

СООБЩЕНИЯ

УДК 519.24

ТЕЗИСЫ О НЕЙРОННЫХ СЕТЯХ*

Д.О. РОМАННИКОВ**

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

При разработке моделей, основанных на искусственных нейронных сетях, выбор модели нейронной сети основывается на эмпирических знаниях исследователя, что крайне негативно сказывается на пороге вхождения, обоснованности и корректности работы модели, времени обучения и других аспектах. Таким образом, задача обоснования выбора модели нейронной сети является актуальной задачей. В статье рассматривается подход к формированию структуры нейронной сети на основе предварительно рассчитанной структуры сети Петри. Приведены реализации данного подхода на примерах построения нейронных сетей для логических функций «и», «или», «исключающего или», выбора максимального числа и другие. По результатам экспериментов сделаны выводы о неоднозначности обучения в нейронных сетях, а именно, что при различных начальных условиях могут получиться различные веса параметров. Также в статье исследуется влияние крутизны сигмоидальной функции на обучение нейронных сетей. По результатам экспериментов сделаны выводы о том, что незначительное увеличение крутизны (в пределах от 1 до 20 в степени) дает ускорение скорости обучения. Дальнейшее увеличение крутизны сигмоидальной функции приводит к деградации скорости обучения, а при дальнейшем увеличении – практически к полному останову. Статья заканчивается примером построения нейронной сети для задачи выбора максимального числа среди трех чисел, представленных в бинарном виде тремя разрядами. Эксперимент показал, что полученная структура сети позволяет выполнить обучение примерно в три раза быстрее без потери точности. Такое различие в скорости обучения может быть объяснено меньшим количеством параметров в нейронной сети: 102 вместо 127.

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

DOI: 10.17212/2307-6879-2017-1-98-108

* Статья получена 18 октября 2016 г.

** В Сборнике научных трудов НГТУ № 4 (86) за 2016 год на с. 102 неправильно дан английский вариант названия статьи Романикова Д.О. Правильное название: On the transformation of Petri nets in neural network.

ВВЕДЕНИЕ

Понимание работы любого инструмента основывается на осознании принципов его действия и практических навыков его использования. В частности, понимание работы нейронных сетей может быть достигнуто за счет моделирования различных примеров и анализа результатов.

В данной статье рассматривается задача синтеза структуры нейронной сети [1–8] для известных алгоритмов. В частности, рассматриваются нейронные сети для реализации арифметических и логических операций, таких как операция «и», операция «или», операция «исключающее или» (*XOR*) [1–5], нахождения максимального элемента. При решении данных задач анализируется влияние начальных условий [9] на скорость обучения, крутизна сигмоидальной нелинейностей и другие параметры, оказывающие влияние.

1. ПОСТАНОВКА И РЕШЕНИЕ ЗАДАЧИ

В статье рассматриваются несколько задач: операция «и», операция «или», операция «исключающее или» (*XOR*), нахождения максимального элемента. При решении вышеприведенных задач ограничимся входными значениями «0» и «1» и фиксированной сигмоидальной нелинейностью. Рассмотрим решение данных задач подробнее.

Для решения задачи операции «и» с использованием сетей Петри была построена сеть, изображенная на рис. 1, с двумя входными местами и третьим местом, в которое через переход поступает метка выполняемой операции (в данном случае «и»). В нейронных сетях данная операция будет иметь вид, представленный на рис. 2. Данная реализация сходна со структурной схемы в терминах сети Петри, а также реализует операции «или», выбор максимального значения. Нейронная сеть на рис. 2 характеризуется несколькими параметрами: значениями w_1 , w_2 , b . Для задачи выбора максимального числа можно подобрать параметры $w_1 = 4$, $w_2 = 4$, $b = -5$ (в данном случае есть неоднозначность: например, параметры $w_1 = 6$, $w_2 = 6$, $b = -8$ также позволяют получить решение).

Для решения задачи *XOR* в терминах сетей Петри можно придерживаться такой же логики, что и для предыдущих задач (рис. 1), но полученная нейронная сеть будет отличаться. Структурная схема нейронной сети для *XOR* задачи приведена в [1–3] и на рис. 3. На рисунке нейроны обозначены кругами n_1 – n_3 . Данные нейроны соответствуют нейрону, выделенному пунктирной линией на рис. 2.

