

СОВРЕМЕННЫЕ ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 519.24

**ПРИМЕР РЕАЛИЗАЦИИ ВЫБОРА
МИНИМАЛЬНОГО ЧИСЛА В БИНАРНОМ ВИДЕ
НА НЕЙРОННЫХ СЕТЯХ***

А.А. ВОЕВОДА¹, Д.О. РОМАННИКОВ²

¹ 630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет, доктор технических наук, профессор кафедры автоматики. E-mail: uscit@uscit.ru

² 630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет, кандидат технических наук, доцент кафедры автоматики. E-mail: dmitry.romannikov@gmail.com

В статье сравниваются два способа построения нейронных сетей. Первый способ является классическим способом построения нейронной сети в виде многослойного перцептрона. Второй основывается на предварительно рассчитанной структуре нейронной сети. В статье рассматривается пример нейронной сети, которая выбирает максимальное число из трех чисел, представленных в бинарном виде двумя разрядами. При первом варианте построения нейронной сети она будет представлена как виде последовательно соединенных слоев, где во входном слое содержится 6 нейронов, во втором – 5 нейронов, в третьем – 4 и в последнем (выходном) – 2 нейрона. Общее количество обучаемых параметров в такой сети – 69. При построении сети исходя из предварительно рассчитанного варианта сеть будет иметь более сложную структуру, но общее количество обучаемых параметров уменьшается до 44. В статье сравниваются скорости обучения двух полученных нейронных сетей. При использовании предварительно рассчитанных коэффициентов скорость обучения нейронной сети возрастает примерно в два раза., но при использовании одинаковых начальных условий, полученных путем выбора случайных значений из диапазона от 0 до 1 с равномерным распределением, скорость обучения многослойной нейронной сети выше, чем у предварительно рассчитанного варианта.

Ключевые слова: нейронные сети, сети Петри, искусственный интеллект, преобразование, функция активации, keras, регуляризация, обучение

DOI: 10.17212/2307-6879-2017-1-64-71

* Статья получена 30 ноября 2016 г.

ВВЕДЕНИЕ И ПОСТАНОВКА ЗАДАЧИ

В теории нейронных сетей существует идея уменьшения количества параметров обучения (в понятие параметров в статье включены все веса, смещения (*bias*) и параметры нелинейности (например, для сигмоида – это сжатие по осям x и y)) в нейронной сети за счет использования схемы с предварительно рассчитанной структурой.

В общем виде данная задача относится к задаче многоклассовой классификации (*multilabel classification*) [1–8] и может быть решена с помощью многослойного персептрона или других многослойных нейронных сетей. Однако при таком подходе число параметров может быть избыточно, что усложняет обучение (из-за того, что методы обучения основаны на градиентном спуске [1–6]) и потенциально приводит к тому, что нейронная сеть сходится не к наилучшему решению. Для предварительно рассчитанной модели нейронной сети число параметров должно значительно уменьшиться. Для иллюстрации данного утверждения в данной статье рассматривается пример построения нейронной сети, с помощью которой выбирается наибольшее число из трех чисел, представленных в бинарном виде тремя разрядами.

1. РЕШЕНИЕ

Первоначально необходимо получить сеть Петри для решения поставленной задачи. Так как сети Петри являются полными по Тьюрингу [2] и из-за того, что можно хранить значения переменных, то в них есть возможность преобразовать поразрядное представление числа в целое путем умножения разрядов, а затем просто сравнить три числа. Но из-за того, что такой подход неприменим к нейронным сетям, необходимо придерживаться поразрядного представления числа в сетях Петри.

В данной работе сеть Петри и нейронная сеть совпадают полностью, и из-за этого кажется целесообразным привести только представление нейронной сети. На рис. 1 представлена структура получившейся нейронной сети, в которой для компактности реализован выбор между тремя числами, представленными двумя разрядами. При этом увеличение количества разрядов не вызывает алгоритмической сложности. На рис. 2 $a_1, a_2; b_1, b_2; c_1, c_2$ представляют собой три числа, в которых меньший номер соответствует старшему разряду. Нейроны обозначены как $n_1 – n_{10}$, а выходной слой как y_1, y_2 . В схеме содержится 44 параметра (33 параметра «веса» и 11 – смещений). В качестве нелинейности нейронов используется сигмоидальная функция. Значения весов для нейронной сети на рис. 1 приведены в таблице. На рис. 1 n_2 – составной нейрон. Он представляет два нейрона в первом слое и один нейрон во втором.

