

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/377400611>

Влияние настраиваемых параметров полносвязной нейронной сети на качество предсказания для задачи классификации литотипов (INFLUENCE OF NEURAL NETWORK PARAMETERS FOR THE QUALITY)

Article · November 2022

DOI: 10.24412/2073-0667-2023-1-48-59

CITATIONS

0

READS

10

2 authors, including:



Georgy Kossov

Moscow Institute of Physics and Technology

12 PUBLICATIONS 11 CITATIONS

SEE PROFILE

INFLUENCE OF NEURAL NETWORK PARAMETERS FOR THE QUALITY OF PREDICTION FOR THE TASKS OF AUTOMATIC LITHOTYPE DESCRIPTION

G. A. Kossov, I. A. Seleznev

LLC “TCS”,
125171, Moscow, Russia

DOI: 10.24412/2073-0667-2023-1-48-59

EDN: QQFRGC

Machine learning methods are widely used for solving problems of interpreting and describing geological and geophysical data. One of them is automatic lithology extraction during the analysis of a whole core photographs. In this paper we propose to analyze the parameters that represent the textural and color features of the images. The advantage of this approach is that it allows online training and retraining of the classification model. Among the existing classification methods, such as boosting, random forests, support vector machines, neural networks are preferred for their universality and implementation in various sets of programming tools. The application of neural networks requires the user to have a clear understanding of the modelling goals, because an important factor is the choice of model architecture.

There are many parameters that are set by the user, and all of them affect the quality of the prediction. Therefore, the purpose of this research is to study the behavior of networks with various configurations and to find any common regularities. The paper considers the problem of classifying lithotypes using fully connected neural networks. The data for processing are color and textural features that were obtained as a result of the processing of whole core images. Thus, we consider the classification task of training examples with 48 features into 20 classes corresponding to certain lithotypes. The test sample consisted of 2998 elements. We trained the model on samples consisting of 10,000 and 1,000 elements, respectively. The hyperparameters of the model include loss function, optimization method, activation function, batch size, number of epochs, number of hidden layers, and number of neurons in a layer. Based on a given issue, it is already possible to explain the choice of one or another parameter or function in advance. For the classification problem the optimal way is using ReLU and LogSoftMax activation function. CrossEntropyLoss was used as a loss function. This loss function combines LogSoftMax and NLLLoss, so the use of LogSoftMax is also justified by simplifying the calculation of CrossEntropyLoss. We use the Adam algorithm as the method of optimization. The quality of the model was evaluated using the f1-score metric. According to the results of training a model with a fixed number of layers and nodes, but with a different batch size, it was figured out that the optimal batch size consists of 256 elements. Based on this assumption we determined that 30 epochs are enough to train the model. All in all among a large set of network hyperparameters it is complicated to determine the exact number of network elements, i.e. the number of layers and neurons. Therefore, in the current research we study the dependence of f1-score and the value of the loss function on the number of nodes in the layer. The paper shows that an increase in the number of neurons definitely leads to a gain in quality. F1-score equals 1 for all cases after 10 neurons in a layer. Moreover, a model with incorrect number of layers can be improved by increasing the amount of neurons in each layer. Increasing the number of layers allows the model to construct a more complex approximation,

which can improve the quality of the prediction. However, as the number of layers increases, there is a risk of network overfitting and the appearance of local minima of the error function that leads to training problems. Thus, the number of nodes in a layer is the defining parameter and we should set this parameter up first. An important factor in the model training is the time spending. In this research, we propose a following estimate of the algorithm complexity. Besides, we have studied the influence of the number of layers (m) and nodes (n). The estimate is given in terms of O -notation. It is shown that the number of performed operations increase linearly $O(m)$ in the number of layers and cubically $O(n^3)$ in the number of neurons. Consequently, with relation to the number of operations it is preferably to increase the number of network layers. However, many elements does not guarantee the rise in the f1-score. The predictions of some classification algorithms (for example, boosting or random forest) are highly dependent on the first initialization of the parameters. In our case, the dependence of the loss value on the random initialization of the neural network weights was investigated. We use the Epps-Pally test to check the normality of the loss value distribution. Tests have shown that the distribution of the value of the loss is not a Gaussian one. This fact should be taken into account in setting the requirement for the reproducibility of experiments result. The starting model weights should be initialized accordingly.

Key words: neural network, lithotype description, core analysis, hyperparameters, supervised learning.