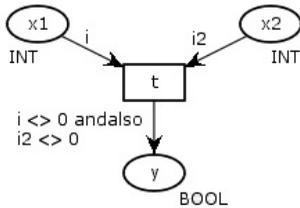


Рис. 1. Сеть Петри, реализующая логическое «и»

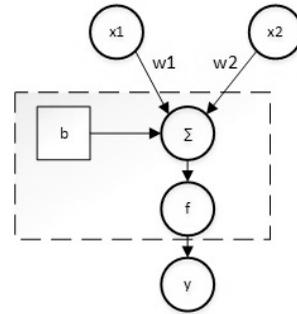


Рис. 2. Нейронная сеть, полученная по построенной сети Петри на рис. 1

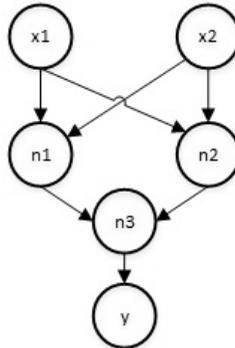


Рис. 3. Нейронная сеть для решения XOR задачи для двух входных переменных

Далее проанализируем обучение приведенных систем с целью исследования однозначности решения, влияния крутизны нелинейности и других параметров.

2. ТЕЗИС О НАЧАЛЬНЫХ УСЛОВИЯХ

При обучении данных нейронных сетей с начальными условиями, полученными выбором случайных значений из диапазона $[0, 1]$, нейронная сеть не всегда обучается (эксперимент выполнялся при 10 000 эпохах и коэффициенте

обучения 0.01), что напрямую связано с выбором начальных условий. Из этого следует *тезис о неоднозначности обучения в нейронных сетях*, что частично подтверждено в [9]. Например, при различных начальных условиях можно для операции «и» получить коэффициенты: $w_1 = 5.271$, $w_2 = 5.146$, $b_1 = -7.845$ или $w_1 = 11.462$, $w_2 = -4.774$, $b_1 = -2.326$ или $w_1 = 7.408$, $w_2 = 5.967$, $b_1 = -10.396$ и т. д. Показательно, что при задании предварительно рассчитанных начальных условий 1) нейронная сеть сходится; 2) обучение проходит за меньшее число итераций (для рассматриваемых операций сходимость достигается сразу).

Исследуем множество решений для нейронной сети, выполняющей логическую операцию «и». Для этого полученную на предыдущем этапе структуру нейронной сети проверим на целых значениях коэффициентов в диапазоне $[-10, 10]$. Результат проверки приведен на рис. 4.

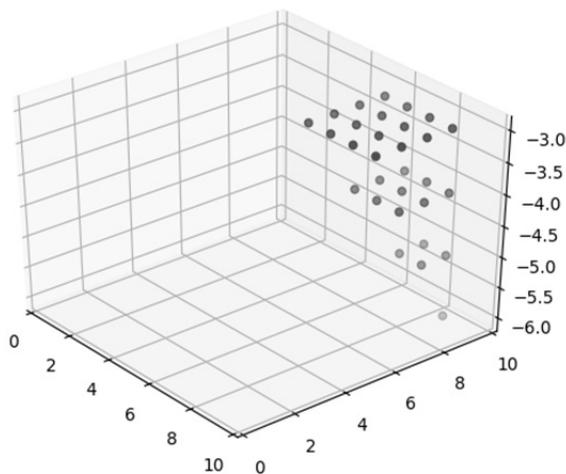


Рис. 4. Точки образуют пространство значений коэффициентов, при котором нейронная сеть выполняет логическую операцию «и»

Осями x , y , z на рис. 4 являются коэффициенты w_1 , w_2 , b соответственно. Критерием наличия решения в данном случае являлась проверка отклонения значения выхода нейронной сети не более чем на 10 %, т. е. на 0.1 в абсолютных значениях при масштабировании выхода нейронной сети к диапазону $[0, 1]$. Из приведенного рисунка видно, что нейронная сеть имеет решение при множестве вариантов значений ее параметров.

3. ТЕЗИС О ВЛИЯНИИ НЕЛИНЕЙНОЙ ФУНКЦИИ

В предыдущих примерах нелинейность была фиксированной $f = 1/(1 + e^{(-x)})$, хотя при анализе известных нейронных схем сочетание нескольких нейронов приводит к тому, что из-за изменения параметров фактически происходит смещение и растяжение f по осям x и y (их тоже можно отнести к параметрам).

