

Алгоритм формирования обучающего множества и уточнения характеристики нейрона сети типа многослойный персептрон

А.А. Пастухов, А.А. Прокофьев

*Национальный исследовательский университет «МИЭТ», г.Москва,
Россия*

pastuhov1992@gmail.com

Алгоритм предварительной обработки данных для обучения нейронной сети типа многослойный персептрон используется для формирования обучающего множества из исходных данных, а также для уточнения вида нейрона за счет вычисления его характеристики на базе выборочных констант Липшица и определения константы Липшица нейронной сети. В работе показано применение алгоритма предварительной обработки данных типа многослойный персептрон для формирования обучающего множества и приведены вычисления характеристики нейрона конкретной нейронной сети. Разработанный алгоритм позволяет сформировать обучающее множество со значением энтропии больше, чем при формировании его случайным образом, а также дать оценку характеристики нейрона снизу и уточнить его вид. При решении задачи с применением описанного алгоритма получены прирост энтропии обучающего множества на 0,14 бит и уменьшение среднеквадратичной ошибки обучения на 0,098.

Ключевые слова: нейронная сеть; алгоритм кластеризации; предварительная обработка данных; многослойный персептрон; самоорганизующиеся карты Кохонена

Для цитирования: Пастухов А.А., Прокофьев А.А. Алгоритм формирования обучающего множества и уточнения характеристики нейрона сети типа многослойный персептрон // Изв. вузов. Электроника. – 2018. – Т. 23. – № 5. – С. 512–520. DOI: 10.24151/1561-5405-2018-23-5-512-520

Algorithm of a Representative Samples Formation and Calculation of a Neuron Characteristics for Multilayer Perceptron

A.A. Pastukhov, A.A. Prokofyev

National Research University of Electronic Technology, Moscow, Russia

pastuhov1992@gmail.com

Abstract: The article considers the algorithm of a representative samples formation for training the neural network of the multilayer perceptron. The purpose of the algorithm under consideration is the training set formation from the factor space and neuron characteristic calculation with application of the Lipschitz constant estimation. In the paper the application of the algorithm to the training set formation has been shown. The algorithm allows to form a training set with the entropy value more than when it is formed randomly, as well as to estimate the characteristics of the neuron from below and refine its form. The approach under consideration was concluded to have an influence on the increase in the entropy of the training set by 0,14 bit, decrease of mean square error by the value of 0,098 and (as a result) to lead to the quality improvement of training of multilayer perceptron with the small dimensionality of the factor space.

Keywords: neural network; clustering algorithm; representative sample; multilayer perceptron; self-organizing map

For citation: Pastukhov A.A., Prokofyev A.A. Algorithm of representative samples formation and calculation of neuron characteristics for multilayer perceptron. *Proc. Univ. Electronics*, 2018, vol. 23, no. 5, pp. 512–520. DOI: 10.24151/1561-5405-2018-23-5-512-520

Введение. Обучение искусственной нейронной сети, как правило, основано на алгоритмах локальной оптимизации [1]. Наиболее используемым для обучения многослойного персептрона (MLP) является алгоритм обратного распространения ошибки. До применения данного алгоритма необходим этап предварительной обработки исходных данных. Стандартными этапами перед обучением нейронной сети являются нормализация, масштабирование, а также начальная инициализация весов [2]. Эти этапы нельзя считать достаточными. С одной стороны, при небольшом количестве исходных данных следует учитывать специфику их распределения [3, 4] для обучения нейронной сети со среднеквадратичной ошибкой обучения, близкой к минимальной. С другой стороны, при большом количестве элементов, составляющих исходные данные, учет специфики распределения является достаточно сложной задачей. Таким образом, целесообразно применять методы и алгоритмы, которые позволяют проводить предварительную обработку исходных данных, учитывающую специфику их распределения.

