

ИНФОРМАТИКА,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И УПРАВЛЕНИЕ

INFORMATICS,
COMPPUTER ENGINEERING
AND CONTROL

УДК 519.233.22

DOI: 10.17212/1814-1196-2018-3-33-42

Синтез нейронной сети для реализации рекуррентного метода наименьших квадратов*

А.А. ВОЕВОДА^а, Д.О. РОМАННИКОВ^б

630073, РФ, г. Новосибирск, пр. Карла Маркса, 20, Новосибирский государственный технический университет

^а voevoda@ucit.ru ^б dmitry.romannikov@gmail.com

В статье рассматривается синтез нейронной сети для решения задачи вычисления параметров P , K , θ статических объектов рекуррентным методом наименьших квадратов для скалярного случая. При этом структура нейронной сети получена из реализуемых соотношений, в которых каждая функция двух переменных аппроксимируется отдельной нейронной сетью. Для вычисления параметров K и P достаточно только аппроксимировать исходные соотношения, так как они сами зависят только от двух переменных (т. е. являются функциями двух переменных). Для вычисления параметра θ было необходимо представить исходное соотношение в виде функций двух переменных, а затем по отдельности построить нейронные сети для их аппроксимации и синтезировать итоговую нейронную сеть. Приведена структура синтезированной нейронной сети, основанная на использовании структуры для аппроксимации функции двух переменных, позволяющей рассчитать параметры нейронной сети и сумматоров, выполненных на нейронах. При этом часть связей является рекуррентной. Проведены эксперименты для подтверждения работы синтезированной нейронной сети для расчета заданных соотношений 1) при рассчитанных параметрах нейронной сети; 2) при параметрах нейронной сети, полученных путем обучения методом градиентного спуска обратного распространения ошибки для синтезированной структуры нейронной сети. Представлены графики, показывающие, что вычисленные с помощью синтезированной нейронной сети искомые параметры в достаточной степени соответствуют полученным в результате вычислений по исходным формулам значениям параметров в заданном диапазоне изменения водных переменных. Полученная нейронная сеть обладает такой же сходимостью, что и реализуемые соотношения.

Ключевые слова: нейронные сети, синтез, классификация, структура нейронных сетей, методы обучения, машинное обучение, обратное распространение ошибки, структурные методы, алгоритмы управления

* Статья получена 10 мая 2018 г.

ВВЕДЕНИЕ

В настоящее время применение нейронных сетей для решения задач, связанных с разработкой искусственного интеллекта, приобрело большую популярность. В частности, они широко применяются для распознавания объектов на изображениях [1], распознавания звуков и речи [2, 3], решения проблем, связанных с синтезом систем оптимального управления [4–6], и т. д. При этом решения каждой из вышеуказанных задач имеют общие недостатки, связанные с общими проблемами нейронных сетей: отсутствие однозначных рекомендаций по выбору структуры сети, типов нейронов, начальных параметров (более подробно рассмотрено в [7–11]). Кроме того, использование методов, основанных на градиентном спуске для оптимизации стоимостной функции в нейронных сетях, не гарантируют сходимость [7–9].

Преодоление вышеприведенных недостатков на практике достигается перебором различных вариантов архитектур, типов нейронов, начальных условий, методов обучения и дальнейшим выбором наиболее удачного решения. С другой стороны, в [12] представлен подход, в котором выполняется синтез нейронной сети, а также расчет параметров таким образом, что итоговая нейронная сеть не требует обучения. В [13, 14] представлен подход к синтезу нейронной сети на основании предварительно рассчитанной сети Петри. Также представлена статья, в которой предлагается метод синтеза для аппроксимации функций двух переменных. Ее особенностью является предлагаемая структура нейронной сети и способ расчета параметров нейронной сети. Данный метод был усовершенствован в части оптимизации количества нейронов в нейронной сети и в части масштабирования предлагаемого метода для произвольного количества входных переменных.

