

ОБУЧЕНИЕ НЕЙРОННОЙ СЕТИ С ПРЕДВАРИТЕЛЬНО РАССЧИТАННОЙ СТРУКТУРОЙ*

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

При разработке моделей, основанных на искусственных нейронных сетях, выбор строится на эмпирических знаниях исследователя, что крайне негативно сказывается на пороге вхождения в технологию, а также на обоснованности и корректности работы модели, времени обучения и других ее аспектах. В статье рассматривается подход к формированию структуры нейронной сети на основе предварительного формирования структурной схемы системы и расчета или подбора ее коэффициентов. Целью является демонстрация того, что синтез структуры нейронной сети может быть выполнен не только интуитивным подбором слоев и связей. Для компактности предлагаемый подход показан на примере выбора максимального значения из двух чисел (используемые в примере числа «-1» и «1»). В результате составлена структурная схема нейронной сети и подобраны значения ее коэффициентов. Экспериментами показано, что скорость обучения нейронной сети при предварительно рассчитанных коэффициентах значительно выше, чем при случайных коэффициентах, взятых из нормального распределения в диапазоне $[-1, 1]$. Также проведены другие эксперименты по обучению нейронной сети с максимальными / минимальными начальными значениями коэффициентов из диапазона $[4, 20]$. В результате эксперимента при всех используемых начальных условиях получены различные значения коэффициентов, что приводит к выводу о том, что существует неединственное решение поставленной задачи, что несомненно является отдельной темой исследования. Также стоит отметить, что скорость обучения нейронной сети с синтезированной структурой выше, чем с у сетей прямого распространения с несколькими слоями (отдельно стоит сказать, что выбор количества слоев и нейронов является неоднозначным).

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

* Статья получена 24 января 2018 г.

ВВЕДЕНИЕ

В настоящее время актуальной задачей является создание систем, способных к обучению или на основании обучения, вместо разработки алгоритмов [1–6]. Для решения данных задач применяют различные инструменты и способы, начиная от случайных лесов (*randomtrees*) и заканчивая опорной машиной векторов (*supportvectormachine*). В настоящее время методы, основанные на использовании нейронных сетей [1–4], имеют наибольшую популярность в силу того, что с их помощью можно создавать более сложные модели. Так, в работах [1–4, 7] используются нейронные сети с возможностью обучения срока хранения информации, в [8] приводятся нейронные сети с памятью. В работе [9] исследуется задача создания нейронной машины Тьюринга, или частный случай использования нейронных сетей в алгоритмических задачах – задача сортировки [10].

Однако в настоящее время в описании моделей нейронных сетей [11–16] существует ряд нерешенных вопросов, к которым относятся выбор и разработка алгоритмов быстрого обучения (в частности, обоснование выбора структуры нейронной сети).

Данная работа посвящена исследованию выбора структуры нейронной сети на основании предварительно рассчитанной модели.

1. ПОСТАНОВКА ЗАДАЧИ И ЕЕ РЕШЕНИЕ

В настоящее время задача выбора структуры ИНС является нерешенной задачей. На практике выбор вида ИНС осуществляется эмпирически на основе опыта исследователя и множества попыток обучения. Таким образом, задача обоснования выбора структуры ИНС и /или ее предварительный расчет на основе других моделей является актуальной задачей.

В работе предлагается подход к формированию ИНС на основе построенной модели. В качестве исследуемого объекта будем использовать функцию выбора максимального числа из двух поступивших на вход системы (в работе ограничимся набором из двух чисел «-1» и «1»), несмотря на то что в классическом варианте используются числа «0» и «1»).