Термины и определения. Введем следующие определения. Факторное пространство – множество исходных данных, на которых будет обучаться и тестироваться нейронная сеть. Обучающее множество – подмножество факторного пространства, на котором проводится обучение нейронной сети. Эффективность обучения – минимизация среднеквадратичной ошибки обучения нейронной сети, которая вычисляется по формуле

$$E = \frac{1}{2} \sum_{i=1}^n (y_i - d_i)^2,$$

где y_i – реальный отклик нейронной сети на i -й пример обучающего множества; d_i – ожидаемый отклик нейронной сети на i -й пример обучающего множества; n – количество элементов в обучающем множестве.

Тестовым множеством будем называть подпространство факторного пространства, на котором проводится тестирование уже обученной нейронной сети для проверки ее работоспособности.

В работе [5] предложен метод, решающий задачу формирования обучающего множества со значением энтропии, приближенным к максимальному, с использованием кластеризации. В работе [6] исследованы различные методы кластеризации применительно к этой задаче. Из выводов [6] следует, что эффективнее применение самоорганизующихся карт Кохонена [7, 8] к решению задачи кластеризации исходных данных.

Перед обучением необходимо определиться с архитектурой нейронной сети, т.е. с количеством слоев и нейронов в каждом слое, а также параметрами нейрона. Под параметрами нейрона подразумевается функция активации, которая вычисляет выходной сигнал. Достаточно часто используется сигмоидальная функция активации, которая имеет вид $y = [1 + e^{-cX}]^{-1}$. Здесь константа c определяет наклон функции активации и является единственным настраиваемым параметром. Оценить данную величину можно на базе соотношения константы Липшица нейронной сети и выборочных констант Липшица факторного пространства.

Константа Липшица сети ограничивает отношение нормы разницы значения функции в двух точках к норме разницы этих двух точек, т.е. определяется соотношением

$$L \geq \frac{\|f(x_1) - f(x_2)\|}{\|x_1 - x_2\|}. \quad (1)$$

Выборочная константа Липшица определяет сложность аппроксимации таблично заданной функции f в точках x_i и вычисляется по формуле [9]

$$\Lambda_i = \max_{i \neq j} \frac{\|y^i - y^j\|}{\|x^i - x^j\|}. \quad (2)$$

Оценка константы Липшица нейронной сети основана на ее иерархическом устройстве. В работе [10] приведен подробный вывод оценки константы Липшица нейронной сети с сигмоидальной функцией активации на базе констант Липшица ее отдельных элементов. Эта оценка имеет вид

$$\Lambda \leq c^m \sqrt{n_0 n_m} \prod_{i=1}^{m-1} n_i. \quad (3)$$

Соотношение между константой Липшица нейронной сети и выборочной константой Липшица элементов обучающего множества задается тем, что для адекватного обучения константа Липшица сети должна превосходить выборочные константы Липшица элементов обучающего множества, т.е. должно выполняться соотношение

$$\Lambda \geq \Lambda_{t_i}, \forall \Lambda_{t_i} \in \Lambda_t \subset \{\Lambda_i\}, \quad (4)$$

где $\{\Lambda_i\}$ – множество выборочных констант Липшица, вычисленных для элементов множества Z ; $\{\Lambda_{t_i}\} = \Lambda_{t_i}$ – множество выборочных констант Липшица, вычисленных для элементов обучающего подмножества T .

На основе соотношения (4) можно оценить величину характеристики нейрона c в формуле (3).

Обобщим результаты работ [5, 6] в виде описания алгоритма предварительной обработки данных на базе самоорганизующихся карт Кохонена. Цель алгоритма – нахождение обучающего подмножества $T \subset Z$, где Z – факторное пространство с точки зрения максимизации энтропии множества T , а также оценка характеристики нейрона c снизу на базе выборочной константы Липшица в зависимости от конфигурации нейронной сети.