Во всех вышеприведенных работах синтезируемые нейронные сети являлись сетями прямого распространения, т. е. не было примеров синтеза рекуррентных нейронных сетей.

Использование многослойного персептрона несколько затруднено для решения данной задачи. К примеру, обучение нейронной сети для расчета параметра K в [15] из одного входного слоя, трех скрытых слоев и одного выходного слоя (число нейронов по слоям: 1, 5, 4, 3, 1 соответственно) с линейными функциями активации методом градиентного спуска и стоимостной функцией суммы квадратов ошибки не всегда выполняется, а приводит к расхождению.

В настоящей статье показывается применение методики синтеза нейронных сетей для решения задачи вычисления параметров статических объектов рекуррентным методом наименьших квадратов [15] для скалярного случая.

1. ПОСТАНОВКА ЗАДАЧИ

В статье рассматривается задача синтеза нейронной сети для вычисления параметров статических объектов рекуррентным методом наименьших квадратов [15] для скалярного случая $y = \theta x$, т. е. вычисления θ_{N+1} , K_{N+1} , P_{N+1} :

$$\begin{cases} \hat{\theta}_{N+1} = \hat{\theta}_N + K_{N+1}(y_{N+1} - x_{N+1}\hat{\theta}_N), \\ K_{N+1} = P_N x_{N+1} / (1 + x_{N+1} P_N x_{N+1}), \\ P_{N+1} = \left(1 - P_N \frac{x_{N+1} x_{N+1}}{1 + x_{N+1} P_N x_{N+1}} \right) P_N. \end{cases}$$

В данной системе уравнений $\theta_N, K_N, P_N, x_N, y_N$ являются скалярами; x – входной переменной; y – выходной переменной; x_N, y_N – обозначают N -е измерение; параметры θ_N, K_N, P_N – вычисляемые параметры, из которых K_N, P_N являются вспомогательными переменными, с помощью которых оценивается параметр θ_N .

Синтезируемая нейронная сеть должна состоять только из нейронов и дифференцируемых функций активаций (т. е. для итоговой нейронной сети должно быть возможным обучение одним из градиентных методов обратного распространения ошибки). В качестве функции активации будет использоваться функция выпрямителя (*rectifier*) – *relu*.

2. СИНТЕЗ НЕЙРОННОЙ СЕТИ

Так как реализуемые соотношения из раздела постановки задачи являются рекуррентными, то очевидно, что и нейронная сеть для их реализации должна иметь обратные связи (т. е. являться рекуррентной), что поднимает вопрос об их устойчивости при обучении [7, 8].

Уравнения для решения соотношений для расчета P_{N+1}, K_{N+1} являются уравнениями двух переменных и могут быть аппроксимированы^{1,2,3}. Для аппроксимации уравнения при расчете θ_{N+1} можно представить уравнение в виде выражений функций двух переменных, а далее суммировать полученный результат. Блок-схема последовательности вычисления приведена на рис. 1. Эта блок-схема является основной структурой синтезируемой нейронной сети.

На рис. 1 функции $F_1 - F_4$ являются функциями двух переменных, аппроксимация которых приведена в статьях^{1,2,3}. Более детально ее структура изображена на рис. 2, 3, а коэффициенты получены путем обучения одним из методов градиентного пуска. Пунктиром выделены соответствующие исходным выражениям части блок-схемы.

¹ Voevoda A.A., Romannikov D.O. Approximations of two-dimensional surfaces using neural networks // The 13th International Forum on Strategic Technology (IFOST 2018) : proc., Harbin, China, 30 May – 1 June 2018.

² Voevoda A.A., Romannikov D.O., Troshina G.V. The Method of Synthesis of a Neural Network for the Approximation of Surfaces of Two Variables // The XXI International Conference on Soft Computing and Measurement (SCM'2018) : proc., St. Petersburg, Russia, May 23–25 2018.

³ Voevoda A.A., Romannikov D.O. Synthesis of a Neural Network for N-Dimension Surfaces Approximation // Actual problems of electronic instrument engineering (APEIE–2018).

