

БЫСТРАЯ ОЦЕНКА КАЧЕСТВА ОБУЧЕНИЯ СЕТИ ИСКУССТВЕННЫХ НЕЙРОНОВ С ИСПОЛЬЗОВАНИЕМ КВАНТОВОГО ОРАКУЛА, ПРЕДСКАЗЫВАЮЩЕГО ВЕРОЯТНОСТИ ПОЯВЛЕНИЯ РЕДКИХ СОБЫТИЙ АЛГОРИТМОМ ГОСТ Р 52633.3

В. И. Волчихин, А. В. Безяев, Н. А. Иванова, Ю. И. Серикова

Введение

В настоящее время в России и за ее пределами активно создаются технологии преобразования биометрического образа человека в его криптографический ключ. За рубежом для этого используют технологии так называемых «нечетких экстракторов» [1–4], Россия и Казахстан идут путем создания нейросетевых преобразователей биометрия-код [5, 6]. Следует отметить то, что «нечеткие экстракторы» являются частным случаем искусственных нейронных сетей и по этой причине все их качественные показатели оказываются хуже, чем у нейросетевых преобразователей [7].

Одной из проблем нейросетевой биометрии является тестирование качества ее работы. Базовый национальный стандарт [8] рекомендует выполнять тестирование нейросетевого преобразователя после каждого обучения. Если тестировать нейронную сеть традиционно, то каждое средство обучения должно иметь в своем составе тестовую базу, состоящую примерно из 1 000 000 образов «Чужой». Это потребует как минимум снабжать средство автоматического обучения нейронной сети памятью для хранения тестовой базы образов «Чужой». Так, если речь идет о голосовой биометрической аутентификации, звуковой файл парольной фразы из трех слов при 8-битной кодировке с частотой дискретизации 8 кГц будет иметь размер порядка 128 Кбайт, т.е. память для хранения тестовой базы образов «Чужой» должна иметь размеры порядка 128 Гбайт. Это сильно удорожает устройство и увеличивает его габариты.

Причиной всего этого является ориентация при тестировании на традиционную технологию, построенную на ожидании и фиксации редких событий. Чем ниже вероятность редкого события, тем больше должна быть тестовая выборка для его обнаружения. Выходом из этого тупика является использование оракула, который не ждет появления редких событий, а предсказывает вероятность их появления по некоторым признакам, анализируемым им данных.

Прогнозирование вероятности появления редких событий в пространстве расстояний Хэмминга

В случае, если используется нейросетевой преобразователь биометрия-код, обученный алгоритмом ГОСТ Р 52633.5 [9], то стандарт по тестированию ГОСТ Р 52633.3 [10] предписывает проводить численный эксперимент по схеме, изображенной на рис. 1.

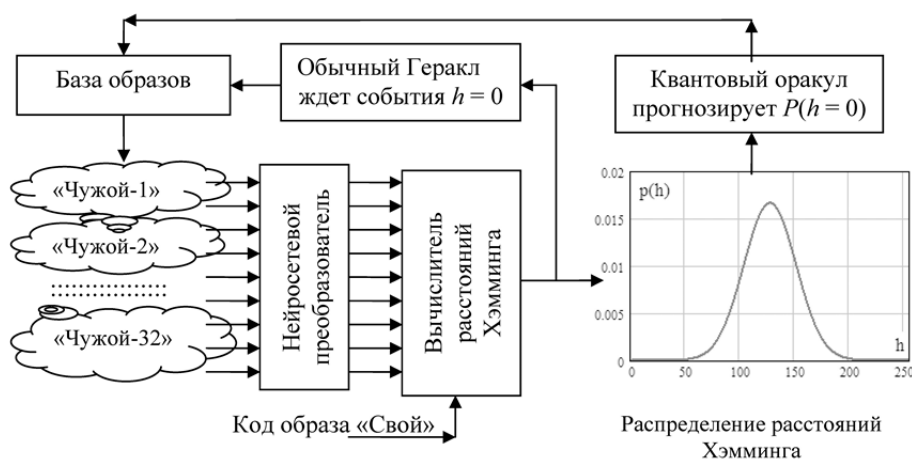


Рис. 1. Схема организации численного эксперимента по оценке вероятности ошибок второго рода (ошибочный пропуск «Чужого»)

Достаточно располагать выборкой всего из 32 разных (случайно выбранных из тестовой базы) образов «Чужой». В этом случае мы можем для каждого кода образа «Чужой» вычислить расстояние Хэмминга по отношению к коду образа «Свой»:

$$h("c", "x") = \sum_{i=1}^{256} ("c_i") \oplus ("x_i"), \quad (1)$$

где 256 – длина криптографического ключа отечественных криптографических стандартов на шифрование и формирование цифровой подписи; "c" – бинарный код образа «Свой»; "x" – бинарный код образа «Чужой».

