

УДК 004.032.26

А. С. Перков, Т. Р. Жангиров, А. А. Лисс
Санкт-Петербургский государственный электротехнический
университет «ЛЭТИ» им. В. И. Ульянова (Ленина)

Н. Ю. Григорьева
СПб НИЦЭБ РАН

Л. В. Чистякова
РЦ «Культивирование микроорганизмов» НП СПбГУ

Сравнение методов обучения нейронных сетей в задаче классификации

На основе анализа серий искусственных нейронных сетей прямого распространения разработан метод определения оптимальной архитектуры нейронной сети для задачи классификации штаммов цианобактерий по спектрам собственной флуоресценции. Проведен анализ шести градиентных методов обучения нейронных сетей и их параметров, найдено оптимальное число нейронов на скрытом слое для нейронных сетей, обученных каждым из методов, проведена оценка различных методов инициализации весовых коэффициентов нейронов и способов разбиения исходной выборки на обучающую, тестовую и контрольную. Выбор оптимальной архитектуры проводился на основе результатов классификации, а именно, на основе графиков точности классификации и ошибки классификации. Исследование проводилось на примере распознавания 16 классов, представляющих собой 16 штаммов цианобактерий. Был выявлен ряд недостатков в методике тестирования нейронных сетей прямого распространения и определены направления дополнительных исследований классифицирующих нейронных сетей в части расширения методики тестирования их внутренней логики.

Нейронные сети, методы обучения, машинное обучение, задачи классификации, методы инициализации

Искусственные нейронные сети в настоящее время широко применяются практически во всех областях, связанных с обработкой больших объемов информации. Одной из самых распространенных задач является задача классификации. Одним из основных направлений применения нейронных сетей в данной области является классификация изображений [1], также нейронные сети активно используются при классификации и прогнозировании числовых и временных рядов, в метеорологии, лингвистике, радиотехнике, пищевой промышленности, в области защиты информации, в геологии, экономике, медицине, астрономии и океанологии.

Основные концепции искусственных нейронных сетей (ИНС) были предложены в 40–50-х гг. XX в. [2]–[4], в частности, многослойный персептрон – нейронная сеть прямого распространения с одним скрытым слоем. Как научное направление теория искусственных нейронных сетей была

определена в классической работе У. МакКаллока и У. Питтса [3], заложившей основы двух направлений исследований нейронных сетей. Первый был ориентирован на изучение биологических процессов в головном мозге, второй – на применение сетей для создания искусственного интеллекта. Дальнейшее развитие заключалось в создании различных типов нейронных сетей для приближения сложных функций и распознавания закономерностей, таких как ассоциативные нейронные сети, радиальные нейронные сети, рекуррентные нейронные сети, сети с самоорганизацией, нечеткие нейронные сети, которые разрабатывались и успешно адаптировались в 1970–1980-х гг. [5]–[7]. В последние десятилетия был разработан широкий спектр нейронных сетей, таких как сверточные нейронные сети (CNN) [8] и рекуррентные сети с длинной кратковременной памятью (LSTM) [9]. Хорошие результаты были достигнуты на основе применения нейронных

сетей в области обработки изображения и видео [10]–[12], в лингвистике и анализе речи [13]–[16]. Эффективное применение различных типов нейронных сетей стало возможным благодаря вычислительной мощности предоставляемой графическими процессорами и другими параллельными вычислительными платформами [16], [17], более гибким и масштабируемым программным средам [18], [19] и появившимся большим объемам доступных данных. Однако многослойный перцептрон остается достаточно популярной архитектурой нейронной сети за счет ее простоты и гибкости.

Основным направлением в изучении нейронных сетей является исследование зависимости между типом нейронной сети, ее архитектурой и результатами ее работы. В процесс построения нейронной сети входит и процесс ее обучения, т. е. задание и корректировка весов сети на каждой итерации для получения требуемого результата ее работы [20]. Обучение нейронной сети прямого распространения основывается на вычислении функции ошибки [21] методом обратного распространения ошибки [22]. В качестве методов обучения используют методы, основанные на оптимизации стохастического градиента и подразделяющиеся на методы первого и второго порядков в зависимости от порядка используемых производных функции ошибки [23].

Базовым градиентным методом первого порядка является метод наискорейшего спуска [24]. Последующие предложенные методы являются его модификациями путем изменения способа вычисления параметров обучения: метод быстрого распространения [25], метод упругого распространения, метод ускоренного градиента Нестерова [26], метод адаптивного градиента [27], метод адаптивной дельты [28], метод адаптивного момента [29].

Также важными параметрами при построении нейронной сети после выбора метода обучения являются оптимальное число нейронов на скрытом слое и способ начальной инициализации весов.

Таким образом, цель данной работы состояла в сравнении эффективности различных методов обучения на примере классифицирующих нейронных сетей и анализе влияния различных параметров сети и метода обучения на результат работы сети. В качестве исходных данных использовались серии спектров (числовых рядов), предварительно обработанных различными мате-

матическими методами для извлечения из них ряда классификационных признаков [30].

Архитектура исследуемой нейронной сети.