Также в статье рассматривается эксперимент по обучению нейронной сети с двумя видами нелинейных функций нейронной сети: сигмоид (*sigmoid*) и сигмоида, приближенного к ступенчатой функции: $1/(1 + e^{(-k*x)})$, где $k \in [1, 50]$. Графики сходимости стоимостных функций при использовании сигмоидальных функций с коэффициентами $\{1, 10, 20, 30, 40\}$ представлены на рис. 4. График с коэффициентом 1 располагается посередине, далее при росте коэффициента сначала наблюдается скачкообразное ускорение обучения (при коэффициентах 10 и 20. На рис. 5 представлены нижними двумя графиками), что связано с большим наклоном сигмоидальной функции, но при увеличении значения до 30 наблюдается деградация скорости обучения, а при 40 – практически полный ее останов.

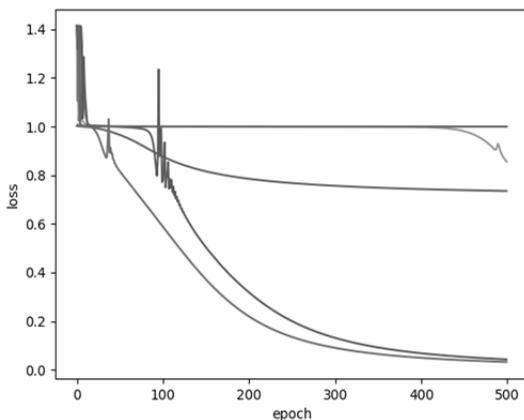


Рис. 5. Графики стоимостной функции в зависимости от значения параметра k в сигмоидальной нелинейности

Таким образом, из вышеприведенного следует **тезис** не просто о том, что скорость обучения зависит от выбора вида нелинейности, но и **о том, что изменение угла наклона нелинейности приводит к изменению скорости обучения**, из чего следует потенциальная возможность обучения нейронных

сетей при помощи изменения угла наклона нелинейных функций, а также ее смещения.

4. ПРИМЕР

Рассмотрим пример еще одной операции – выбора максимального числа из массива чисел. Для компактности будем рассматривать задачу как поиск максимального числа из трех целых чисел, представленных в бинарном виде тремя разрядами. При решении данной задачи в терминах сетей Петри будет получена представленная на рис. 6 структура, на которой представлено два числа. Первое число состоит из разрядов x_1 – x_3 (x_1 – старший), второе – из разрядов x_4 – x_6 (x_4 – старший).

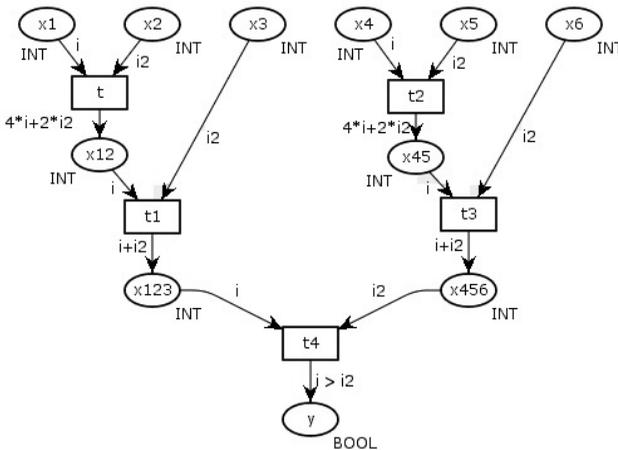


Рис. 6. Сеть Петри, реализующая задачу выбора максимального числа из двух чисел, представленных тремя разрядами (для более компактного представления приведены два числа)

Нейронная сеть, построенная по приведенной на рис. 6 структуре, изображена на рис. 7¹. Сеть была обучена на 10 000 примерах за 4000 эпох. Точность обучения на обучающих данных составила 98,72 %. График стоимостной функции представлен на рис. 8.

¹ Изображение автоматически сгенерировано программным пакетом keras 2.0.2. На рис. 7 нейроны представлены в виде прямоугольников и подписаны как Dense. Также на рисунке приведены подписи входных и выходных размерностей.