Выборки из 32 примеров достаточно, чтобы получить оценку математического ожидания – $E(h)$ и стандартного отклонения – $\sigma(h)$ расстояний Хэмминга. В свою очередь, знание этих двух статистических моментов позволяет спрогнозировать вероятность появления ошибок второго рода, когда выходной код случайно выбранного образа «Чужой» совпадет с кодом «Свой» или $h = 0$:

$$P_2 = \frac{1}{\sigma(h)\sqrt{2\pi}} \int_{-\infty}^1 \exp\left\{-\frac{(E(h)-u)^2}{2\sigma^2(h)}\right\} \cdot du. \quad (2)$$

Вместо того, чтобы ждать появления редкого события $h = 0$, расходуя на численный эксперимент время, и вместо расходования памяти на хранение большой тестовой базы, мы создаем оракула (2), который предсказывает появление редкого события в рамках гипотезы о нормальности распределения значений расстояний Хэмминга. Априорная информация о нормальном законе заранее получена и зафиксирована в руководящем документе [10]. Гипотеза нормальности работает только для искусственных нейронных сетей, обученных алгоритмом ГОСТ Р 52633.5 [9], для других алгоритмов обучения, а также для «нечетких экстракторов» [1–4] гипотеза нормальности не работает.

Оценка выигрыша от перехода к использованию оракула, предсказывающего вероятность появления редких событий

Обращение к оракулам – это одна из древнейших традиций. Она была у всех народов, однако наиболее яркие письменные источники, дошедшие до нас, оказались греческими. В связи с этим уместно при исследованиях пользоваться греческими источниками, названиями, именами.

Очевидно, что Геракл сильнее любого оракула, однако он не видит будущего и вынужден обращаться к оракулу за советом. При этом Геракл доверяет оракулу, опираясь на предшествующий опыт своих соседей. В Древней Греции не было рейтинговых агентств (сейчас много рейтинговых агентств, таких как Standart & Poor's, Moody's, Fitch Ratings, ...), в связи с этим Геракл не мог обратиться в одно из рейтинговых агентств и заказать оценку уровня доверия тому или иному оракулу. Геракл должен был сам оценивать уровень доверия к тому или иному оракулу, практикующему на текущий момент.

В нашем случае Геракл должен собрать тестовую базу из 1 000 000 случайных образов «Чужой». Далее Геракл должен запустить численный эксперимент по схеме, отображенной на рис. 1, по силовой атаке «тупого» перебора образов тестовой базы образов «Чужой». Если на N -м шаге силовой атаки Геракла в базе обнаружится коллизия образов «Свой»/«Чужой», то вероятность ошибок второго рода Геракл оценит по классической формуле

$$P_2 \approx \frac{1}{N}. \quad (3)$$

Очевидно, что оценка (3) является случайной, оцениваемому оракулу могло просто не повезти со случайной точкой старта силового перебора адресов базы образов «Чужой». Если Геракл заинтересован в «правде», то он должен повторить численный эксперимент много раз, каждый раз забывая результаты предыдущего теста и стартуя со случайно выбранного адреса. Перебор адресов должен осуществляться каждый раз в новой системе адресации примеров образов «Чужой». Тогда Геракл имеет возможность достаточно надежно оценить ускорение вычислений, достижимое за счет обращения к оракулу, по следующей формуле:

$$W \approx \frac{E(N)}{32}. \quad (4)$$

Чем выше показатель ускорения (4), тем больше будет очередь из желающих получить «предсказание» того или иного оракула. При этом затраты на «рейтингование» оракулов могут быть весьма и весьма значительными, Геракл, наткнувшись на лжеоракула, может пожелать сокрытия факта своих затрат (своих ошибочных действий) и будет и далее обращаться к лжеоракулу.