В рамках данного исследования для решения задачи классификации используется стандартная трехслойная архитектура нейронной сети прямого распространения. На вход подаются объекты – серии спектров, описываемые 63 признаками. Соответственно, число входных нейронов равно 63. Количество целевых классов в данной задаче 16, следовательно, на выходном слое число нейронов равно 16. Рассматривается нейронная сеть с одним скрытым слоем, размер которого первоначально определяется как среднегеометрическое от размеров внешних слоев [28] и равен 31 нейрону. На рис. 1 показана схема исследуемой нейронной сети (где X – входные признаки, W – весовые коэффициенты нейронов, S – состояние нейронов, Y – выход сети). В качестве функций активаций на скрытом и выходном слое используются гиперболический тангенс и softmax соответственно. Для обучения нейронной сети используются три подвыборки, на которые делится исходная выборка в соотношении 70:10:20 – обучающая: контрольная: тестовая (L:C:T). В совокупности, вся выборка состоит из 230 объектов, распределенных равномерно по всем классам. В качестве метода инициализации весовых коэффициентов нейронов (W) могут быть использованы такие методы, как инициализация Лекуна (LeCun) [31], Ксавьера [32], Хи (He) [33], Сакса [34], Мишкина–Матаса [35].

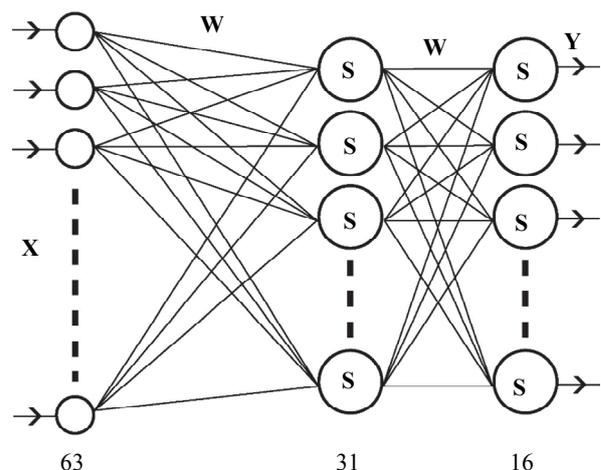


Рис. 1

Для оценки исследуемых параметров и закономерностей используется результат классификации или ошибка классификации, которые были получены на сериях из 10 нейронных сетей после

их усреднения. В данной работе ошибка вычисляется по формуле кроссэнтропии [20]. На графиках отображается ошибка по всей выборке на каждой эпохе обучения.

Градиентные методы обучения. Особенности процессов обучения нейронных сетей исследуются на примере шести наиболее распространенных методов обучения: метод наискорейшего спуска (GD), метод быстрого распространения (qProp), метод ускоренного градиента Нестерова (NAG), метод адаптивного градиента (AdaGrad), метод адаптивной дельты (AdaDelta), метод адаптивного момента (Adam).

Для всех перечисленных методов обучения вычисление изменения весов на одной итерации производится по формуле

$$(\Delta W)_t = \eta [f_t(\nabla E) + \rho W_{t-1}] + \mu (\Delta W)_{t-1},$$

где η – параметр начальной скорости обучения; $f_t(\nabla E)$ – функция от градиента ошибки; μ – параметр момента; $(\Delta W)_{t-1}$ – изменение весов на предыдущей итерации; ρ – параметр регуляризации; W_{t-1} – значения весов на предыдущей итерации;

$E = \frac{1}{n} \sum_{i=1}^n x_i \ln x_i$ – среднее значение кроссэнтропии.

Перечисленные методы обучения различаются способом вычисления двух параметров, оказывающих наибольшее влияние на процесс обучения, а именно, параметра скорости обучения и вида функции от градиента ошибки [29].

1. *Метод наискорейшего спуска (GD).*

Скорость обучения: $0 < \eta < 1$. Функция от градиента ошибки: $(\nabla E)_t$.

2. *Метод быстрого распространения (qProp).*

Скорость обучения:

$$\eta = \begin{cases} \eta_0, & (\Delta W = 0) \vee (-\Delta W(\nabla E + \rho W_{t-1}) > 0), \\ \text{при } \eta_0 \in (0.01, 0.6); \\ 0, & \text{при } \eta_0 \notin (0.01, 0.6). \end{cases}$$

Функция от градиента ошибки: $(\nabla E)_t$.

3. *Метод ускоренного градиента Нестерова (NAG).*

Скорость обучения: $0 < \eta < 1$. Функция от градиента ошибки: $(\nabla E)_t(W_{t-1} + \mu(\Delta W)_{t-1})$.

4. *Метод адаптивного градиента (AdaGrad):*

Скорость обучения: $0 < \eta < 1$. Функция от градиента ошибки: $\frac{(\nabla E)_t}{\sqrt{\sum_{i=1}^t (\nabla E)_i^2}}$.

5. *Метод адаптивной дельты (AdaDelta).*

Скорость обучения: $0 < \eta < 1$. Функция от

градиента ошибки: $\left(\frac{\sqrt{D_t}}{\sqrt{S_t}}\right)(\nabla E)_t$, где $S_t = \alpha S_{t-1} + (1 - \alpha)(\nabla E)_t^2$; $D_t = \beta D_{t-1} + (1 - \beta)(\Delta W)_{t-1}^2$, $S_0 = 0$, $D_0 = 0$, $\alpha = 0.9$, $\beta = 0.9$.

6. *Метод адаптивного момента (Adam).*

Скорость обучения: $0 < \eta < 1$. Функция от

градиента ошибки: $\frac{D_t}{1 - \beta} \sqrt{\frac{1 - \alpha}{S_t}}$, где $S_t = \alpha S_{t-1} + (1 - \alpha)(\nabla E)_t^2$; $D_t = \beta D_{t-1} + (1 - \beta)(\nabla E)_{t-1}$, $S_0 = 0$, $D_0 = 0$, $\alpha = 0.999$, $\beta = 0.9$.