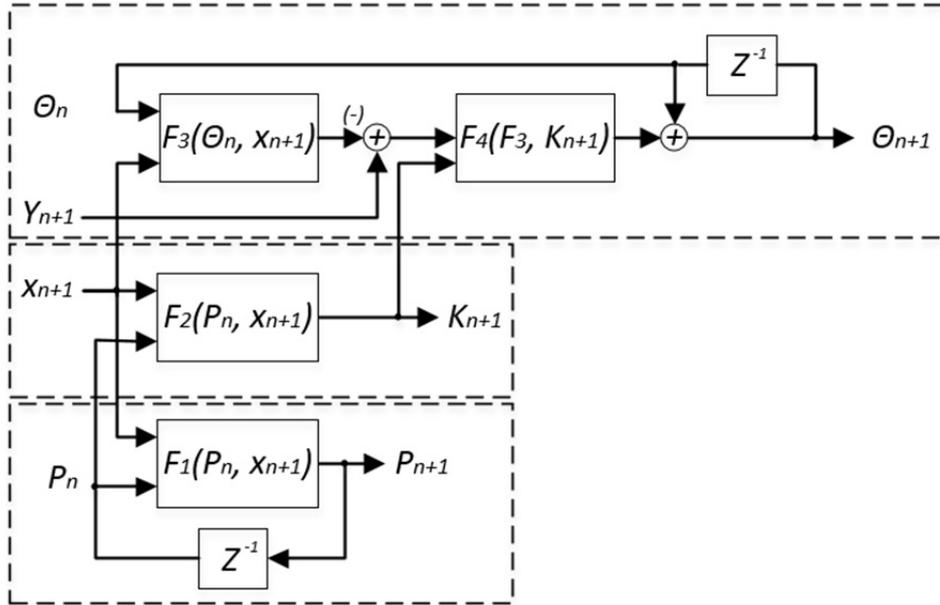


Рис. 1. Блок-схема алгоритма вычислений рассматриваемых соотношений

Fig. 1. The block diagram of the computation algorithm of relationships under study

На рис. 2 приведена часть нейронной сети, где кругами обозначены нейроны. Первые расположенные параллельно четыре нейрона с функцией активации выпрямителя (*relu*) необходимы для построения трапеции, с помощью которой выделяется поддиапазон входной координаты. Далее фигурой трапеции обозначается данная часть нейронной сети (рис. 3). Выходной нейрон необходим для суммирования выходных значений нейронов первого слоя. Его выходное значение символизирует то, что значение входной переменной принадлежит поддиапазону. Например, если x_1 имеет значение 0.7, а нейронная сеть на рис. 2 «улавливает» поддиапазон от 1 до 2, то на выходе будет значение, близкое к нулю, а если x_1 будет принадлежать рассматриваемому поддиапазону, то близкое к единице.

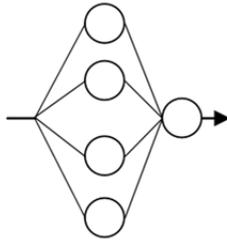


Рис. 2. Структура нейронной сети для выделения признака принадлежности к поддиапазону

Fig. 2. The neural network structure to extract the attribute of the membership to the sub-band

Диапазон каждого значения входной переменной x_1, x_2 разбивается на поддиапазоны, и определяется принадлежность значения входной переменной одному из поддиапазонов (второй слой, рис. 2). Третий слой формируется путем пересечения каждого из выходных нейронов поддиапазона одной размерности (например, x_1) с каждым из выходных нейронов поддиапазона следующей размерности (например, x_2).