References

1. Thomas A., et al. Automated lithology extraction from core photographs // First Break. 2011. V. 29. N 6.
2. Baraboshkin E. E., et al. Deep convolutions for in-depth automated rock typing // Computers and Geosciences. 2020. V. 135. P. 104330.
3. Abashkin V. V., et al. Quantitative analysis of whole core photos for continental oilfield of Western Siberia // SPE Russian Petroleum Technology Conference, OnePetro, 2020.
4. Seleznev I. A., et al. Joint Usage of Whole Core Images Obtained in Different Frequency Ranges for the Tasks of Automatic Lithotype Description and Modeling of Rocks' Petrophysics Properties // Geomodel 2020, European Association of Geoscientists and Engineers, 2020. V. 2020. N 1. P. 1–5.
5. Amirgaliev E. N., i dr. Integratsiya algoritmov raspoznavaniya litologicheskikh tipov // Problemy informatiki. 2013. № 4 (21). S. 11–20.
6. Chang B. T. T., i dr. Klassifikatsiya izobrazhenii na osnove primeneniya tsvetovoi informatsii, veivlet-preobrazovaniya Khaara i mnogosloinoi neironnoi seti // Problemy informatiki. 2011. № 5. S. 81–86.
7. Mukhamedgaliev A. F., Razakova M. G., Smirnov V. V. Sozdanie i razvitiye geoinformatsionnykh tekhnologii tematicheskoi interpretatsii dannykh radiolokatsionnogo zondirovaniya s ispol'zovaniem matematicheskikh metodov i vychislitel'nykh algoritmov teksturnoi klassifikatsii i neironnykh setei // Problemy informatiki. 2012. № 3. S. 69–73.
8. Manurangsi, P., Reichman, D. The computational complexity of training ReLU (s). arXiv:1810.04207v2 [cs.CC]. 2018.
9. Kingma, D. P., Ba, J. Adam: A method for stochastic optimization. arXiv:1412.6980 [cs.LG]. 2014.
10. Maksimushkin V. V., Arzamascev A. A. Sravnitel'naya ocenka vychislitel'noj slozhnosti obucheniya iskusstvennoj neironnoj seti s zhestkim yadrom i seti s klassicheskoy strukturoj // Vestnik Tambovskogo universiteta. Seriya: Estestvennye i tekhnicheskie nauki. 2006. T. 11. № 2. S. 190–197.
17. Makienko D., Seleznev I., Safonov I. The effect of the imbalanced training dataset on the quality of classification of lithotypes via whole core photos // Creative Commons License Attribution. 2020. V. 4.

11. Bernard, S., Heutte, L., Adam, S. Influence of hyperparameters on random forest accuracy // International workshop on multiple classifier systems, Springer, Berlin, Heidelberg, 2009. P. 171–180.
12. Epps, T. W., Pulley, L. B. A test for normality based on the empirical characteristic function // Biometrika. 1983. V. 70. N 3. P. 723–726.
13. GOST R. 5479–2002. Statisticheskie metody. Proverka otkloneniya raspredeleniya veroyatnosti ot normal'nogo raspredeleniya / M.: Izd-vo standartov, 2002.
14. Lemeshko B. YU. Kriterii proverki otkloneniya raspredeleniya ot normal'nogo zakona. Rukovodstvo po primeneniyu / B. YU. Lemeshko. M.: OOO «Nauchno-izdatel'skij centr INFRA-M», 2015. 160 s.



ВЛИЯНИЕ НАСТРАИВАЕМЫХ ПАРАМЕТРОВ ПОЛНОСВЯЗНОЙ НЕЙРОННОЙ СЕТИ НА КАЧЕСТВО ПРЕДСКАЗАНИЯ ДЛЯ ЗАДАЧИ КЛАССИФИКАЦИИ ЛИТОТИПОВ

Г. А. Коссов, И. А. Селезнев

ООО «ТКШ»,
125171, Москва, Россия

УДК 519.7

DOI: 10.24412/2073-0667-2023-1-48-59

EDN: QQFRGC

В работе рассматривается задача классификации литотипов с помощью полносвязной нейронной сети. Тренировочными и тестовыми данными являются цветовые и текстурные признаки, полученные в результате анализа полноразмерных изображений керна. Преимущества такого подхода заключаются в возможности как обучать модель в реальном времени, так и адаптировать ее к новому набору данных посредством дообучения. Число признаков каждого тренировочного примера равнялось 48, число классов, соответствующих определенным литотипам — 20. В работе показано, что для задачи классификации с помощью нейронных сетей наиболее значимым параметром архитектуры модели является число слоев и узлов. В работе была предложена оценка сложности алгоритма в терминах O -нотации. Показано, что число выполняемых операций растет линейно $O(m)$ по числу слоев и кубически $O(n^3)$ по числу нейронов в слое. Однако с точки зрения качества предсказания модели увеличение числа слоев не приводит к лучшим результатам. При анализе зависимости метрики f1-score от числа узлов для различных слоев было получено, что увеличение числа нейронов приводит к выигрышу в качестве предсказания.