Для сравнения рассмотрим классически построенную схему нейронной сети. Исходный код схемы приведен на рис. 2. В данной сети четыре слоя, во входном слое шесть нейронов, в первом скрытом пять нейронов, во втором – четыре и в выходном слое – два.

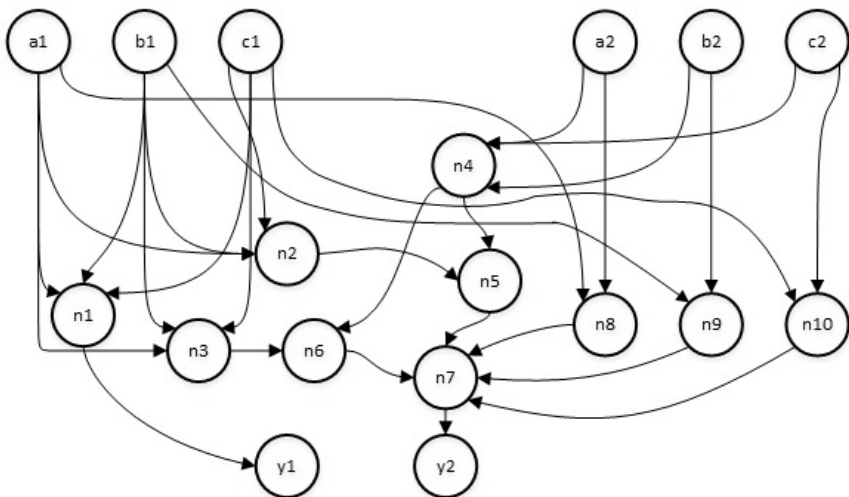


Рис. 1. Нейронная сеть для выбора максимального числа из трех представленных двумя разрядами

Значение весов для нейронной сети

Название нейрона	Значение весов (веса указаны для связей слева направо)	Значение смещения
n1	7; 7; 7	-3
n2	3.7; -3.6; 3.36; -2.5; -4.97; -4.71; -3.32; 4.91	-2.48; 1.08 -1.5
n3	4; 4; 4	-10
n4	7; 7; 7	-3
n5	7; 7	-3
n6	7; 7	-3
n7	7; 7; 7; 7; 7	-3
n8	5; 5	-6
n9	5; 5	-6
n10	5; 5	-6

```
ret_model.add(Dense(5, input_shape=(6,), activation='sigmoid'))
ret_model.add(Dense(4, input_shape=(5,), activation='sigmoid'))
ret_model.add(Dense(2, activation='sigmoid'))
ada_grad = Adagrad(lr=0.1, epsilon=1e-08, decay=0.0)
ret_model.compile(optimizer=ada_grad, loss='binary_crossentropy')
```

Рис. 2. Исходный код для реализации нейронной сети

Всего в нейронной сети (на рис. 2) содержится 69 обучаемых параметров, что в полтора раза больше нейронной сети, построенной по сети Петри. Графики обучения (100 эпох) обеих нейронных сетей приведены на рис. 3 и 4.