Если тестовая база в 1 000 000 образов «Чужой» оказалась мала, то Геракл не сможет обнаружить коллизию образов «Свой»/«Чужой». В этом случае он должен запоминать значения расстояний Хэмминга всех образов «Чужой» и найти минимальное из них – $\min(h)$. При этом оценка способностей оракула должна вычисляться по формуле, учитывающей конечный размер базы тестовых образов:

$$W \approx \frac{E(N)}{32} \sum_{i=1}^{E(\min(h))} (256 + (1-i))^i. \quad (5)$$

Практика применения общедоступной среды моделирования «БиоНейроАвтограф» [11] позволяет показать, что выигрыш от привлечения оракула, выполненного по ГОСТ Р 52633.3 [10], составляет от 100 000 для «слабых, нестабильных» рукописных образов с низкой «уникальностью», до 1 000 000 000 000 для «уникальных и стабильных» рукописных биометрических образов. Чем надежнее биометрико-нейросетевая защита, тем больше получается выигрыш от обращения к тому или иному оракулу. При этом Геракл всегда может положиться на общественное мнение рейтинговых агентств, контролирующих «рынок» оракулов или самостоятельно оценить качество данных лично ему предсказаний.

Соотношение между квантовой механикой и квантовой нейродинамикой

Столь впечатляющие оценки возможностей оракула, выполненного по ГОСТ Р 52633.3 [10], – это объективная реальность. Причина столь высокой эффективности оракула в том, что он обладает очень большим объемом априорной информации о том, что объект тестирования – это искусственная нейронная сеть, обученная по ГОСТ Р 52633.5 [9], так как только в этом случае распределение расстояний Хэмминга будет нормальным. Если оракулу, выполненному по ГОСТ Р 52633.3 [10], Геракл намеренно даст данные «нечеткого экстрактора», то качество предсказаний оракула катастрофически упадет. При тестировании оракулов рейтинговые агентства и Геракл должны пользоваться достоверными данными. При тестировании оракула Геракл не должен менять объявленные условия. Отечественному оракулу, ориентированному на предсказания по алгоритму ГОСТ Р 52633.3 [10], не следует давать данные «нечетких экстракторов». Для предсказания оценки качества работы «нечетких экстракторов» нужно создавать специальных оракулов.

Принципиально важным является также то, что рассматриваемый в данной статье оракул ориентирован только на тестирование нейронных сетей, находящихся в режиме поддержки выходной квантовой суперпозиции [12]. Либо оракул, либо Геракл (схема рис. 1) должен постоянно изменять состояния данных на входах нейронной сети. Только в этом случае возникает нейродинамика, и мы можем наблюдать квантовую суперпозицию выходных состояний нейронной сети.

Нейронные сети кардинально отличаются от объектов квантовой механики (например, молекулы водорода) [13]. «Нейросетевая молекула», находясь в статике, обеспечивает «легкий» режим наблюдения своего выходного кода, нет нужды прилагать специальные усилия для того, чтобы увидеть выходной код «нейросетевой молекулы» или номер одного из ее возможных состояний. «Нейросетевые молекулы» – это программные макрообъекты [12], для них не работает принцип неопределенности Гейзенберга. «Коты Шредингера» не мешают использованию «нейросетевых молекул», так как искусственно поддерживаемая нейродинамика «нейросетевых молекул» не описывается уравнением Шредингера. Спектры выходных состояний «нейросетевых молекул» и «молекул хи-квадрат Пирсона» вполне наблюдаемы, однако эти объекты кто-то должен «загнать» в циклическое состояние [14–17] и держать их в циклическом состоянии необходимое для осуществления прогноза время.

Микрообъекты квантовой механики ведут себя совершенно иначе. Синхронизовать между собой двух и более «котов Шредингера» крайне сложно. Для квантовой микромеханики физически существующих микрообъектов принцип неопределенности Гейзенберга является гигантским техническим ограничением при попытках организации квантовых вычислений и квантовых предсказаний [16, 17].

Заключение

Вопрос о том, можно ли, опираясь на постулаты квантовой механики, создать квантового оракула, остается открытым. В данной статье авторы постарались показать, что создать квантового оракула, опираясь на нейросетевую динамику, вполне возможно. Более того, стандарт по тестированию нейросетевых преобразователей биометрия-код ГОСТ Р 52633.3 [10] фактически узаконил применение квантовых оракулов шесть лет назад в 2011 г.

Формулы (4) и (5), приведенные в данной статье, инвариантны к принципам, на которых построен квантовый оракул. Они могут быть применены для оценки качества любых предсказаний. Очевидно также то, что формулы (4), (5) могут быть уточнены с учетом реализации того или иного оракула, ориентированного под решение той или иной практической задачи. Авторы статьи уверены в том, что тематика объективной оценки качества предсказаний квантовых оракулов актуальна и вызовет интерес у специалистов, занимающихся проектированием и тестированием сложных систем.