Здесь и далее сравнение методов обучения проводится при следующих значениях параметров: начальная скорость обучения $\eta = 0.01$, параметр момента $\mu = 0.01$, параметр регуляризации $\rho = 0.001$. Далее эти параметры оптимизируются в зависимости от метода обучения.

Анализ влияния параметров метода обучения. В первую очередь необходимо сравнить нейронные сети, обученные различными методами при одних и тех же параметрах на относительно большом числе эпох.

На данном этапе исследуемая нейронная сеть имеет три слоя (входной, скрытый, выходной) с размерами 63, 31 и 16 нейронов соответственно; в качестве метода инициализации W используется инициализация Ксавьера; входящая выборка делится на три подвыборки (обучающая, контрольная и тестовая) в соотношении 70:10:20.

Как следует из графика зависимости ошибки выхода сети для различных методов обучения (рис. 2), наименьшую ошибку классификации демонстрирует метод Adam. Почти все сети, обученные разными методами, достигают оптимального значения ошибки на 600-й эпохе обучения, поэтому при дальнейшем анализе будем ограничиваться обучением до 600-й эпохи.

Далее производится выбор оптимального параметра скорости обучения (η) для каждого метода. Он проводился на дискретном ряде значений [0.001, 0.005, 0.01, 0.05, 0.1, 0.5]. Соответствующие оптимальные значения скорости обучения для каждого метода представлены в таблице.