2. ИССЛЕДОВАНИЕ

Рассмотрим задачу определения наибольшего из двух чисел. Будем решать данную задачу путем расчета (или подбора) параметров нейронной сети, а саму структуру будем представлять в виде рисунка (см. ниже). На данном рисунке представлена структура нейронной сети, в которой входные данные поступают на входы «x1» и «x2». Далее умноженные на коэффициенты W_1 и

Таблица с начальными условиями и обученными коэффициентами

Начальные условия (W1, W2, B1, W3)				Коэффициенты после обучения (W1, W2, B1, W3)			
1				2			
4	4	4	4	7.619	7.619	7.619	12.104
4	4	4	20	6.986	6.986	6.986	27.694
4	4	20	4	9.763	9.763	14.236	11.157
4	4	20	20	8.830	8.830	15.169	21.452
4	20	4	4	9.763	14.236	9.763	11.157
4	20	4	20	8.830	15.169	8.830	21.453
4	20	20	4	7.610	20.000	20.000	12.108
4	20	20	20	6.931	20.003	19.996	27.748
20	4	4	4	14.236	9.763	9.763	11.158
20	4	4	20	15.169	8.830	8.832	1.453
20	4	20	4	20.000	7.610	20.000	12.108
20	4	20	20	20.003	6.931	19.996	27.748
20	20	4	4	20.000	20.000	7.611	12.108
20	20	4	20	20.000	20.000	6.933	27.748
20	20	20	4	21.121	21.121	18.878	12.116
20	20	20	20	20.012	20.012	19.987	27.882

Из результатов, представленных в таблице, можно сделать вывод о том, что при различных начальных условиях получаются различные коэффициенты итоговой нейронной сети.

ЗАКЛЮЧЕНИЕ

В данной статье проанализирован пример решения задачи выбора максимального числа из двух с ограниченным диапазоном значений «-1» и «1». При решении данной задачи была предварительно разработана структура нейронной сети и рассчитаны коэффициенты. Итоговая структура нейронной сети содержат два нейрона. Выполнен ряд экспериментов по обучению полученной нейронной сети. В частности, за 20 000 итераций обучения при начальных значениях, выбранных случайным равномерным образом из диапазона [0, 1], получены результаты (в формате «x1/x2 входы: выход»): «1/1 : 0.80», «-1/1 : 0.71», «1/-1 : 0.71», «-1/-1 : -0.75». При использовании рассчитанных начальных условий за 3000 итераций получены более точные результаты, а именно: «1/1 : 0.974», «-1/1 : 0.972», «1/-1 : 0.972», «-1/-1 : -0.971».

По результатам эксперимента по обучению из возможных начальных условий были получены различные значения коэффициентов нейронной сети. Из этого можно сделать вывод о существовании множества решений в нейронной сети (даже состоящей только из двух нейронов).

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. *Haykin S.* Neural networks: a comprehensive foundation. – New York: MacMillan Publ., 1994. – 1104 p.
8. End-to-end memory networks / S. Sukhbaatar, A. Szlam, J. Weston, R. Fergus // Advances in Neural Information Processing Systems 28. – Montreal, Canada, 2015. – P. 2440–2448.
9. *Graves A. Wayne G., Danihelka I.* Neural Turing machines // ArXiv.org. – 2014. – arXiv:1410.5401. – P. 1–26.
10. *Воевода А.А., Полубинский В.Л., Романников Д.О.* Сортировка массива целых чисел с использованием нейронной сети // Научный вестник НГТУ. – 2016. – № 2 (63). – С. 151–157.
11. *Hastie T., Tibshirani R., Friedman J.* The elements of statistical learning: data mining, inference, and prediction. – New York: Springer. 2001. – 533 p.
12. *Hagan M., Demuth H., Jesús O.* A neural network predictive control system for paper mill wastewater treatment // Engineering Applications of Artificial Intelligence. – 2003. – Vol. 16 (2). – P. 121–129.
13. *Touretzky D.S., Pomerleau D.A.* What is hidden in the hidden layers? // Byte. – 1989. – Vol. 14. – P. 227–233.
14. *LeCun Y., Bengio Y., Hinton G.* Deep learning // Nature. – 2015. – Vol. 521, N 7553. – P. 436–444.

15. Dropout: a simple way to prevent neural networks from overfitting / N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, R. Salakhutdinov // Journal of Machine Learning Research. – 2014. – Vol. 15. – P. 1929–1958.

16. Early stopping without a validation set / M. Mahsereci, L. Balles, C. Lassner, P. Hennig // ArXiv.org. – 2017. – arXiv:1703.09580.