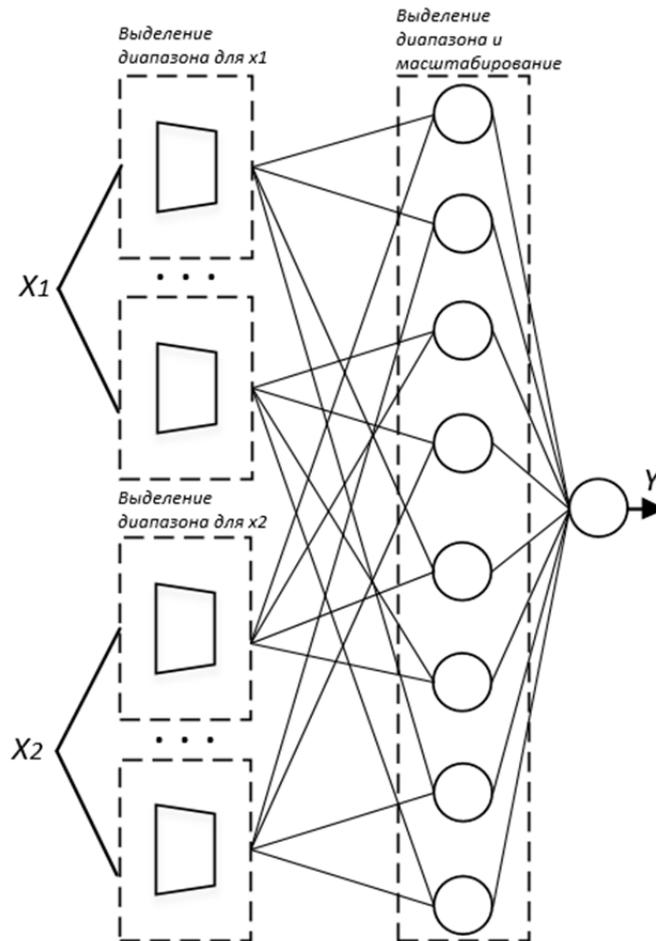


Рис. 3. Структура нейронной сети для аппроксимации функций $F_1 - F_4$

Fig. 3. The neural network structure to approximate the $F_1 - F_4$ functions

В четвертом слое выполняется объединение значений поддиапазонов по логической функции «и» для определения итогового поддиапазона. В последнем слое выполняется суммирование выходных значений предыдущего слоя. Особенностью данной структуры является то, что коэффициенты данной нейронной сети можно получить как путем градиентного обучения, так и расчетом.

3. МОДЕЛИРОВАНИЕ НЕЙРОННОЙ СЕТИ

При моделировании значения переменных x_{N+1} и y_{N+1} изменялись по функции $f(x) = \sin(x)$. Таким образом, значение коэффициента θ должно сходиться к единице.

Результаты значений, полученных нейронной сетью и рассчитанных по формальным выражениям из постановки задачи, приведены на рис. 4.