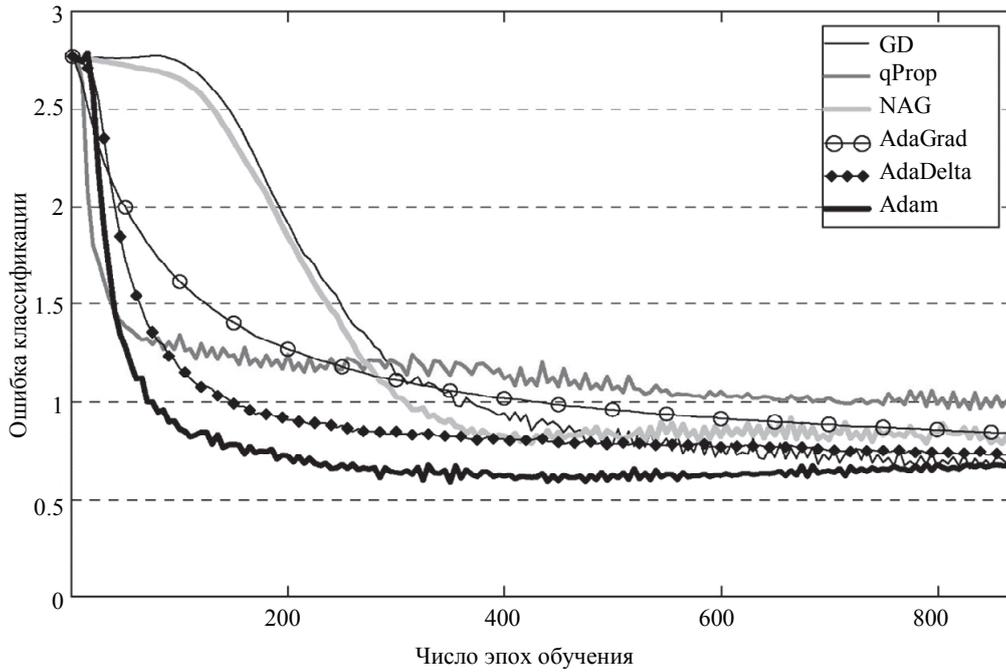


Рис. 2

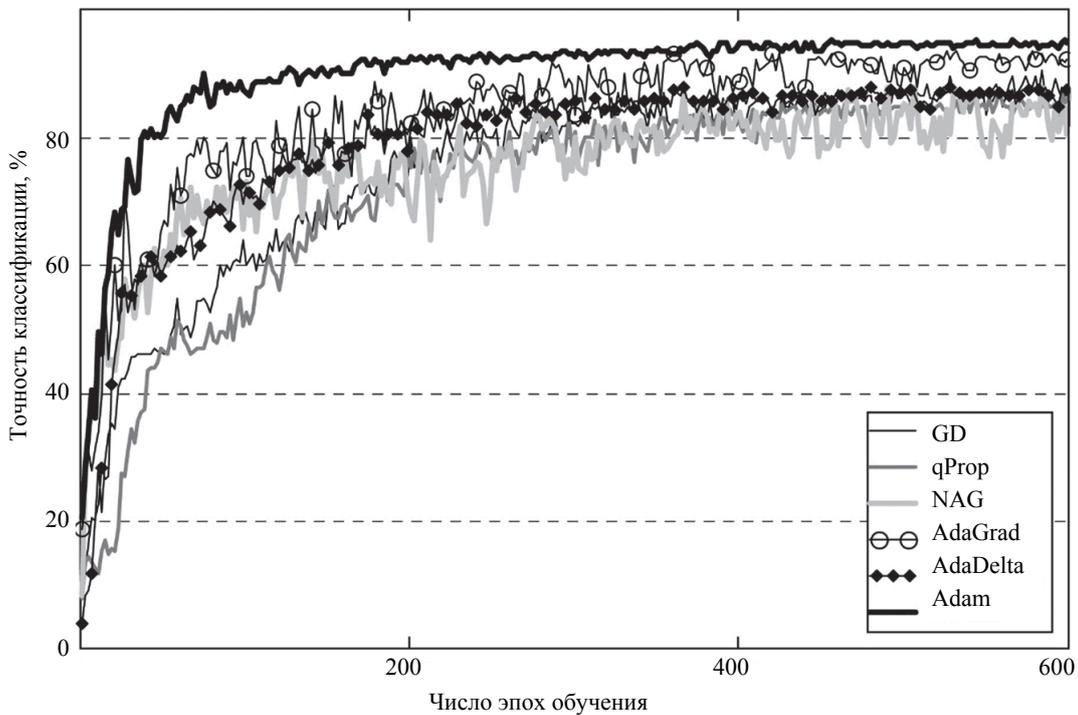


Рис. 3

Метод обучения	Оптимальная скорость обучения
GD	0.1
qProp	0.05
NAG	0.5
AdaGrad	0.1
AdaDelta	0.05
Adam	0.01

Основным показателем корректности и эффективности работы нейронной сети является точность классификации. Результаты расчета зависимости точности классификации от эпохи

обучения при выбранном оптимальном параметре скорости обучения (см. таблицу) для каждого метода представлены на рис 3. Расчеты проводились на трехслойных нейронных сетях с размером слоев 63, 31 и 16 соответственно. Значения усреднены по результатам обучения 10 нейронных сетей в серии. Общее число эпох обучения равно 600. Метод инициализации весов, деление выборки на подвыборки и значения параметров момента и регуляризации соответствуют указанным ранее для рис. 1.