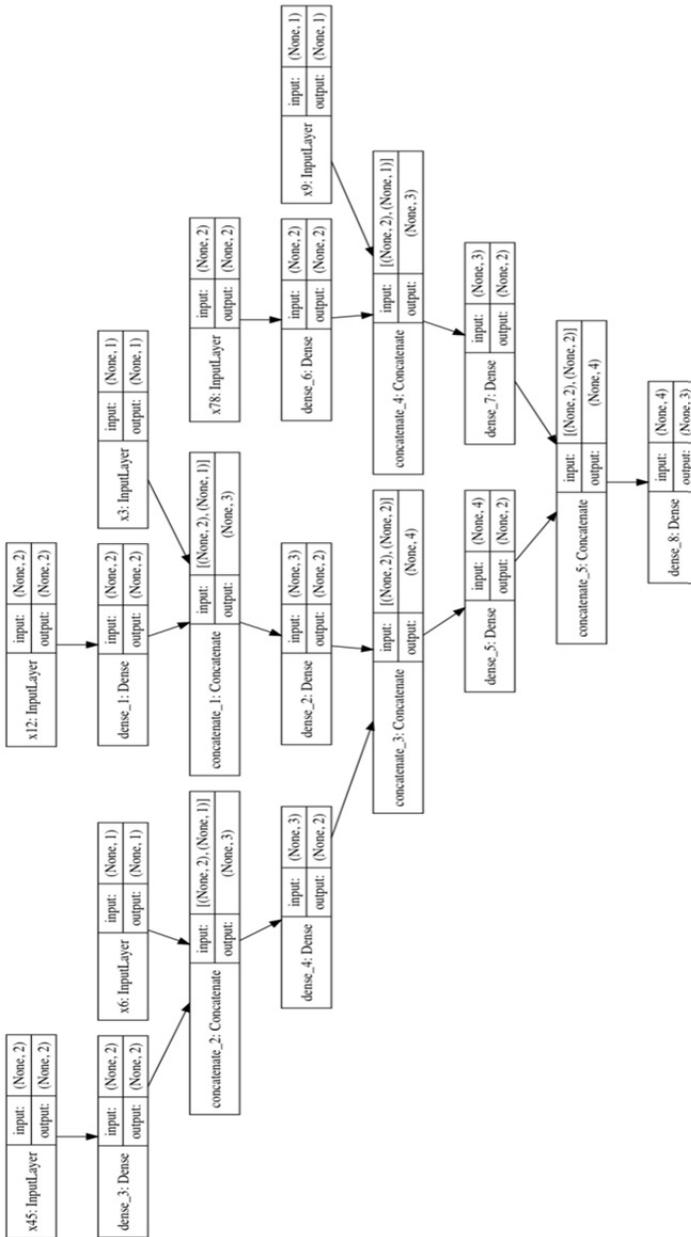


Рис. 7. Нейронная сеть для задачи выбора максимального числа из двух чисел, представленных тремя разрядами (соответствующая сеть Петри изображена на рис. 6)

Для сравнения приведен график стоимостной функции для нейронной сети, построенной классическим способом (рис. 9), а именно в слоях размерностью в $9 \times 7 \times 5 \times 3$. По графикам видно, что скорость обучения нейронной сети, построенной по структуре сети Петри и приведенной на рис. 6, выше, и при этом сходятся они к одному значению. Например, после 1500 экспериментов сети имели значения стоимостных функций 1.5 и 0.6, что может быть объяснено меньшим числом параметров в сетях (число параметров в многослойном персептроне 128 (а при выборе персептрона с равным количеством нейронов в слое 9 общее число параметров будет 300), в нейронной сети на рис. 7 – 107).