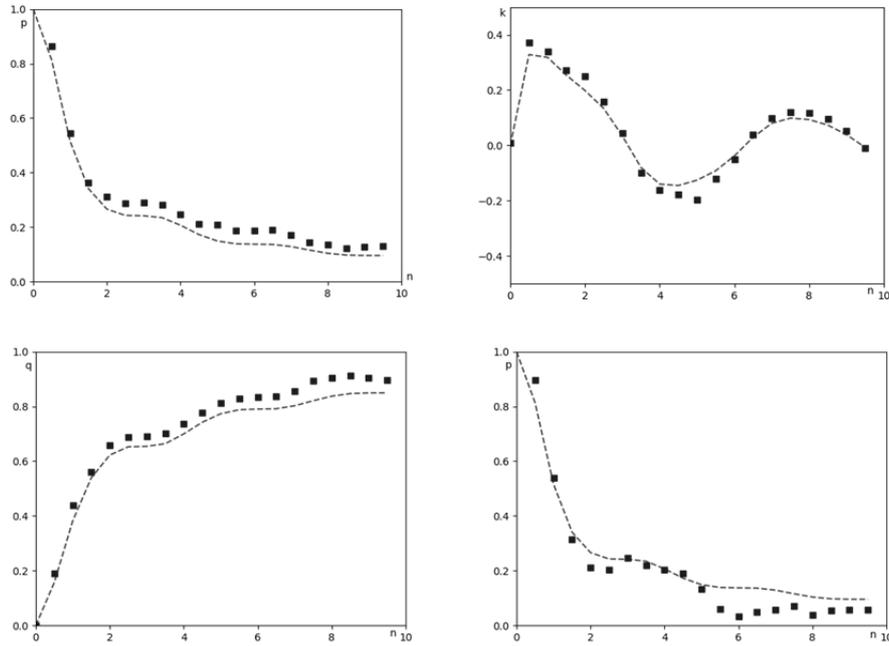


Рис. 4. Графики функций для значений P_{N+1} (левый верхний), K_{N+1} (правый верхний), θ_{N+1} (левый нижний), P_{N+1} (правый нижний). Пунктирной линией отмечены данные, полученные по формулам из постановки задачи, квадратами – при помощи нейронной сети

Fig. 4. Function graphs for the values P_{N+1} (upper left), K_{N+1} (upper right), θ_{N+1} (lower left) and P_{N+1} (lower right). Dotted lines show data obtained by the formulas from the problem statement and squares show data obtained with the help of a neural network

Из приведенных результатов (рис. 4) видно, что полученные при помощи нейронной сети значения коэффициентов θ_{N+1} , K_{N+1} , P_{N+1} в достаточной степени близки к данным, посчитанным по формулам. Левый верхний и правый нижний графики отличаются тем, что первый получен от нейронной сети, в которой параметры рассчитаны, а второй – путем обучения.

Также выполнялись эксперименты по расчету P_{N+1} при помощи нейронной сети, но с параметрами обучения, полученными путем обучения, а не расчета. Результаты приведены на нижнем правом графике на рис. 4. Нейронная сеть для расчета P_{N+1} была обучена при помощи адаптивного метода субградиента для онлайн-обучения и стохастической оптимизации (*AdamGrad*) [8] за 100 эпох с шагом обучения 0.1 без уменьшения шага обучения (*weight decay*).

На рис. 5 приведены результаты моделирования синтезированной нейронной сети при зашумленном входном сигнале. В качестве источника шума, как и в [15], взят генератор шума с нормальным законом распределения с отклонением $\sigma = 0.1$.

По приведенным графикам (рис. 5) видно, что, несмотря на то что график параметра P_{N+1} меняется под воздействием шума, на остальные параметры шум оказывает меньшее воздействие.