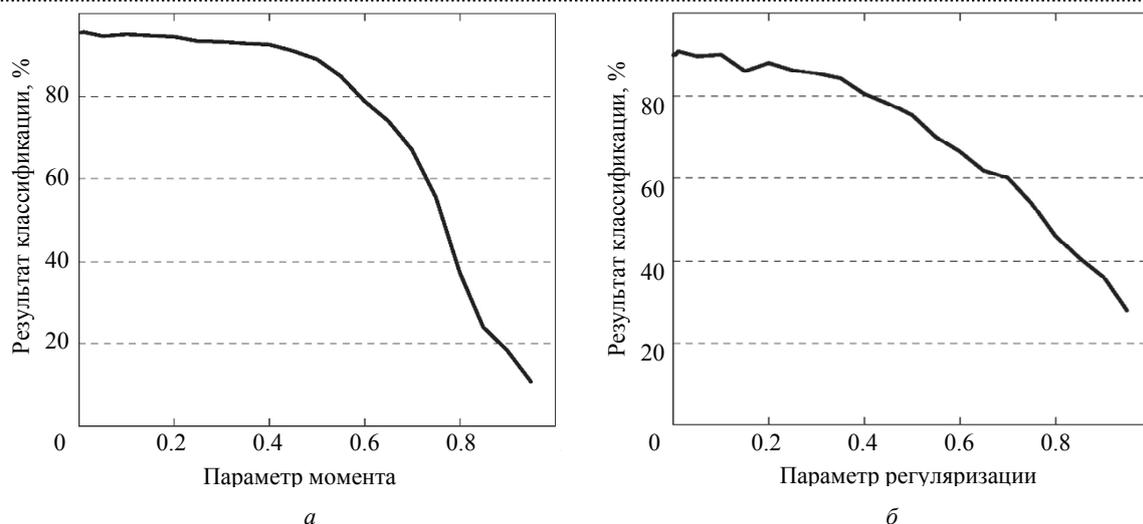


Рис. 4

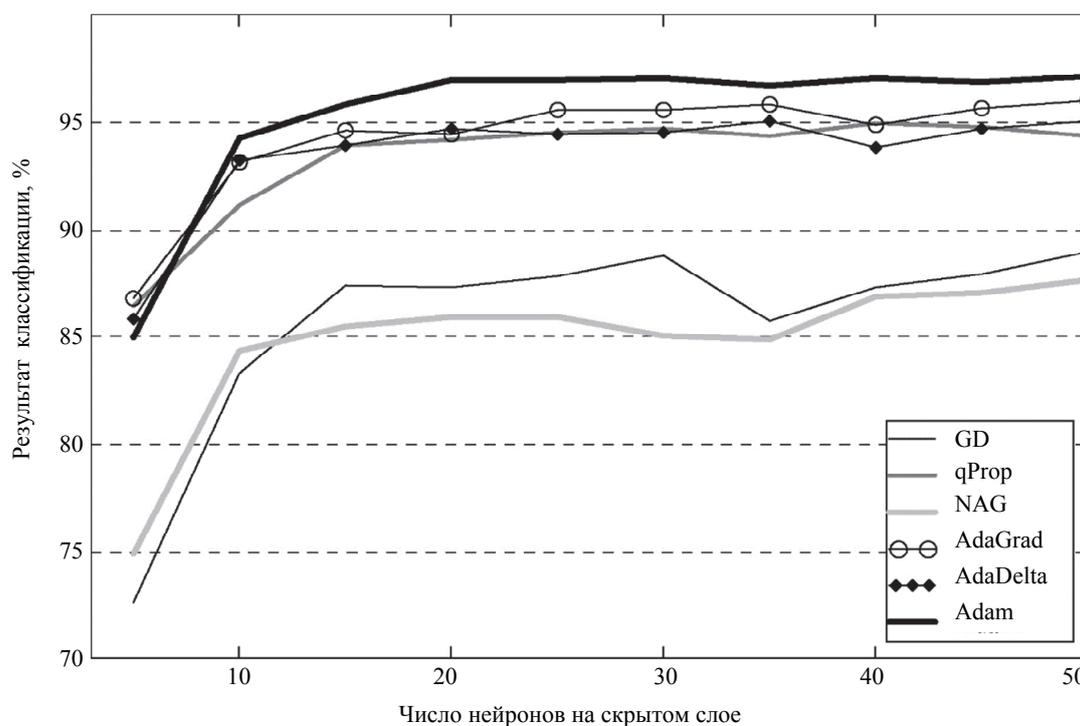


Рис. 5

Из представленных на рис. 3 графиков видно, что нейронная сеть, обученная методом адаптивного момента (Adam), имеет более высокую точность классификации и достигает 93 % уже на сотой эпохе. Нейронные сети, обученные другими методами, достигают точности 85–90 % только на последних эпохах (500–600-й), т. е. обучение происходит гораздо медленнее. Это является следствием того, что метод адаптивного момента, являясь по существу методом первого порядка, имеет эффективность, близкую к методам второго порядка.

Для дальнейшего анализа необходимо удостовериться, что выбранные ранее параметры обучения являются оптимальными. Для наиболее эф-

фективного метода обучения – метода адаптивного момента (Adam) – были построены графики зависимости точности классификации от значения параметра момента (рис. 4, а) и от параметра регуляризации (рис. 4, б). Из представленных графиков следует, что оба параметра выбраны верно (0.01 и 0.001 соответственно), так как точность классификации при этих значениях достигает максимального значения.

Анализ влияния других параметров на процесс обучения. После выбора оптимальных параметров обучения для каждого метода следует проверить, насколько методы остаются эффективными при изменении архитектуры сети. Напри-

мер, рассмотрим зависимость точности классификации нейронной сети от размера скрытого слоя при соответствующих оптимальных параметрах обучения для каждого метода. Для этого рассчитывается график зависимости точности классификации нейронной сети от размера скрытого слоя, показанный на рис. 5. Очевидно, что метод Adam показывает наилучший результат при количестве нейронов больше 10, а также, что изначально выбранное среднегеометрическое значение для количества нейронов на скрытом слое (31 нейрон) действительно является оптимальным для данного метода в рамках решаемой задачи.

Таким образом, оптимальным методом обучения нейронной сети для решения задачи классификации на заданной выборке является метод адаптивных моментов с параметрами обучения $\eta = 0.01$, $\mu = 0.01$, $\rho = 0.001$, и для дальнейших исследований будет использован именно этот метод.

Следующим важным этапом в исследовании эффективности работы классификатора является выбор метода инициализации весовых коэффициентов нейронов. В данной работе рассмотрены три метода:

- инициализация Лекуна (LeCun) – $W_i \sim U\left(-\frac{1}{\sqrt{n_i}}, \frac{1}{\sqrt{n_i}}\right)$;

- инициализация Ксавьера (Xavier) – $W_i \sim U\left(-\frac{\sqrt{6}}{\sqrt{n_i + n_{i+1}}}, \frac{\sqrt{6}}{\sqrt{n_i + n_{i+1}}}\right)$;

- инициализация Хи (He) – $W_i \sim N\left(0, \frac{2}{\sqrt{n_i}}\right)$,

где i – номер слоя нейронной сети; n – размер слоя; U – равномерное распределение; N – нормальное распределение.

Результаты применения различных методов инициализации представлены на рис. 6. Нейронные сети с инициализацией Ксавьера показывают наилучший результат с разницей в 3 %, возможно, это связано с тем, что данный вид инициализации учитывает размеры двух ближайших слоев.

Считается, что на результат классификации также достаточно сильно может повлиять способ деления исходной выборки, которая подается на вход нейронной сети при обучении. Обычно выборка делится на три части: обучающую, контрольную и тестовую (три подвыборки). Однако тестовая подвыборка может состоять и из всей выборки, а исходная выборка тогда делится лишь на две части (две подвыборки). При выбранном методе обучения (Adam) на конечной 600-й эпохе обучения нейронная сеть на двух подвыборках дает точность классификации 94 %, а на трех