Ключевые слова: нейронные сети, классификация литотипов, анализ керна, гиперпараметры, обучение с учителем.

Введение. Методы машинного обучения широко применяются для решения задач интерпретации и описания геолого-геофизических данных. Одной из таких задач является автоматическое литотипирование горных пород при анализе фотографий полноразмерного керна [1]. Решение данной задачи классификации значительно упростит работу привлекаемых специалистов — геологов и петрофизиков. Распространенным подходом для решения вышеупомянутой задачи является использование сверточных сетей [2]. В данной работе предлагается применять полносвязную нейронную сеть, на вход которой подаются признаки, отражающие текстурные и цветовые свойства изображений. Методика получения признаков из полноразмерного изображения керна описана в работах [3, 4]. Выбор такого подхода обоснован тем, что возможно проводить быстрое, оперативное обучение

Статья по докладу на Международной конференции «Марчуковские научные чтения–2022», Россия, Новосибирск, 03.10.2022–07.10.2022.

и дообучение модели. Применение НС требует от пользователя четкого понимания поставленных целей, т. к. важным фактором является подбор архитектуры модели, которая зависит от решаемой задачи. Существует множество параметров, которые задаются пользователем, и все они влияют на качество предсказания. Несмотря на активное применение НС [5–7] для решения задачи классификации, в работах подробно не рассматривается методика выбора той или иной модели. Поэтому целью данной работы является исследование поведения сетей различных конфигураций и выявление каких-либо общих закономерностей. Это позволит более глубоко понять принцип действия алгоритмов, а также разработать методику по подбору оптимальных гиперпараметров НС для задачи классификации литотипов.

1. Используемая модель. В работе рассматривается задача классификации литотипов с помощью НС. Тестовая выборка состояла из 2998 элементов. Обучение модели проводилось на выборках, состоящих из 10000 и 1000 элементов соответственно. Количество признаков, т. е. размер входного слоя — 48, количество классов — 20. Настраиваемые параметры, позволяющие управлять процессом обучения модели, называются гиперпараметрами. К ним относятся: функция потерь, метод оптимизации, функция активации нейрона, размер батча, количество эпох обучения, количество скрытых слоев и количество нейронов в слое. Безусловно, все они оказывают существенное влияние на результаты работы классификатора, и задача построения предсказывающей модели сводится к поиску оптимальных параметров и функций. Однако, исходя из постановки задачи, можно уже заранее обосновать выбор того или иного параметра или функции. К примеру, функция, определяющая выходной сигнал нейрона, т. е. функция активации, выбирается с точки зрения вычислительной сложности алгоритма. В работе [8] показано, что для скрытых слоев с точки зрения вычислительной сложности оптимально применять ReLU. В качестве функции активации выходного слоя мы использовали SoftMax:

$$\text{SoftMax}(y_i) = \frac{\exp(y_i)}{\sum_j^N (\exp(y_j))}, \text{ где } N \text{ — число выходов сети.}$$

Выходное значение каждого нейрона уже отнормировано в диапазоне от 0 до 1, причем сумма значений всех выходных нейронов равняется 1, так что выход сети можно считать вероятностью принадлежности одному из классов. Однако в настоящей работе применяется модификация данной функции активации LogSoftMax:

$$\text{LogSoftMax}(y_i) = \frac{\ln(\exp(y_i))}{\sum_j^N (\exp(y_j))}$$

С помощью вышеупомянутого подхода решается проблема потери значения, когда отрицательные входные данные SoftMax велики по модулю, что может приводить к округлению выхода до нуля. В качестве функции потерь используется CrossEntropyLoss:

$$CE(\mathbf{p}, \mathbf{t}) = - \sum_{c=1}^N t_c \cdot \ln(p_c),$$

где \mathbf{t} — метка класса, а \mathbf{p} — значение функции активации.

Эта функция потерь объединяет LogSoftMax и NLLLoss, поэтому применение LogSoftMax обосновано еще и упрощением расчета CrossEntropyLoss. Для оптимизации функции потерь применяется алгоритм оптимизации Adam [9], являющийся модифицированной версией SGD:

$$w^{t+1} = w^t - a \frac{EMA_{b_1}(\nabla f)^t}{\sqrt{EMA_{b_2}(\nabla f^2)^t} + \epsilon},$$

где w^t, w^{t+1} — веса до и после обновления соответственно, $a = 1 * 10^{-3}$, $b_1 = 0,9$, $b_2 = 0,999$, $(\nabla f)^t$ — градиент оптимизируемой функции в точке w^t ,

$$(\nabla f^2)^t = \left[\left(\frac{\partial f}{\partial w_1} \right)^2 \dots \left(\frac{\partial f}{\partial w_N} \right)^2 \right],$$

$$EMA_b(f)^t = (1 - b) * f^t + b * EMA_b(f)^{t-1} — скользящее среднее в точке f^t.$$

Для оценки качества предсказания в работе использовалась метрика f1-score, т. е. среднее гармоническое precision и recall с множителем 2:

$$f_1 = \frac{2 * precision * recall}{precision + recall}$$

$$precision(\text{точность}) = \frac{TP}{TP + FP}$$

TP — верноположительные срабатывания
 FP — ложноположительные срабатывания

$$recall(\text{полнота}) = \frac{TP}{TP + FN}$$

FN — ложноотрицательные срабатывания.