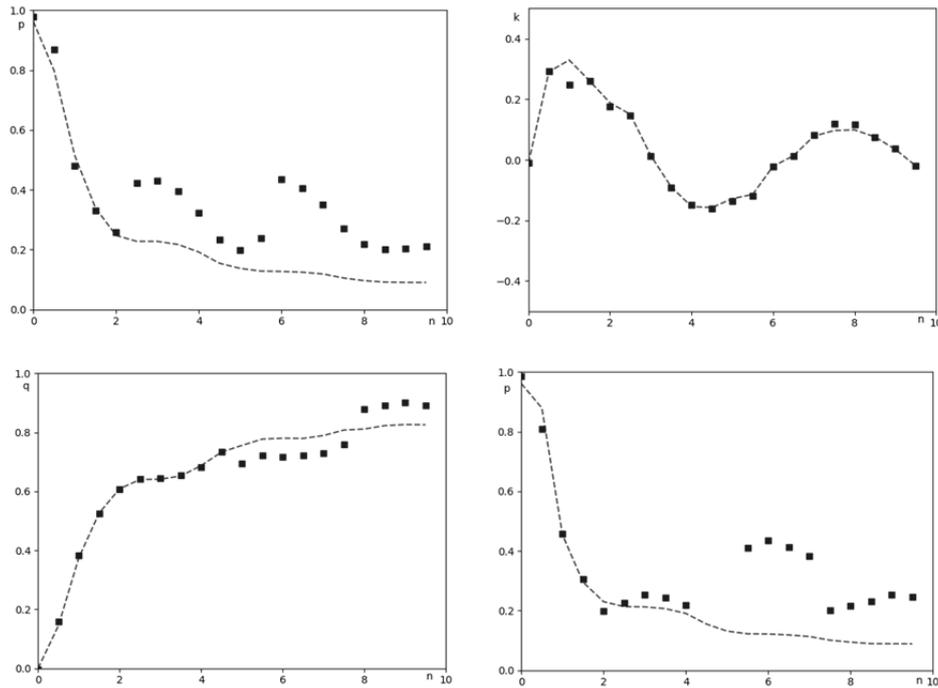


Рис. 5. Графики функций для значений P_{N+1} (левый верхний), K_{N+1} (правый верхний), θ_{N+1} (левый нижний), P_{N+1} (правый нижний) при зашумленном входном сигнале. Пунктирной линией отмечены данные, полученные по формулам из постановки задачи, квадратами – при помощи нейронной сети

Fig. 5. Function graphs for the values P_{N+1} (upper left), K_{N+1} (upper right), θ_{N+1} (lower left) and P_{N+1} (lower right) with noisy input signal. Dotted lines show data obtained by the formulas from the problem statement and squares show data obtained with the help of a neural network

ЗАКЛЮЧЕНИЕ

В статье приводится реализация структуры нейронной сети для вычисления параметров статических объектов рекуррентным методом наименьших квадратов [15] для скалярного случая: q_{N+1} , K_{N+1} , P_{N+1} . Полученная нейронная сеть построена путем использования структуры для аппроксимации функции двух переменных. Причем соотношения для вычисления K_{N+1} , P_{N+1} являются функциями двух переменных и могут быть аппроксимированы без дополнительных модификаций; соотношение для расчета θ_{N+1} было представлено в виде выражения над функциями двух переменных и далее аппроксимировано таким же способом, что и предыдущие соотношения. Моделирование показало, что предлагаемая структура нейронной сети позволяет решить поставленную задачу. Исследуемый процесс зависит от начальных условий. В частности, начальное значение для параметра P нельзя выбирать близким к нулю (должно быть положительное).

Преимуществом предлагаемого подхода к построению структуры нейронной сети заключается в том, что множество недостатков, среди которых есть такие, как неопределенность при выборе количества слоев, типов

нейронов, нелинейностей и другие, теряют свою актуальность, так как сама структура нейронной сети получается из алгоритмического решения поставленной задачи. Полученная структура нейронной сети состоит из восьми слоев, 240 нейронов и 727 параметров обучения. При этом присутствуют обратные связи не только к входному слою, но и к скрытым.

Одним из направлений дальнейшего развития является минимизация количества параметров обучения.