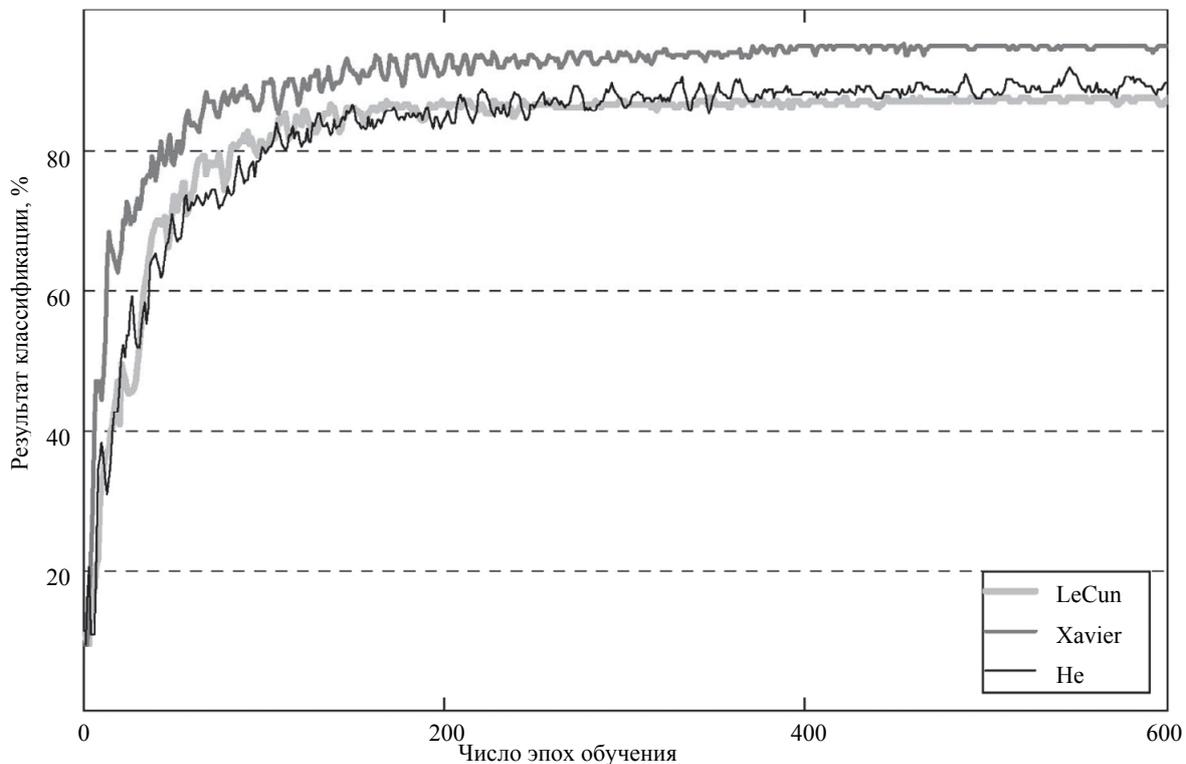


Рис. 6

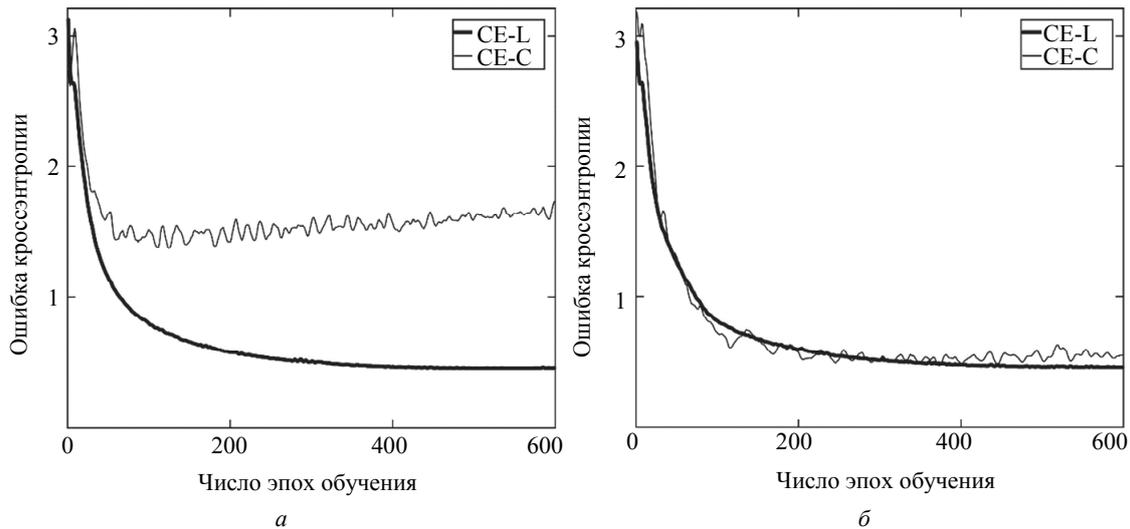


Рис. 7

подвыборках 92 %. Лучший результат при делении на две двух подвыборок показывает, что увеличение размера выборки для обучения нейронной сети дает положительный результат в виде увеличения точности классификации.

Обсуждение. Известно, что все типы искусственных нейронных сетей в той или иной степени страдают нестабильностью результатов, поэтому для исследования различных закономерностей результаты работы нейронных сетей обычно усредняются по серии сетей. Однако для практического использования обученных сетей и подачи на них новых объектов полную серию нейронных сетей использовать невозможно, особенно, если обученные нейросети имплементировать в аппаратные технологии. Кроме того, если в серии окажется одна сильно отличающаяся по результатам обучения нейронная сеть, она может значительно повлиять на результат работы всей серии. Непосредственно классифицирующие нейронные сети стандартной архитектуры (многослойный персептрон) обладают такими традиционными недостатками, как появление ложных решений и плохая способность к обобщению. Однако методик выявления в серии нейронных сетей с наилучшей способностью к обобщению, т. е. сети, которая классифицирует имеющиеся и новые объекты наиболее корректно, на сегодняшний день не существует.

Для демонстрации подобных ситуаций на рис. 7, а и б представлены две нейронные сети (А и Б соответственно). Здесь CE – кроссэнтропия, L – обучающая выборка, С – контрольная выборка. Графики изменения ошибки классификации в процессе обучения методом Adam и достигнутая точность классификации показывают, что нейросеть А обучена лучше. Однако только по графику

ошибки сети невозможно в полной мере оценить, например, качество обобщения полученной сети, поскольку по данному графику не видно, насколько уверенно нейросеть относит объект к тому или иному классу. При подаче на вход нейронной сети объектов, не принадлежащих к классам, на которых данная сеть обучалась, возможен вариант, когда нейронная сеть Б покажет лучший результат распознавания предложенных объектов, чем нейронная сеть А. Следовательно, необходимо разработать методику более глубокого тестирования нейронной сети на качество обобщения для выявления наиболее корректно обученной сети.