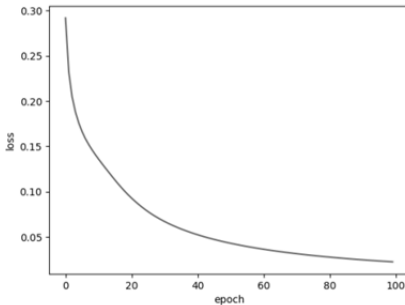


Рис. 3. График стоимостной функции обучения нейронной сети на рис. 1

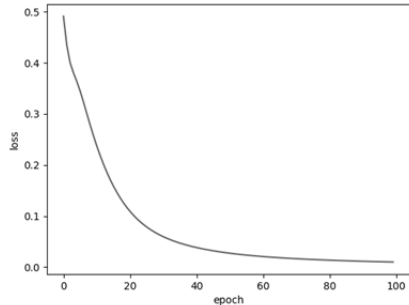


Рис. 4. График стоимостной функции обучения нейронной сети на рис. 2

Из графиков на рис. 3 и 4 видно, что нейронная сеть с предварительно рассчитанными параметрами обучается быстрее и за 100 эпох сошлась к меньшей ошибке, чем нейронная сеть на рис. 2. И такой результат является ожидаемым. Поэтому более интересным является эксперимент, в котором нейронная сеть с предварительно рассчитанной структурой обучается с «нулевых» (под «нулевыми» начальными условиями здесь и далее понимается равномерно распределенные значения из диапазона от нуля до единицы) начальных условий. Результат такого обучения представлен на рис. 5.

По графику можно сделать вывод, что скорость обучения нейронной сети на рис. 2 выше, чем скорость нейронной сети на рис. 1. В частности, по графику видно, что на рис. 4 после 40 эпох значение стоимостной функции меньше, чем стоимостной функции на рис. 5.

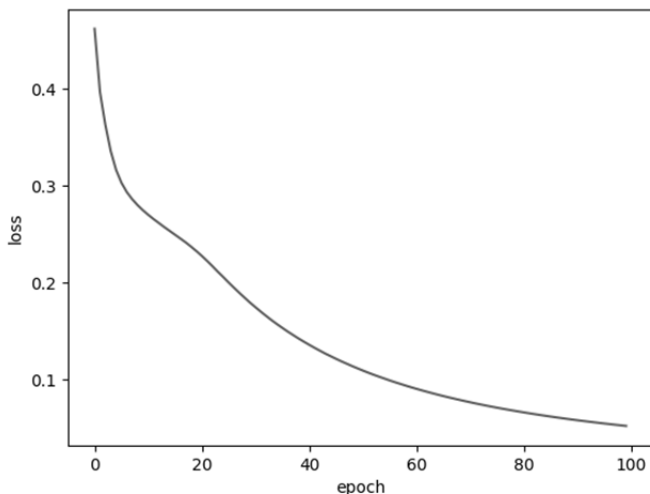


Рис. 5. График стоимостной функции обучения нейронной сети на рис. 1 с «нулевыми» начальными условиями

Также стоит отметить интересную особенность, связанную с реализацией нейрона $n2$ на рис. 1. Данный нейрон реализует функцию исключающего «или» с инверсией (т. е. единица на выходе данного нейрона появляется только в случае всех нулей или всех единиц на входе). Данная функция может быть реализована на нейронной сети, имеющей структуру в три входных нейрона, три скрытых и один выходной. Однако в составе большей нейронной сети на рис. 1 нейрон $n2$ имеет структуру из трех входных нейронов и одного выходного. Для объяснения данного факта необходимо дальнейшее исследование, но предположительно это может быть объяснено тем, что часть «информации» при реализации данной функции передается через другие нейроны сети.

ЗАКЛЮЧЕНИЕ

В статье сравниваются два подхода к построению нейронной сети. Первый подход является классическим – нейронная сеть представляет собой несколько слоев. Второй способ заключается в предрасчете структуры нейронной сети и дальнейшем ее обучении. В статье приведены графики стоимостных функций. Предварительно рассчитанная нейронная сеть содержит меньшее количество нейронов. Для рассматриваемого примера выбора максималь-

ного числа среди трех чисел, представленных в бинарном виде двумя разрядами, число обучаемых параметров для предварительно рассчитанного варианта составило 44 против 69 для классического представления. Сравнивая скорость обучения для многослойного представления нейронной сети с представленным в статье предварительно рассчитанным вариантом, можно сказать, что при использовании предварительно рассчитанных коэффициентов нейронная сеть обучается гораздо быстрее, но при случайном равномерном выборе значений начальных параметров обучение в классическом представлении происходит быстрее.