Формализованная постановка задачи имеет следующий вид. Пусть $X = \{X^1, \dots, X^M, Y^1, \dots, Y^M\}$ – факторное пространство, где $X^i = \{x_1, x_2, \dots, x_{n-1}, x_n\}$; $Y^i = \{y_1, \dots, y_k\}$; x_i – i -й входной параметр; y_j – j -й выходной параметр; M – количество векторов в факторном пространстве; n – количество входных параметров; k – количество выходных параметров. Требуется построить алгоритм, который позволяет найти такое разбиение факторного пространства на три множества (обучающее, проверочное и тестовое), для которого выполняются условия:

$$H_0(T) < H(T) \leq H_{\max}(T), \\ S_T = S_{\min},$$

где $H(T)$, $H_0(T)$ – соответственно энтропия обучающего множества с использованием кластеризации и для случайного разбиения факторного пространства на представительскую выборку; $H_{\max}(T) = \log_2 N_t$ – максимальная энтропия этого множества; N_t – размер обучающего множества, составляющего 80 % от факторного пространства; S_T – среднеквадратичная ошибка обучающего множества.

Описание алгоритма предварительной обработки данных. Пусть $P = \{x_i\}$ – множество векторов-входов, $R = \{y_i\}$ – множество откликов на соответствующие векторы множества R , $x_i \in P$ – вектор-вход, $y_i \in R$ – отклик на вектор-вход $x_i \in P$, $Z = \{z_i\}$ – множество пар вида $z_i = \{x_i, y_i\}$, составляющих факторное пространство, N – полносвязная нейронная сеть типа многослойный персептрон, имеющая n_0 входных векторов, m слоев и n_i нейронов в i -м слое. Все нейроны сети имеют сигмоидальную функцию активации. Схематически алгоритм имеет вид, представленный на рис.1. Элементы множества $\{\Lambda_i\}$ Λ – константы, поэтому условие (4) может быть записано в виде $\Lambda \geq \max(\Lambda_{t_i})$, $\Lambda_{t_i} \in \Lambda_i$. Значение $L_{op} = \max(\Lambda_{t_i})$ назовем опорной точкой алгоритма.

На этапе 1 алгоритма вычисляются выборочные константы Липшица, формирующие множество $\{\Lambda_i\}$, элементы множества Z , а также константа Липшица сети Λ .

На этапе 2 проводится кластеризация множества Z с применением самоорганизующихся карт Кохонена (Self-organizing map, SOM) [7]. Результат данного этапа – разбиение множества $Z = \{K_j\}$, где K_j – подмножество, содержащее элементы, принадлежащие кластеру j . Тогда $z_i^j \in K_j \subset Z$, где z_i^j – элемент множества Z , принадлежащий кластеру j .