В описываемой работе на основе построения серий нейронных сетей было изучено влияние отдельных компонентов архитектуры нейронной сети на результаты классификации. Важным фактором при построении сети оказывается выбор метода обучения и его оптимальных параметров, в частности, параметра скорости обучения. При правильном выборе структуры сети и метода обучения точность классификации даже на малых выборках с большой дисперсией признаков может достигать 95 %, и уже на первых 200 эпохах достигает 80 %. Также показано, что на конечный результат существенно могут повлиять такие параметры, как метод инициализации весовых коэффициентов нейронов и способ деления входной выборки. По результатам проведенных исследований сделан вывод о необходимости разработки новой расширенной методики тестирования классифицирующих нейронных сетей на качество обобщения для обеспечения возможности их использования для корректной классификации новых объектов, не принадлежащих к классам, на которых проводилось обучение.

СПИСОК ЛИТЕРАТУРЫ

1. Deep residual learning for image recognition / K. He, X. Zhang, S. Ren, J. Sun // Proc. of the IEEE conf. on computer vision and pattern recognition. Las Vegas, 2016. P. 770–778.
2. Hebb D. O The organization of behavior. New York: Wiley, 1949.
3. McCulloch W. S., Pitts W. A logical calculus of the ideas immanent in nervous activity // The bulletin of mathematical biophysics. 1943. Vol. 5, № 4. P. 115–133.
4. Rosenblatt F. The perceptron: a probabilistic model for information storage and organization in the brain // Psychological review. 1958. № 65. P. 386–408.
5. Elman J. L. Finding structure in time // Cognitive science. 1990. Vol. 14, № 2. P. 179–211.
6. Jordan M. I. Serial order: a parallel distributed processing approach // Advances in Psychology. 1997. Vol. 121. P. 471–495.
7. Werbos P. J. Maximizing long-term gas industry profits in two minutes in Lotus using neural network methods // IEEE Transactions on Systems, Man, and Cybernetics. 1989. Vol. 19, № 2. P. 315–333.
8. Chua L. O., Roska T. The CNN paradigm // IEEE Transactions on Circuits and Systems I: Fundamental Theory and Applications. 1993. Vol. 40, № 3. P. 147–156.
9. Hochreiter S., Schmidhuber J. Long short-term memory // Neural computation. 1997. Vol. 9, № 8. P. 1735–1780.
10. Large-scale video classification with convolutional neural networks / A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, L. Fei-Fei // Proc. of the IEEE conf. on Computer Vision and Pattern Recognition. 2014. P. 1725–1732. URL: https://www.cv-foundation.org/openaccess/content_cvpr_2014/html/Karpathy_Large-scale_Video_Classification_2014_CVPR_paper.html (дата обращения: 12.03.19).
11. Going deeper with convolutions / C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich // Proc. of the IEEE conf. on computer vision and pattern recognition. 2015. P. 1–9. URL: https://www.cv-foundation.org/openaccess/content_cvpr_2015/html/Szegedy_Going_Deeper_With_2015_CVPR_paper.html (дата обращения: 12.03.19).
12. Convolutional learning of spatio-temporal features / G. W. Taylor, R. Fergus, Y. LeCun, C. Bregler // European conf. on computer vision. Berlin, Heidelberg: Springer, 2010. P. 140–153.
13. Google's neural machine translation system: Bridging the gap between human and machine translation / Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, L. Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. Corrado, M. Hughes, J. Dean // arXiv preprint arXiv:1609.08144. 2016. URL: <https://arxiv.org/abs/1609.08144> (дата обращения: 02.04.19).
14. Deep speech: Scaling up end-to-end speech recognition / A. Hannun, C. Case, J. Casper, B. Catanzaro, G. Diamos, E. Elsen, R. Prenger, S. Satheesh, S. Sengupta, A. Coates, A. Ng // arXiv preprint arXiv:1412.5567. 2014. URL: <https://arxiv.org/abs/1412.5567> (дата обращения: 21.02.19).
15. Efficient estimation of word representations in vector space / T. Mikolov, K. Chen, G. Corrado, J. Dean // arXiv preprint arXiv:1301.3781. 2013. URL: <https://arxiv.org/abs/1301.3781> (дата обращения: 21.02.19).
16. Schmidhuber J. Deep learning in neural networks: An overview // Neural networks. 2015. Vol. 61. P. 85–117.
17. Goodfellow I., Bengio Y., Courville A. Deep learning. Cambridge: MIT press, 2016. Vol. 1. P. 12–26
18. Boyd S., Vandenberghe L. Convex optimization. Cambridge Massachusetts: Harvard University Press, 2004. 730 p.
19. cuDNN: Efficient primitives for deep learning / S. Chetlur, C. Woolley, P. Vandermerch, J. Cohen, J. Tran, B. Catanzaro, E. Shelhamer // arXiv preprint arXiv:1410.0759. 2014. URL: <https://arxiv.org/abs/1410.0759> (дата обращения: 14.03.19).
20. Осовский С. Нейронные сети для обработки информации / пер. с польск. М.: Финансы и статистика, 2002. 344 с.
21. Hagan M. T., Menhaj M. B. Training feedforward networks with the Marquardt algorithm // IEEE transactions on Neural Networks. 1994. Vol. 5, № 6. P. 989–993.
22. Rumelhart D. E., McClelland J. L. Parallel distributed processing: explorations in the microstructure of cognition. Cambridge: MA, MIT Press. 1986. Vol. 1.
23. Battiti R. First-and second-order methods for learning: between steepest descent and Newton's method // Neural computation. 1992. Vol. 4, № 2. P. 141–166.
24. Amari S. Backpropagation and stochastic gradient descent method // Neurocomputing. 1993. Vol. 5, № 4–5. P. 185–196.
25. Vrahatis M. N., Magoulas G. D., Plagianakos V. P. Convergence analysis of the quickprop method // Neural Networks, 1999. IJCNN'99. Intern. Joint Conf. on. IEEE. Washington, USA, 1999. Vol. 2. P. 1209–1214.
26. Botev A., Lever G., Barber D. Nesterov's accelerated gradient and momentum as approximations to regularised update descent // Neural Networks (IJCNN), 2017 Intern. Joint Conf. on. IEEE. AK, USA, 2017. P. 1899–1903.
27. Duchi J., Hazan E., Singer Y. Adaptive subgradient methods for online learning and stochastic optimization // J. of Machine Learning Research. 2011. Vol. 12. P. 2121–2159.
28. Zeiler M. D. ADADELTA: an adaptive learning rate method // arXiv preprint arXiv:1212.5701. 2012. URL: <https://arxiv.org/abs/1212.5701> (дата обращения: 12.03.19).
29. Kingma D. P., Ba J. Adam: A method for stochastic optimization // arXiv preprint arXiv:1412.6980. 2014. URL: <https://arxiv.org/abs/1412.6980> (дата обращения: 21.02.19).
30. Применение линейного дискриминантного анализа для классификации цианобактерий по спектрам собственной флуоресценции / Т. Р. Жангиров, А. С. Перков, Н. Ю. Григорьева, Л. В. Чистякова, А. А. Лисс // Изв. СПбГЭТУ «ЛЭТИ». 2018. № 5. С. 45–55.