Воевода Александр Александрович, доктор технических наук, профессор кафедры автоматки 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 автоматки Новосибирского государственного технического университета. Основное направление научных исследований – нейронные сети, сети Петри. Имеет более 50 публикаций. E-mail: dmitry.romannikov@gmail.com

DOI: 10.17212/2307-6879-2018-2-93-100

Training of neural network with preliminary calculated structure*

А.А. Voevoda¹, D.O. Romannikov²

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*

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 has a very negative effect on the threshold of entry into the technology, the validity and correctness of the model's work, training time and other aspects. The article deals with the approach to the formation of the neural network structure on the basis of the preliminary formation of the structural scheme of the system and the calculation or selection of its coefficients. The goal is to demonstrate that the synthesis of the neural network structure can be performed not only by intuitively selecting layers and links. For compactness, the proposed approach is illustrated by the example of selecting the maximum value from two numbers (used in the example numbers - "-1", "1"). As a result, a structural diagram of the neural network is made up and the values of its coefficients are chosen. Experiments show that the speed of training a neural network

*Received 24 January 2018.

with previously calculated coefficients is much higher than for random coefficients taken from the normal distribution in the range $[-1, 1]$. Also, other experiments were conducted on training the neural network with the maximum / minimum initial values of the coefficients from the range $[4, 20]$. As a result of the experiment, for all the initial conditions used, different values of the coefficients were obtained, which leads to the conclusion that there is a non-unique solution to the problem posed, which is undoubtedly a separate topic of the study. It should also be noted that the learning speed of a neural network with a synthesized structure is higher than that of networks with direct propagation with several layers (separately it is necessary to say that the choice of the number of layers and neurons is ambiguous).

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

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.
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. Haykin S. *Neural networks: a comprehensive foundation*. New York, MacMillan Publ., 1994. 1104 p.
8. Sukhbaatar S., Szlam A., Weston J., Fergus R. End-to-end memory networks. *Advances in Neural Information Processing Systems 28*, Montreal, Canada, 2015, pp. 2440–2448.
9. Graves A. Wayne G., Danihelka I. Neural Turing machines. *The Computing Research Repository* 1410.5401, 2014, pp. 1–24.
10. Voevoda A.A., Polubinsky V.L., Romannikov D.O. Sortirovka massiva tseilykh chisel s ispol'zovaniem neironnoi seti [Sorting the array of integers using a neural network]. *Nauchnyi vestnik Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Science bulletin of the Novosibirsk state technical university*, 2016, no. № 2 (63), pp. 151–157.
11. Hastie T., Tibshirani R., Friedman J. *The elements of statistical learning: data mining, inference, and prediction*. New York, Springer, 2001, 533 p.

12. Hagan M., Demuth H., Jesús O. A neural network predictive control system for paper mill wastewater treatment. *Engineering Applications of Artificial Intelligence*, 2003, vol. 16 (2), pp. 121–129.

13. Touretzky D.S., Pomerleau D.A. What is hidden in the hidden layers? *Byte*, 1989, vol. 14, pp. 227–233.

14. LeCun Y., Bengio Y., Hinton G. Deep learning. *Nature*, 2015, vol. 521, no. 7553, pp. 436–444.

15. Srivastava N., Hinton G., Krizhevsky A., Sutskever I., Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 2014, vol. 15, pp. 1929–1958.

16. Mahsereci M., Balles L., Lassner C., Hennig P. Early stopping without a validation set. *ArXiv.org*, 2017, arXiv:1703.09580.

Для цитирования:

Воевода А.А., Романников Д.О. Обучение нейронной сети с предварительно рассчитанной структурой // Сборник научных трудов НГТУ. – 2018. – № 2 (92). – С. 93–100. – doi: 10.17212/2307-6879-2018-2-93-100.

For citation:

Voevoda A.A. Romannikov D.O. Obuchenie neironnoi seti s predvaritel'no rasschitannoi strukturoi [Training of neural network with preliminary calculated structure]. *Sbornik nauchnykh trudov Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Transaction of scientific papers of the Novosibirsk state technical university*, 2018, no. 2 (92), pp. 93–100. doi: 10.17212/2307-6879-2018-2-93-100.