Библиографический список

1. Dodis, Y. Fuzzy Extractors: How to Generate Strong Keys from Biometrics and Other Noisy / Y. Dodis, L. Reyzin, A. Smith // Proc. EUROCRYPT. – 2004. – P. 523–540.
2. Monrose, F. Cryptographic key generation from voice / F. Monrose, M. Reiter, Q. Li, S. Wetzel // Proc. IEEE Symp. on Security and Privacy, 2001. – P. 202–213.
3. Ramírez-Ruiz, J. Cryptographic Keys Generation Using FingerCodes / J. Ramírez-Ruiz, C. Pfeiffer, J. Nolasco-Flores // Advances in Artificial Intelligence – IBERAMIA-SBIA (LNCS 4140). – 2006. – P. 178–187.
4. Hao, F. Crypto with Biometrics Effectively / F. Hao, R. Anderson, J. Daugman // IEEE TRANSACTIONS ON COMPUTERS. – 2006. – Vol. 55, № 9. – P. 1073–1074.
5. Технология использования больших нейронных сетей для преобразования нечетких биометрических данных в код ключа доступа : моногр. / Б. С. Ахметов, А. И. Иванов, В. А. Фунтиков, А. В. Безяев, Е. А. Малыгина. – Алматы. : Изд-во LEM, 2014. – 144 с. – URL: <http://portal.kazntu.kz/files/publicate/2014-06-27-11940.pdf>
6. Ахметов, Б. С. Алгоритмы тестирования биометрико-нейросетевых механизмов защиты информации / Б. С. Ахметов, В. И. Волчихин, А. И. Иванов, А. Ю. Малыгин. – Алматы : КазНТУ им. Сатпаева, 2013. – 152 с. – URL: <http://portal.kazntu.kz/files/publicate/2014-01-04-11940.pdf>.
7. Иванов, А. И. Нечеткие экстракторы: проблема использования в биометрии и криптографии / А. И. Иванов // Первая миля. – 2015. – № 1. – С. 40–47.
8. ГОСТ Р 52633.0-2006. Защита информации. Техника защиты информации. Требования к средствам высоконадежной биометрической аутентификации.
9. ГОСТ Р 52633.5-2011. Защита информации. Техника защиты информации. Автоматическое обучение нейросетевых преобразователей биометрия-код доступа.
10. ГОСТ Р 52633.3-2011. Защита информации. Техника защиты информации. Тестирование стойкости средств высоконадежной биометрической защиты к атакам подбора.
11. Иванов, А. И. Среда моделирования «БиоНейроАвтограф» / А. И. Иванов, О. С. Захаров. – Пенза : АО «ПНИЭИ», 2009. – URL: <http://пниэи.рф/activity/science/noc.htm>
12. Иванов, А. И. Многомерная нейросетевая обработка биометрических данных с программным воспроизведением эффектов квантовой суперпозиции / А. И. Иванов. – Пенза : АО «ПНИЭИ», 2016. – 133 с. – URL: <http://пниэи.рф/activity/science/BOOK16.pdf>
13. Нильсон, М. Квантовые вычисления и квантовая информация / М. Нильсон, И. Чанг. – М. : Мир, 2006. – 821 с.
14. Кулагин, В. П. Циклические континуально-квантовые вычисления: усиление мощности хи-квадрат критерия на малых выборках / В. П. Кулагин, А. И. Иванов, А. И. Газин, Б. Б. Ахметов // Аналитика. – 2016. – № 5 (30). – С. 22–29.
15. Перспективы создания циклической континуально-квантовой хи-квадрат машины для проверки статистических гипотез на малых выборках биометрических данных и данных иной природы / В. И. Волчихин, А. И. Иванов, Д. В. Пашенко, Б. Б. Ахметов, С. Е. Вятчанин // Известия высших учебных заведений. Поволжский регион. Технические науки. – 2017. – № 1. – С. 3–7.
16. Использование среднего геометрического, ожидаемой и наблюдаемой функций вероятности как статистического критерия оценки качества биометрических данных / Б. С. Ахметов, А. И. Иванов, К. А. Перфилов, Е. Д. Проценко, Д. С. Пашенко // Труды Международного симпозиума Надежность и качество. – 2015. – Т. 2. – С. 283–285.
17. Быстрый алгоритм оценки высокоразмерной энтропии биометрических образов на малых выборках / Б. Б. Ахметов, А. И. Иванов, А. Ю. Малыгин, А. В. Безяев, А. И. Газин // Труды Международного симпозиума Надежность и качество. – 2015. – Т. 2 – С. 285–287.