СПИСОК ЛИТЕРАТУРЫ

1. Krizhevsky A., Sutskever I., Hinton G.E. ImageNet classification with deep convolutional neural networks // *Advances in Neural Information Processing Systems 25 (NIPS 2012)*. – Lake Tahoe, Nevada, 2012. – Vol. 1. – P. 1097–1105.
2. Graves A., Mohamed A., Hinton G. Speech recognition with deep recurrent neural networks // *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2013*. – Vancouver, Canada, 2013. – P. 6645–6649.
3. Deng L., Hinton G.E., Kingsbury B. New types of deep neural network learning for speech recognition and related applications: an overview // *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2013*. – Vancouver, Canada, 2013.
4. Playing Atari with deep reinforcement learning / V. Mnih, K. Kavukcuoglu, D. Silver, A. Graves, I. Antonoglou, D. Wierstra, M. Riedmiller // *NIPS 2013 Workshops*. – Lake Tahoe, 2013.
5. Learning from demonstrations for real world reinforcement learning / T. Hester, M. Vecerik, O. Pietquin, M. Lanctot, T. Schaul, B. Piot, D. Horgan, J. Quan, A. Sendonaris, G. Dulac-Arnold, I. Osband, J. Agapiou, J.Z. Leibo, A. Gruslys // *ArXiv.org*. – 2017. – arXiv:1704.03732.
6. Prioritized experience replay / T. Schaul, J. Quan, I. Antonoglou, D. Silver // *Proceeding ICLR 2016*. – P. 1260–1268.
7. Bishop C. *Pattern recognition and machine learning*. – New York: Springer, 2007. – 738 p. – (Information science and statistics).
8. Goodfellow I., Bengio Y., Courville A. *Deep learning*. – Cambridge: MIT Press, 2016. – 800 p.
9. LeCun Y., Bengio Y., Hinton G. Deep learning // *Nature*. – 2015. – Vol. 521, N 7553. – P. 436–444.
10. 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.
11. Early stopping without a validation set / M. Mahsereci, L. Balles, C. Lassner, P. Hennig // *ArXiv.org*. – 2017. – arXiv:1703.09580.
12. Воевода А.А., Романников Д.О. Синтез нейронной сети для решения логико-арифметических задач // *Труды СПИИРАН*. – 2017. – Вып. 54. – С. 205–223.
13. Романников Д.О. О синтезе нейронных сетей // *Сборник научных трудов НГТУ*. – 2018. – № 1 (91). – С. 104–111.
14. Воевода А.А., Романников Д.О. Синтез нейронных сетей с несколькими переменными // *Сборник научных трудов НГТУ*. – 2018. – № 1 (91). – С. 86–94.
15. Воевода А.А., Трошина Г.В. Оценивание параметров линейных статистических объектов с использованием рекуррентного метода наименьших квадратов в среде Simulink // *Сборник научных трудов НГТУ*. – 2016. – № 3 (85). – С. 33–48.

Воевода Александр Александрович, доктор технических наук, профессор кафедры автоматизации Новосибирского государственного технического университета. Область научных интересов: полиномиальный синтез, сети Петри, UML-диаграммы. Число научных публикаций – более 200. E-mail: voevoda@ucit.ru

Романников Дмитрий Олегович, кандидат технических наук, доцент кафедры автоматизации Новосибирского государственного технического университета. Область научных интересов: машинное обучение, нейронные сети, сети Петри. Число научных публикаций – 51. E-mail: dmitry.romannikov@gmail.com

DOI: 10.17212/1814-1196-2018-3-33-42

Synthesis of a neural network for the implementation of the recursive least-squares method*

A.A. VOEVODA^a, D.O. ROMANNIKOV^b

Novosibirsk State Technical University, 20, K. Marx Prospekt, Novosibirsk, 630073, Russian Federation

^a voevoda@ucit.ru ^b dmitry.romannikov@gmail.com

Abstract

The article deals with the synthesis of a neural network for solving the problem of calculating the P, K, θ parameters of static objects by the recursive least-squares method for the scalar case. In doing so, the structure of the neural network is derived from realizable relationships in which each function of two variables is approximated by a separate neural network. To calculate the parameters K and P, it is sufficient only to approximate the original relations as they depend on only two variables (that is, they are functions of two variables). To calculate the θ parameter, it was necessary to represent the initial relation in the form of an expression over the functions of two variables, and then separately build neural networks to approximate them and synthesize the final neural network. The structure of a synthesized neural network is based on the use of a structure to approximate a function of two variables that allows one to calculate the parameters of a neural network and summators implemented on neurons. In this case, some of the links are recurrent. Experiments were performed to confirm the operation of the synthesized neural network for calculating the specified relationships 1) with the calculated neural network parameters; 2) with the neural network parameters obtained by training the method of gradient descent of the backward propagation error for the synthesized structure of the neural network. The graphs show that the parameters calculated with the help of a synthesized neural network correspond quite sufficiently to the values of the parameters obtained in the calculation of the initial formulas in the given range of water variables variation. The resulting neural network has the same convergence as the realizable relations.

