго рода изменяется в зависимости от изменения качества биометрических параметров. Таким образом, уровень равенства ошибок $EER \approx 7 \cdot 10^{-3}$ при использовании недополненной синтетическими примерами базы «Чужой» и $EER \approx 7 \cdot 10^{-4}$ при использовании дополненной синтетическими примерами базы «Чужой».

Полученные результаты показывают, что исследуемая модель нейросетевого преобразователя "Биометрия - код доступа" показывает допустимые результаты для выходных параметров и даже показывает лучшие результаты по сравнению с существующими разработками в данной области. Также спрогнозированные значения вероятностей ошибок первого и второго рода позволяют определить необходимое количество входных параметров, используемых в нейросетевом преобразователе, для достижения определенного уровня вероятности ошибки второго рода.

Для проведения дальнейших исследований необходимо расширить базу естественных образов для проведения не теоретического расчета показателей вероятностей ошибок первого и второго рода, а получения значений данных вероятностей для экспериментов, проведенных на сравнимой с величиной ошибок выборке.

Литература

1. ГОСТ Р 52633.0-2006. Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации. – Введен впервые; Введ. 27.12.2006. – М.: Стандартинформ, 2007. – 25 с.

2. Гончаров С.М. Нейросетевой преобразователь «Биометрия – код доступа» на основе электроэнцефалограммы в современных криптографических приложениях [Текст] / С.М. Гончаров,

А.Е. Боршевников // Вестник. – 2016. – № 1. – С. 17-22.

3. Гончаров С.М. Использование вейвлет-преобразования для выделения биометрических характеристик потенциала Р300 в задачах высоконадежной биометрической аутентификации [Текст] / С.М. Гончаров, А.Е. Боршевников // Информация и безопасность. – 2016. – Т. 19, № 4. – С. 527-530.

4. Abo-Zahhad M. A new multi-level approach to EEG based human authentication using eye blinking [Text] / M. Abo-Zahhad, S.M. Ahmed, S.N. Abbas // In Pattern Recognition Letters. – 2016. – Vol. 82, No. 2, – P. 216-225.

5. Abo-Zahhad M. A New EEG Acquisition Protocol for Biometric Identification Using Eye Blinking Signals [Text] / M. Abo-Zahhad, S.M. Ahmed, S.N. Abbas // International Journal of Intelligent Systems and Applications (IJISA). – 2015. – P. 48-54.

6. Koike-Akino T. High-accuracy user identification using EEG biometrics [Text] / T. Koike-Akino // 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). – 2016. – P. 854-858.

7. Palaniappan R. Method of Identifying Individuals using VEP Signals and Neural Network [Text] / R. Palaniappan // IEEE Proceedings: Science, Measurement and Technology. – 2004. – Vol. 151, No. 1. – P. 16-20.

8. ГОСТ Р 52633.5-2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия - код доступа. – Введен впервые; Введ. 01.12.2011. – М.: Стандартинформ, 2012. – 20 с.

9. ГОСТ Р 52633.2-2010. Защита информации. Техника защиты информации. Требования к формированию синтетических биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации. – Введен впервые; Введ. 30.09.2010. – М.: Стандартинформ, 2011. – 17 с.

10. Ахметов Б.С. Оценка вероятностей появления ошибок нейросетевых преобразователей биометрия-код на основе малых выборок [Текст] / Б.С. Ахметов, А.И. Иванов и др. // Труды II Международной научной конференции «Высокие технологии – залог устойчивого развития». – 2013. – Т. 1. – С. 234-237.

11. ГОСТ Р 52633.1-2009. Защита информации. Техника защиты информации. Требования к формированию баз естественных биометрических образов, предназначенных для тестирования средств высоконадежной биометрической аутентификации. – Введен впервые; Введ. 15.12.2009. – М.: Стандартинформ, 2010. – 24 с.

ФГБОУ ВО «Морской государственный университет имени адмирала Г.И. Невельского»
Maritime state university named after admiral G.I. Nevelskoy
ФГАОУ ВО «Дальневосточный федеральный университет»
Far eastern federal university

THE FORECAST FOR OUTPUT PARAMETERS OF A NEURAL NET TRANSFORMER «BIOMETRICS – ACCESS CODE» BASED ON ELECTRONETSEPHALOGRAM

S.M. Goncharov, A.E. Borshevnikov

The article discusses the output parameters of the mathematical model of a neural net transformer «Biometrics - access code» based on the electroencephalogram. The forecast of the output parameters of the mathematical model of a neural net transformer «Biometrics – Access code» based on the electroencephalogram is describe.

Keywords: neural net transformer «Biometrics – access code», key recovery, electroencephalogram, P300, biometric authentication