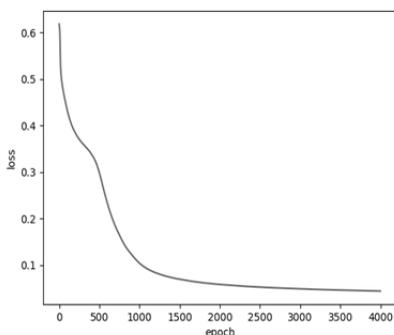


Рис. 8. График стоимостной функции обучения нейронной сети на рис. 6

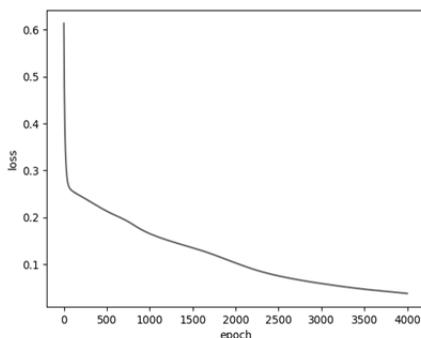


Рис. 9. График стоимостной функции обучения многослойного персептрона

Необходимо отметить, что уменьшение числа параметров было сравнительно небольшим: со 127 до 102, но это привело к приросту скорости обучения. Однако при значительном уменьшении числе параметров в сети обучение не будет достигать той точности, что и сети с избыточными параметрами. Также вероятно, что уменьшение параметров – это не единственный фактор, влияющий на скорость обучения, что является дальнейшим аспектом изучения.

ЗАКЛЮЧЕНИЕ

В статье исследуется принцип построения нейронной сети на основе заранее построенной сети Петри. Приведены примеры построения нейронных сетей для функций «и», «или», «исключающее или» и других предварительно рассчитанных сетей Петри. На вышеприведенных примерах были исследованы некоторые свойства нейронных сетей: показано, что выбор начальных

условий влияет на сходимость нейронных сетей (при «неудачном» выборе начальных условий нейронная сеть для XOR задачи не сходится). Также наблюдалось, что при различных начальных условиях нейронная сеть сходится к различным коэффициентам, что говорит о ее неоднозначности.

Кроме того, исследовался вопрос о влиянии крутизны нелинейности при обучении нейронных сетей. По результатам экспериментов было выявлено, что при незначительном увеличении угла наклона сначала наблюдается увеличение скорости обучения, но при достижении сигмоидальной функции некоторой крутизны (когда сигмоидальная функция приближается к ступенчатой) наблюдается значительная деградация скорости обучения.

В статье приведен пример нейронной сети для выбора максимального числа из чисел, представленных тремя разрядами в бинарном виде. По результатам экспериментов видно, что нейронная сеть, построенная по структуре сети Петри, обучается быстрее, чем обычная нейронная сеть, что может быть объяснено меньшим числом параметров: 127 против 102.

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.
9. *Романников Д.О.* О преобразовании сети Петри в нейронную сеть // Сборник научных трудов НГТУ. – 2016. – № 4 (86). – С. 98–103.

Романников Дмитрий Олегович, кандидат технических наук, доцент кафедры автоматики Новосибирского государственного технического университета. Основное направление научных исследований – нейронные сети, сети Петри. Имеет более 50 публикаций. E-mail: dmitry.romannikov@gmail.com

The formation of the correlated noises*

D.O. Romannikov

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

In the development of models based on artificial neural networks, the choice of the neural network model is based on the empirical knowledge of the researcher, which negatively affects the threshold of entry, validity and correctness of the model's work, training time and other aspects. Thus, the task of justifying the choice of a neural network model is an urgent task. The article considers the approach to the formation of the structure of the neural network based on the pre-calculated structure of the Petri net. The implementation of this approach is presented using examples of constructing neural networks for logical functions "and", "or", "excluding or", selecting the maximum number, and others. Based on the results of the experiments, conclusions are drawn about the ambiguity of learning in neural networks, namely, that under different initial conditions different weights of the parameters can turn out. The article also investigates the influence of the steepness of the sigmoid function on the training of neural networks. Based on the results of the experiments, conclusions are drawn that a slight increase in the steepness (within the range of 1 to 20 degrees) gives an acceleration in the speed of training. A further increase in the steepness of the sigmoid function leads to a degradation in the speed of training, and with further increase - to almost complete stop. The article ends with an example of constructing a neural network for the problem of selecting the maximum number among three numbers represented in binary form in three digits. The experiment showed that the obtained network structure makes it possible to perform training about 3 times faster without loss of accuracy. This difference in learning speed can be explained by a smaller number of parameters in the neural network: 102 instead of 127.

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

DOI: 10.17212/2307-6879-2017-1-98-108

REFERENCES

1. Bishop C. Pattern recognition and machine learning. Information science and statistics. New York, Springer, 2007. 738 p.

* Received 18 October 2016.

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. 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, Mac-Millan Publ., 1994. 1104 p.
9. Romannikov D.O. O preobrazovanii seti Petri v neironnyuyu set' [On the transformation of Petri nets in neural network]. Sbornik nauchnykh trudov Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Transaction of scientific papers of the Novosibirsk state technical university, 2016, no. 4 (86), pp. 98–103.