Наибольшее значение метрика достигает при максимальных precision и recall и равняется 1. F1-score близка к нулю, если один из аргументов близок к нулю.

2. Подбор размера батча. По результатам обучения модели с фиксированным числом слоев и узлов, но с различным размером батча в течение 1200 итераций регистрировалось значение функции потерь loss.

Размер батча — это количество тренировочных примеров, загружаемых в сеть для одноразового прогона. Как следует из рис. 1, при достаточно большом размере батча (40) влияния на качество предсказания нет. Если же $1 < \text{размер батча} < 40$, то присутствует незначительное проседание метрики. В данной работе размер батча = 256. Выбрав размер батча, определяем число эпох обучения, т. е. проходов сети по всем данным во время обучения. Следует сказать, что число эпох зависит от того, как быстро оптимизируются параметры, экспериментально было показано, что для данного оптимизатора и функции потерь достаточно порядка 30 эпох. Из вышесказанного следует, что среди большого набора гиперпараметров сети из общих соображений затруднительно определить именно количество элементов сети, т. е. число слоев и нейронов.

3. Оценка асимптотической сложности алгоритма. При создании алгоритма важной составляющей является время работы программы, т. к. мы можем создать точную предсказывающую модель, но с большим количеством вычислительных операций, что повлечет за собой значительные временные затраты в процессе обучения. Однако время работы программы не является универсальной мерой, т. к. оно зависит от нескольких факторов, помимо структуры алгоритма (языка программирования, вычислительных возможностей и т. п.). Поэтому введение термина асимптотическая сложность алгоритма

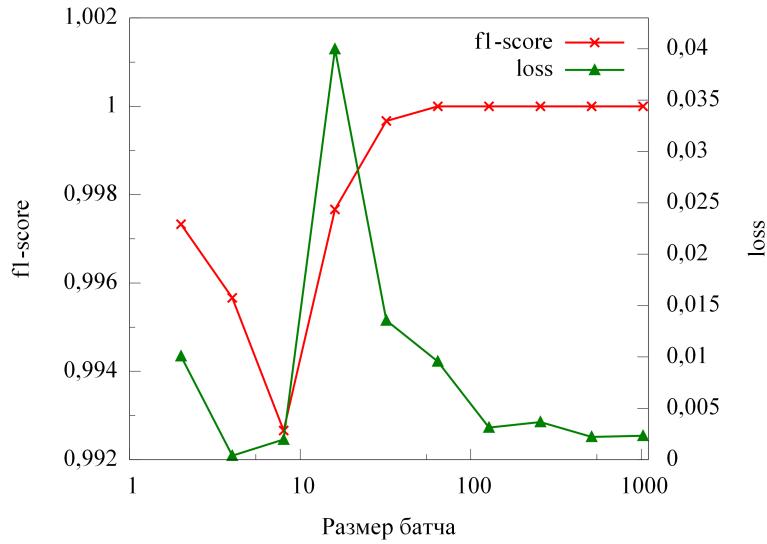


Рис. 1. Зависимость f1-score и loss от размера батча для обучения в течение 1200 итераций

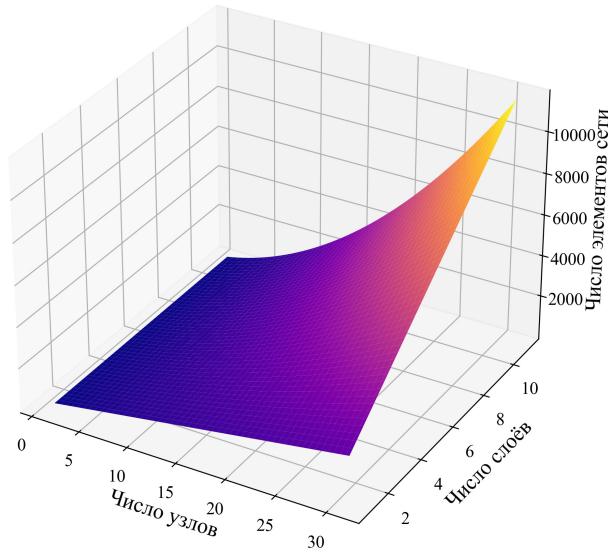


Рис. 2. Зависимость числа оптимизируемых весов от количества узлов и слоев

позволяет описать эффективность в терминах количества выполняемых операций, т. е. O -нотации. $O(n)$ означает, что число операций зависит линейно от количества данных n .