Keywords: neural networks, synthesis, classification, structure of neural networks, learning techniques, machine learning, back propagation errors, structural methods, control algorithms

REFERENCES

1. Krizhevsky A., Sutskever I., Hinton G.E. ImageNet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, Lake Tahoe, Nevada, 2012, vol. 1, pp. 1097–1105.
2. Graves A., Mohamed A., Hinton G. Speech recognition with deep recurrent neural networks. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2013*, Vancouver, Canada, 2013, pp. 6645–6649.

* Received 10 May 2018.

3. Deng L., Hinton G.E., Kingsbury B. New types of deep neural network learning for speech recognition and related applications: an overview. *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2013*, Vancouver, Canada, 2013.
4. Mnih V., Kavukcuoglu K., Silver D., Graves A., Antonoglou I., Wierstra D., Riedmiller M. Playing Atari with deep reinforcement learning. *NIPS 2013 Workshops*, Lake Tahoe, 2013.
5. Hester T., Vecerik M., Pietquin O., Lanctot M., Schaul T., Piot B., Horgan D., Quan J., Sendonaris A., Dulac-Arnold G., Osband I., Agapiou J., Leibo J.Z., Gruslys A. Learning from demonstrations for real world reinforcement learning. *ArXiv.org*, 2017, arXiv:1704.03732.
6. Schaul T., Quan J., Antonoglou I., Silver D. Prioritized experience replay. *Proceeding ICLR 2016*, pp. 1260–1268.
7. Bishop C. *Pattern recognition and machine learning. Information science and statistics*. New York, Springer, 2007. 738 p.
8. Goodfellow I., Bengio Y., Courville A. *Deep learning*. Cambridge, MIT Press, 2016. 800 p.
9. LeCun Y., Bengio Y., Hinton G. Deep learning. *Nature*, 2015, vol. 521, no. 7553, pp. 436–444.
10. 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.
11. Mahsereci M., Balles L., Lassner C., Hennig P. Early stopping without a validation set. *ArXiv.org*, 2017, arXiv:1703.09580.
12. Voevoda A.A., Romannikov D.O. Sintez neironnoi seti dlya resheniya logiko-arifmeticheskikh zadach [Synthesis of neural network for solving logical-arithmetic problems]. *Trudy SPIIRAN – SPIIRAS proceedings*, 2017, vol. 54, no. 5, pp. 205–223.
13. Romannikov D.O. O sinteze neironnykh setei [About of synthesis of neural networks]. *Sbornik nauchnykh trudov Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Transaction of scientific papers of the Novosibirsk state technical university*, 2018, no. 1 (91), pp. 104–111.
14. Voevoda A.A., Romannikov D.O. Sintez neironnykh setei s neskol'kimi peremennymi [Synthesis of neural networks with several variables]. *Sbornik nauchnykh trudov Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Transaction of scientific papers of the Novosibirsk state technical university*, 2018, no. 1 (91), pp. 86–94.
15. Voevoda A.A., Troshina G.V. Otsenivanie parametrov lineinykh statisticheskikh ob"ektov s ispol'zovaniem rekurrentnogo metoda naimen'shikh kvadratov v srede Simulink [Estimation of the parameters of linear statistical objects using the recursive least-squares method in the Simulink environment]. *Sbornik nauchnykh trudov Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Transaction of scientific papers of the Novosibirsk state technical university*, 2016, no. 3 (85), pp. 33–48.

Для цитирования:

Воевода А.А., Романников Д.О. Синтез нейронной сети для реализации рекуррентного метода наименьших квадратов // Научный вестник НГТУ. – 2018. – № 3 (72). – С. 33–42. – doi: 10.17212/1814-1196-2018-3-33-42.

For citation:

Voevoda A.A., Romannikov D.O. Sintez neironnoi seti dlya realizatsii rekurrentnogo metoda naimen'shikh kvadratov [Synthesis of a neural network for the implementation of the recursive least-squares method]. *Nauchnyi vestnik Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Science bulletin of the Novosibirsk state technical university*, 2018, no. 3 (72), pp. 33–42. doi: 10.17212/1814-1196-2018-3-33-42.