СПИСОК ЛИТЕРАТУРЫ

1. *Bishop C.* Pattern recognition and machine learning. – New York: Springer, 2007. – 738 p. – (Information science and statistics).
2. *Richert W., Coelho L.* Building machine learning systems with Python. – Birmingham: Packt Publ., 2013. – 290 p.
3. *Hastie T., Tibshirani R., Friedman J.* The elements of statistical learning: data mining, inference, and prediction. – 2nd ed. – New York: Springer, 2013. – 745 p. – (Springer series in statistics).
4. *Lantz B.* Machine learning with R. – Birmingham: Packt Publ., 2013. – 396 p.
5. *Mohri M., Rostamizadeh A., Talwalkar A.* Foundations of machine learning. – Cambridge, MA: MIT Press, 2012. – 432 p. – (Adaptive computation and machine learning series).
6. *Conway D., White J.M.* Machine learning for hackers. – Sebastopol, CA: O'Reilly, 2012. – 324 p.
7. Welcome to the Deep Learning tutorial [Electronic resource]. – URL: <http://deeplearning.stanford.edu/tutorial/> (accessed: 31.05.2017).
8. *Haykin S.* Neural networks: a comprehensive foundation. – New York: MacMillan Publ., 1994. – 1104 p.

Воевода Александр Александрович, доктор технических наук, профессор кафедры автоматике Новосибирского государственного технического университета. Основное направление научных исследований – управление многоканальными объектами. Имеет более 200 публикаций. E-mail: ucit@ucit.ru

Романников Дмитрий Олегович, кандидат технических наук, доцент кафедры автоматике Новосибирского государственного технического универси-

тета. Основное направление научных исследований – нейронные сети, сети Петри. Имеет более 50 публикаций. E-mail: dmitry.romannikov@gmail.com

An example of realization of the choice of the minimum number in a binary form on neural networks*

А.А. Воевода¹, Д.О. Романиков²

¹ *Novosibirsk State Technical University, 20 Karl Marks Avenue, Novosibirsk, 630073, Russian Federation, doctor of Technical Sciences, professor of the automation department. E-mail: ucit@ucit.ru*

² *Novosibirsk State Technical University, 20 Karl Marks Avenue, Novosibirsk, 630073, Russian Federation, candidate of Technical Sciences, associate professor of the automation department. E-mail: dmitry.romannikov@gmail.com*

The article compares two methods of constructing neural networks. The first method is a classical way of constructing a neural network in the form of a multilayer perceptron. The second is based on the pre-calculated structure of the neural network. The article considers an example of a neural network that selects the maximum number of three numbers represented in binary form in two digits. In the first variant of constructing a neural network, it will be represented as a series of connected layers, where 6 neurons are contained in the input layer, in the second 5 neurons, in the third 4 and in the last (output) - 2 neurons. The total number of learning parameters in such a network is 69. When building a network based on the pre-calculated version, the network will have a more complex structure, but the total number of learning parameters decreases to 44. The article compares the learning rates of the two neural networks obtained. Using the pre-calculated coefficients, the learning speed of the neural network increases approximately by two times, but when using the same initial conditions obtained by selecting random values from the range from 0 to 1 with a uniform distribution, the learning rate of the multilayer neural network is higher than in the pre-calculated version .

Keywords: neural networks, Petri nets, artificial intelligence, transformation, activation function, keras, regularization, training

DOI: 10.17212/2307-6879-2017-1-64-71

REFERENCES

1. Bishop C. *Pattern recognition and machine learning. Information science and statistics*. New York, Springer, 2007. 738 p.
2. Richert W., Coelho L. *Building machine learning systems with Python*. Birmingham, Packt Publ., 2013. 290 p.

* Received 30 November 2016.

3. Hastie T., Tibshirani R., Friedman J. *The elements of statistical learning: data mining, inference, and prediction*. 2nd ed. Springer series in statistics. New York, Springer, 2013. 745 p.
4. Lantz B. *Machine learning with R*. Birmingham, Packt Publ., 2013. 396 p.
5. Mohri M., Rostamizadeh A., Talwalkar A. *Foundations of machine learning. Adaptive computation and machine learning series*. Cambridge, MA, MIT Press, 2012. 432 p.
6. Conway D., White J.M. *Machine learning for hackers*. Sebastopol, CA, O'Reilly, 2012. 324 p.
7. *Welcome to the Deep Learning tutorial*. Available at: <http://deep-learning.stanford.edu/tutorial/> (accessed 31.05.2017).
8. Haykin S. *Neural networks: a comprehensive foundation*. New York, MacMillan Publ., 1994. 1104 p.