В работе [8, 10] была предложена оценка количества операций алгоритма полносвязной НС, а также вычислена асимптотическая сложность. Воспользуемся предложенным методом, учитывая различия используемых архитектур моделей.

Также в расчетах не будем рассматривать алгоритмы оптимизации, вычисления ошибки и обучения модели, т. к. исследуется поведение сети в зависимости от числа слоев и нейронов. Обозначим за S полное количество операций, совершающееся за прямое S_{fp} и обратное распространение S_{bp} . Тогда $S = S_{fp} + S_{bp}$.

Введем следующие обозначения: m — количество слоев (с учетом входного и выходного слоев), n_i — количество узлов в слое i , $n = \max n_i$, n_0 — количество входов сети, n_m — количество выходов сети, a, b и g — коэффициенты для уравновешивания веса вычислительных операций сложения (вычитания), умножения (деления), взятию экспоненты (натурального логарифма) соответственно.

Асимптотическая сложность алгоритма функции активации ReLU равняется $O(1)$ в силу того, что выбор максимального значения осуществляется за константное время. Число операций для вычисления функции активации

$$\text{LogSoftMax} = g(2 + n_i) + b + a(n_i - 1).$$

Вычисление взвешенной суммы входов одного узла первого слоя выполняется следующим образом:

$$\sum_{i=1}^{n_0} w_{1i}x_{1i} + q.$$

Отсюда, количество умножений n_0 , количество сложений равняется $(n_0 - 1)$. Для всего первого слоя имеем: $n_1 n_0(a + b)$. Для слоя i , ($i = 2, m - 1$) получим $S_i \leq n^2(a + b)$. Для слоя m : $S_m \leq nn_m(2a + 2b) + n_m(2g + b - a)$.

В итоге

$$S_{fp} \leq nn_0(a + b) + (m - 2)n^2(a + b) + nn_m(2a + b + g) + n_m(2g + b - a).$$

Аналогично оценивая количество операций для обратного распространения, получаем:

$$\begin{aligned} S_{bp} \leq nn_0(2a + 4b + n(a + b)) + n^2(2a + 4b + n_m(a + b)) + \\ + nn_m(4b + 3a) + (m - 3)n^2(2a + 4b + n(a + b)). \end{aligned}$$

Суммарное число операций:

$$\begin{aligned} S \leq nn_0(a + b) + (m - 2)n^2(a + b) + nn_m(2a + b + g) + \\ + n_m(2g + b - a) + nn_0(2a + 4b + n(a + b)) + n^2(2a + 4b + n_m(a + b)) + \\ + nn_m(4b + 3a) + (m - 3)n^2(2a + 4b + n(a + b)). \end{aligned}$$

Данная формула справедлива для $m \geq 3$. Выражение имеет более простой вид, если предположить, что $a \approx b \approx g$:

$$S \leq a[8nn_0 + (m - 2)2n^2 + 11nn_m + 2n_m + 6n^2 + 2n^2n_m + 2n^2n_0 + (m - 3)n^2(6 + 2n)].$$

Для асимптотической сложности имеем:

$$S \leq O(n^3) + O(n^2n_m) + O(n^2n_0) + O(n^2) + O(nn_0) + O(nn_m) + O(n_m).$$

Получаем, что количество операций прямого и обратного прохода растет пропорционально кубу от максимального количества нейронов в слое и линейно по числу слоев. Также число оптимизируемых весов сети линейно зависит от числа слоев и квадратично от числа нейронов, рис. 1. Из чего следует, что использование глубоких сетей более предпочтительное как с точки зрения сложности алгоритма, так и количества настраиваемых весов.