31. Efficient backprop / Y. A. LeCun, L. Bottou, G. B. Orr, K. R. Muller // Neural networks: Tricks of the trade. Berlin, Heidelberg: Springer, 2012. P. 9–48.

32. Glorot X., Bengio Y. Understanding the difficulty of training deep feedforward neural networks // Proc. of the thirteenth intern. conf. on artificial intelligence and statistics. Sardinia: Domus de Maria, 2010. P. 249–256.

33. The loss surfaces of multilayer networks / A. Choromanska, M. Henaff, M. Mathieu, G. B. Arous,

Y. LeCun // Artificial Intelligence and Statistics. 2015. № 2. P. 192–204.

34. Saxe A. M., McClelland J. L., Ganguli S. Exact solutions to the nonlinear dynamics of learning in deep linear neural networks // arXiv preprint arXiv:1312.6120. 2013. URL: <https://arxiv.org/abs/1312.6120> (дата обращения: 14.03.19).

35. Mishkin D., Matas J. All you need is a good init // arXiv preprint arXiv:1511.06422. 2015. URL: <https://arxiv.org/abs/1511.06422> (дата обращения: 21.02.19).

A. S. Perkov, T. R. Zhangirov, A. A. Liss
Saint Petersburg Electrotechnical University «LETI»

N. Yu. Grigoryeva
Saint Petersburg SRCES RAS

L. V. Chistyakova
Center for Culture Collection of Microorganisms SPSU

COMPARISON OF NEURAL NETWORK TRAINING METHODS IN THE CLASSIFICATION PROBLEM

Based on the analysis of a series of feedforward artificial neural networks, a method has been developed for determining the optimal neural network architecture for the task of classifying cyanobacterial strains according to the fluorescence spectra. The analysis of six gradient methods of training neural networks and their parameters was carried out, the optimal number of neurons in the hidden layer for neural networks trained by each of the methods was found, various methods of initializing the weights of neurons and methods for splitting the initial sample into training, test and control samples were evaluated. The choice of the optimal architecture was carried out on the basis of the classification results, namely, on the basis of the classification accuracy graphs and the classification error graphs. The research was conducted on the example of recognition of 16 classes, representing 16 strains of cyanobacteria. A number of shortcomings were identified in the method of testing feedforward neural networks and directions for additional researches of neural networks for classification in terms of extending the testing methodology of their internal logic were determined.

Neural networks, training methods, machine learning, classification problems, initialization methods

УДК 519.688

И. А. Посов, В. Е. Допира
Санкт-Петербургский государственный электротехнический
университет «ЛЭТИ» им. В. И. Ульянова (Ленина)

Методы поиска плагиата в кодах программ

Изучены такие виды представления данных для поиска плагиата в кодах программ, как текст без преобразований, n-граммы и токены. На языке программирования Python реализовано разбиение текстов программ из массива решений студентов на токены. Сформулированы требования к алгоритмам обнаружения плагиата. Проведен обзор метрик для обнаружения плагиата в текстах программ. Выделены преимущества и недостатки для каждой метрики. Сравнение проведено по критериям: время, память, вероятность найти пару похожих программ, вероятность того, что найденная пара будет действительно похожа. После сравнения метрик: численных значений атрибутов, наибольшей общей подпоследовательности, расстояния Жаккара, расстояния Левенштейна и расстояния Колмогорова, между собой для реализации выбран расчет расстояния Левенштейна. На языке программирования Python реализован алгоритм расчета расстояния Левенштейна для списка токенов. Полученные результаты показывают, насколько тексты программ похожи между собой.

Плагиат, метрики для обнаружения плагиата, токен

Цифровой контент все чаще подвергается копированию. Люди цитируют друг друга, также соавторство порождает увеличение копий работ.

Соответственно, возникает необходимость установления авторских прав на интеллектуальный труд, а также проверки новых работ на плагиат.