Волчихин Владимир Иванович

доктор технических наук, профессор,
президент Пензенского государственного
университета
(440026, Россия, г. Пенза, ул. Красная, 40)
E-mail: vvi@pnzgu.ru

Безяев Александр Викторович

кандидат технических наук,
ведущий специалист,
Пензенский филиал ФГУП НТЦ «Атлас»
(440026, Россия, г. Пенза, ул. Советская, 9)
E mail: Bezyaev_Alex@mail.ru

Иванова Надежда Александровна

аналитик,
ООО «БиоКрипт»
(440031, Россия, г. Пенза, ул. Окружная, 111)
E-mail: ivan@pnici.penza.ru

Серикова Юлия Игоревна

магистрант,
Пензенский государственный университет
(440026, Россия, г. Пенза, ул. Красная, 40)
E-почта: julia-ska@yandex.ru

Аннотация. *Актуальность и цели.* Целью работы является оценка выигрыша от использования квантового оракула при тестировании качества обучения сети искусственных нейронов. Актуальность работы обусловлена необходимостью тестирования нейронной сети после каждого ее обучения или дообучения. *Материалы и методы.* Алгоритм тестирования нейронной сети на малой выборке по ГОСТ Р 52633.3 рассматривается с позиций квантовой кибернетики и с позиций классической статистики. В пространстве расстояний Хэмминга квантовая суперпозиция выходных состояний нейронной сети хорошо описывается нормальным законом распределения значений. *Результаты.* Показано, что квантовый оракул, предсказывающий вероятность появления редких событий, случайного угадывания «Чужим» кода «Свой», дает ускорение тестирования пропорционально обратной величине вероятности ошибок второго рода. Чем надежнее средство биометрико-нейросетевой аутентификации, тем больше выигрыш от использования квантового оракула. *Выводы.* Отказ от классического тестирования, построенного на ожидании редких событий, позволяет снизить размеры тестовой базы с 1 000 000 образов «Чужой» до 32 образов, что эквивалентно снижению затрат на сбор и подготовку данных примерно на пять порядков.

Ключевые слова: статистический анализ малых выборок, качество обучения искусственных нейронных сетей, квантовая суперпозиция.

Volchikhin Vladimir Ivanovich

doctor of technical sciences, professor,
president of Penza State University
(440026, 40 Krasnaya street, Penza, Russia)

Bezyaev Aleksandr Viktorovich

candidate of technical sciences,
leading specialist of STC «Atlas» Penza branch
(440026, 9 Sovetskaya street, Penza, Russia)

Ivanova Nadezhda Aleksandrovna

analyst,
BioCrypt LLC
(440031, 111 Okruzhnaya street, Penza, Russia)

Serikova Julia Igorevna

master degree student,
Penza State University
(440026, 40 Krasnaya street, Penza, Russia)

Abstract. *Background.* The aim of the paper is to estimate the gain from using a quantum oracle when testing the quality of learning the network of artificial neurons. The urgency of the work is due to the need to test the neural network after each of its training or after-training. *Materials and methods.* The algorithm for testing a neural network for a small sample according to GOST R 52633.3 is considered from the positions of quantum cybernetics and from the standpoint of classical statistics. In the Hamming distance space, the quantum superposition of the output states of a neural network is well described by the normal law of distribution of values. *Results.* It is shown that a quantum oracle predicting the probability of occurrence of rare events, random guessing of «Alien» code «Svoy», gives the acceleration of testing in proportion to the reciprocal of the probability of errors of the second kind. The more reliable the means of biometric-neuronet authentication, the greater the gain from using a quantum oracle. *Conclusions.* The abandonment of classical testing, built on the expectation of rare events, allows reducing the size of the test database from 1,000,000 images of «Alien» to 32 images, which is equivalent to reducing the cost of collecting and preparing data by about five orders of magnitude.

Key words: statistical analysis of small samples, quality of training of artificial neural networks, quantum superposition.

УДК 519.24; 53; 57.017

Волчихин, В. И.

Быстрая оценка качества обучения сети искусственных нейронов с использованием квантового оракула, предсказывающего вероятности появления редких события алгоритмом ГОСТ Р 52633.3 / В. И. Волчихин, А. В. Безяев, Н. А. Иванова, Ю. И. Серикова // Надежность и качество сложных систем. – 2017. – № 1 (17). – С. 25–29. DOI 10.21685/2307-4205-2017-1-4.