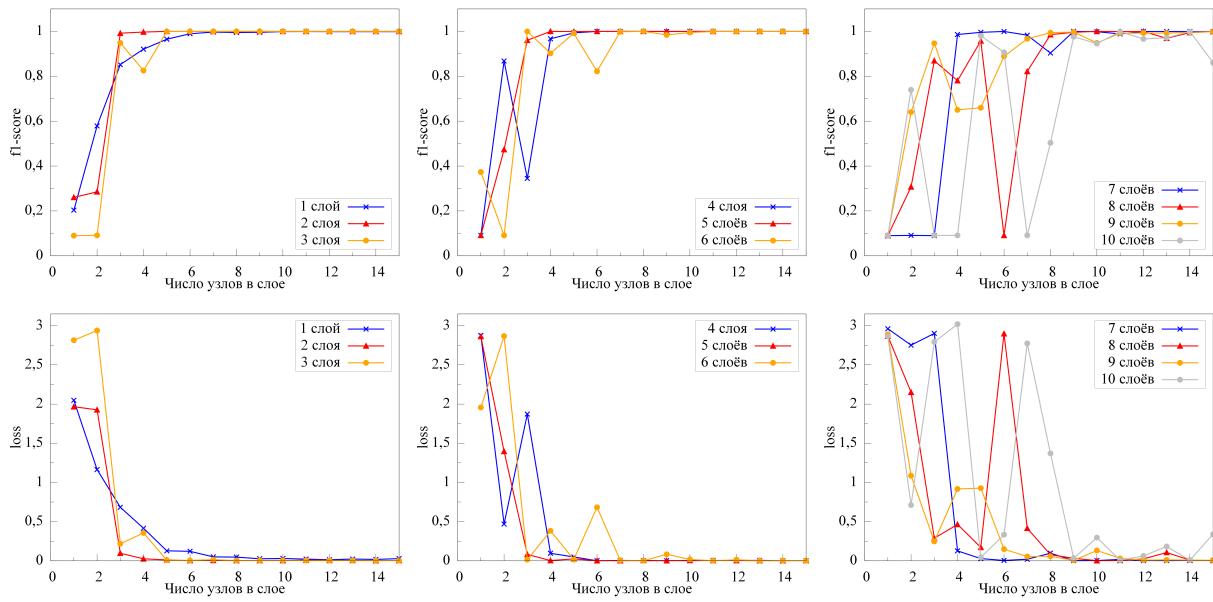


Рис. 3. Зависимость f1-score и loss от количества узлов в сети для тренировочного сета в 10000 элементов

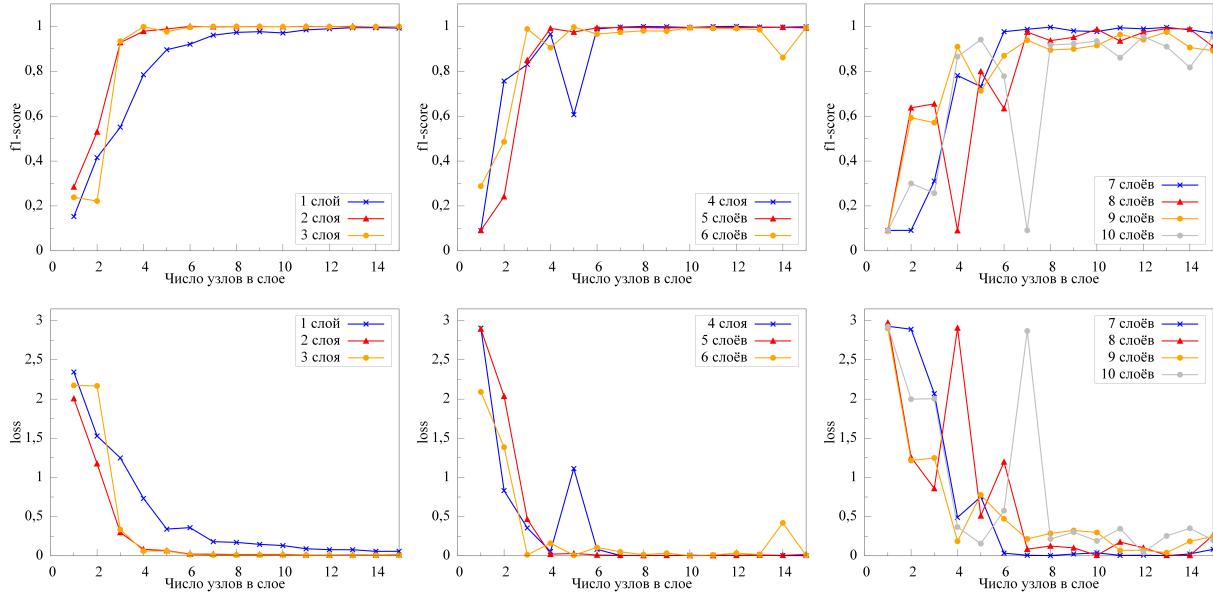


Рис. 4. Зависимость f1-score и loss от количества узлов в сети для тренировочного сета в 1000 элементов

4. Влияние количества нейронов на качество предсказания. В работе исследованы зависимость f1-score и значение функции потерь от числа узлов в слое, при обучении модели на тренировочной выборке в 10000 элементов и в 1000 элементов (рис. 3). Рис. 4).

Увеличение числа узлов в слое приводит к улучшению качества предсказаний модели, т.е. метрика f1-score растет. Выход на плато для всех случаев происходил после 10 узлов. В случае двухслойной модели выход на плато достигается при наименьшем количестве узлов в слое. Дальнейшее увеличение количества слоев сети не приводит к значительному увеличению качества предсказания. А в некоторых случаях (число слоев $m > 6$) приводит к худшим результатам. Наблюдается качественная зависимость между ве-

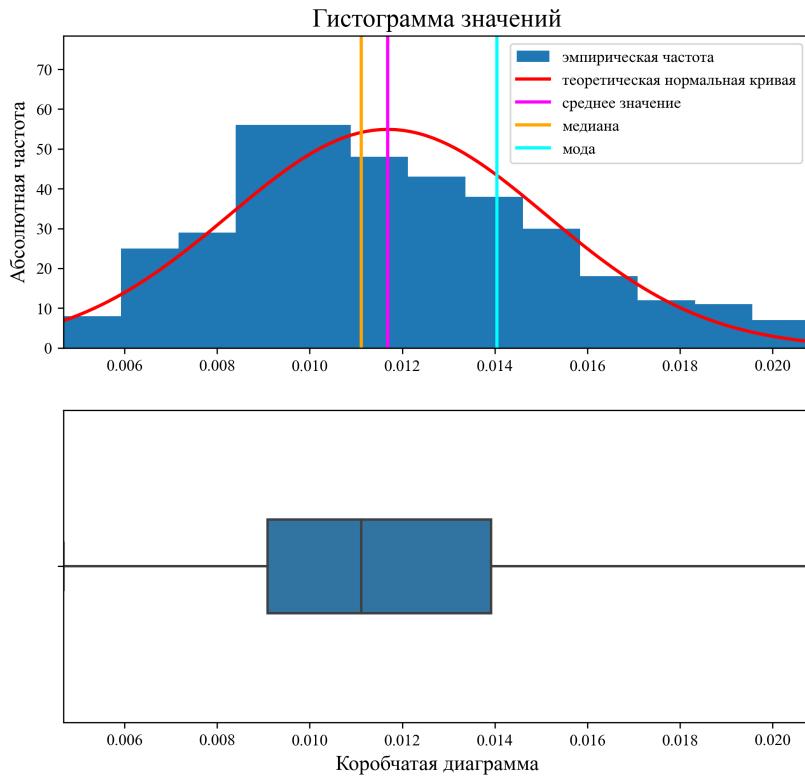


Рис. 5. Гистограмма значений loss и коробчатая диаграмма

личиной loss и f1-score. Высокий loss соответствует низкому значению f1-score. Увеличение числа слоев позволяет модели строить более сложные взаимосвязи, что может повысить качество предсказания. Однако с возрастанием количества слоев возрастает риск переобучения сети и могут возникать локальные минимумы функции ошибок, что приводит к проблемам при обучении. Поэтому с точки зрения метрики f1-score архитектуру сети нецелесообразно составлять из большого числа слоев для данной задачи классификации.

5. Влияние начальной инициализации. Предсказания некоторых алгоритмов, которые применяются для задачи классификации липотипов, сильно зависят от начальной инициализации параметров (например, случайный лес [11, 12]). В нашем случае была исследована зависимость значения loss от начальной инициализации весов НС. При создании модели веса инициализируются случайно из промежутка:

$$(-\sqrt{k}, \sqrt{k}), k = \frac{1}{(\text{num} \cdot \text{input features})}$$

По результатам 400 обучений модели с одним скрытым слоем были построены гистограмма и коробчатая диаграмма значений loss (рис. 5).

Проверка нормальности распределения проводилась с помощью критерия Эпса-Палли [13]. Критерий рекомендован к использованию в ГОСТ Р ИСО 5479-2002 [14]. В работе [15] было установлено, что по мощности критерий Эпса-Палли превосходит кри-

терии Шапиро-Уилка, Д'Агостино, Дэвида-Хартли-Пирсона. Тесты показали, что распределение значения loss не является распределением случайной величины. Данный факт следует учитывать при постановке требования воспроизводимости экспериментов и фиксировать начальную инициализацию весов модели. Также следует упомянуть, что в данной работе мерой качества предсказания сети является значение метрики f1-score, которое по результатам всех 400 предсказаний было близко к 1 и напрямую не связано с количественным значением loss.

Выводы. Согласно вышеизложенным рассуждениям, было показано, что для задачи классификации линотипов с помощью НС нецелесообразно использовать глубокие сети. Несмотря на тот факт, что с точки зрения количества выполняемых операций предпочтительнее увеличивать именно число слоев сети, большое число элементов однозначно не гарантирует рост метрики. Напротив, качество предсказания модели на тестовой выборке оставалось стабильно высоким при небольших слоях ($2 < m < 6$). А вот количество узлов в слое значительно влияет на результаты работы классификатора. При числе узлов в слое $n < 5$ почти все модели показали неудовлетворительные результаты, независимо от глубины сети. Однако увеличение числа нейронов однозначно приводит к выигрышу в качестве. Отсюда следует, что число узлов в слое является определяющим параметром, и в первую очередь необходимо настраивать именно его.

Список литературы

1. Thomas A., et al. Automated lithology extraction from core photographs // First Break. 2011. V. 29. N 6.
2. Baraboshkin E. E., et al. Deep convolutions for in-depth automated rock typing // Computers and Geosciences. 2020. V. 135. P. 104330.
3. Abashkin V. V., et al. Quantitative analysis of whole core photos for continental oilfield of Western Siberia // SPE Russian Petroleum Technology Conference, OnePetro, 2020.
4. Seleznev I. A., et al. Joint Usage of Whole Core Images Obtained in Different Frequency Ranges for the Tasks of Automatic Lithotype Description and Modeling of Rocks' Petrophysics Properties // Geomodel 2020, European Association of Geoscientists and Engineers, 2020. V. 2020. N 1. P. 1–5.
5. Амиргалиев Е. Н., и др. Интеграция алгоритмов распознавания литологических типов // Проблемы информатики. 2013. № 4 (21). С. 11–20.
6. Чанг Б. Т. Т., и др. Классификация изображений на основе применения цветовой информации, вейвлет-преобразования Хаара и многослойной нейронной сети // Проблемы информатики. 2011. № 5. С. 81–86.
7. Мухамедгалиев А. Ф., Разакова М. Г., Смирнов В. В. Создание и развитие геоинформационных технологий тематической интерпретации данных радиолокационного зондирования с использованием математических методов и вычислительных алгоритмов текстурной классификации и нейронных сетей // Проблемы информатики. 2012. № 3. С. 69–73.
8. Manurangsi, P., Reichman, D. The computational complexity of training ReLU (s). arXiv:1810.04207v2 [cs.CC]. 2018.
9. Kingma, D. P., Ba, J. Adam: A method for stochastic optimization. arXiv:1412.6980 [cs.LG]. 2014.
10. Максимушкин В. В., Арзамасцев А. А. Сравнительная оценка вычислительной сложности обучения искусственной нейронной сети с жестким ядром и сети с классической структурой // Вестник Тамбовского университета. Серия: Естественные и технические науки. 2006. Т. 11. № 2. С. 190–197.

11. Makienko D., Seleznev I., Safonov I. The effect of the imbalanced training dataset on the quality of classification of lithotypes via whole core photos // Creative Commons License Attribution. 2020. V. 4.
12. Bernard, S., Heutte, L., Adam, S. Influence of hyperparameters on random forest accuracy // International workshop on multiple classifier systems, Springer, Berlin, Heidelberg, 2009. P. 171–180.
13. Epps, T. W., Pulley, L. B. A test for normality based on the empirical characteristic function // Biometrika. 1983. V. 70. N 3. P. 723–726.
14. ГОСТ Р 5479-2002. Статистические методы. Проверка отклонения распределения вероятностей от нормального распределения / М.: Изд-во стандартов, 2002.
15. Лемешко Б. Ю. Критерии проверки отклонения распределения от нормального закона. Руководство по применению / Б. Ю. Лемешко. М.: ООО «Научно-издательский центр ИНФРА-М», 2015. 160 с.



Коссов Георгий Андреевич — студент магистратуры Московского Физико-Технического Института. Окончил бакалавриат МФТИ в 2021 году по специальности «Прикладная математика и физика». Область научных интересов включает в себя компьютерное зрение, нейронные сети и механику флюидов. E-mail: kossov.ga@phystech.edu.

Kossov Georgy Andreevich is MIPT master's student (Moscow Institute of Physics and Technology). He received the bachelor's degree in 2021. Currently — student-intern SLB, Moscow Research Center. His current research interest includes computer vision, neural networks and fluid mechanics.

Селезнев Илья Александрович окончил МИРЭА, факультет кибернетики. Ра-

бота: ВНИГеосистем, ПетроАльянс, ТК Schlumberger, в настоящее время — старший научный сотрудник SLB, Московский Исследовательский Центр. Область научных интересов включает в себя разработку ПО, аналитику данных, обработку изображений, машинное обучение. E-mail: ISeleznev@slb.com, телефон: +79685114902.



Seleznov Ilya Aleksandrovich graduated from MIREA, Faculty of Cybernetics. Work: Geosystem institute, PetroAliance, TC Schlumberger, currently — Senior research scientist SLB, Moscow Research Center. His current research interest includes software development, data analytics, image processing, ML. Postal address: Vasilisy Kozhinoy st, 14-3-283, Moscow, 121096. E-mail: ISeleznev@slb.com, phone number: +79685114902.

Дата поступления — 07.11.2022