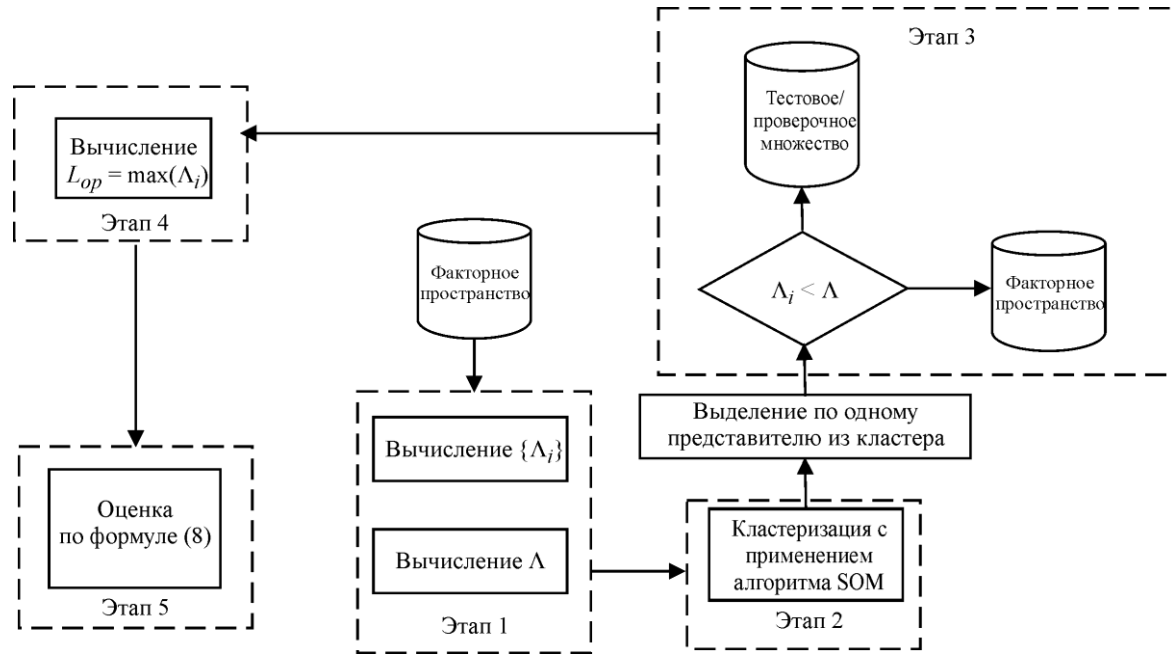


Рис.1. Алгоритм формирования представительской выборки на базе кластеризации
Fig.1. Algorithm of representation set forming

На *этапе 3* формируется тестовое множество, т.е. решение первой задачи алгоритма:

$$T = \{t_i\}, t_i \in K_j : \Lambda_{t_i} = \min(\Lambda_j^j), \quad (5)$$

где Λ_i^j – выборочная константа Липшица элемента $t_i \in K_j$; t_i – элемент множества K_j с минимальной выборочной константой Липшица Λ_{t_i} . Множество $\Lambda_t = \{\Lambda_{t_i}\}$ назовем множеством выборочных констант Липшица обучающего множества.

На *этапе 4* вычисляется опорная точка алгоритма по формуле

$$L_{op} = \max(\Lambda_{t_i}), \Lambda_{t_i} \in \Lambda_t. \quad (6)$$

На *этапе 5* с применением формулы (2) вычисляется значение характеристики нейрона, т.е. решается вторая задача алгоритма. Для удовлетворения условию (3) необходимо и достаточно выполнить условие $\Lambda \geq L_{op}$. Тогда условие (2) будет иметь вид

$$L_{op} \leq c^m \sqrt{n_0 n_m} \prod_{i=1}^{m-1} n_i. \quad (7)$$

Таким образом, оценка характеристики нейрона снизу для сети N , которая будет обучена на обучающем множестве T , вычисляется по формуле

$$c \geq \sqrt[m]{L_{op} \left[\sqrt{n_0 n_m} \prod_{i=1}^{m-1} n_i \right]^{-1}}. \quad (8)$$

Применение алгоритма. Рассмотрим применение описанного алгоритма к решению задачи обучения нейронной сети типа многослойный персептрон с целью оценки влияния алгоритма на качество обучения. Для этого предлагается синтезировать и обучить две одинаковые по архитектуре нейронные сети типа многослойный персептрон. Первую сеть необходимо обучить на случайно сформированном обучающем множестве

с нейронами, имеющими характеристику, выбранную случайным образом, вторую – на множестве, сформированном с применением алгоритма. Далее надо оценить прирост энтропии обучающего множества для случая, когда используется рассматриваемый алгоритм, а также сравнить среднеквадратичные ошибки обучения для обоих случаев и дать оценку влиянию энтропии на качество обучения нейронной сети.

Факторное пространство содержит четыре входных параметра, сформированных случайным образом, и один выходной параметр – отклик. Связь между входными и выходными параметрами задается нелинейной функцией

$$y = e^{x_1} + e^{x_2} + 2e^{x_3} - 3e^{x_4},$$

где y – выходной параметр; x_1, x_2, x_3, x_4 – соответствующие входные параметры.

Всего факторное пространство включает в себя 100 элементов. Обе нейронные сети имеют одинаковую архитектуру, представленную на рис.2. Для всех нейронов использована сигмоидальная функция активации, но для второй нейронной сети значение характеристики вычислено с применением оценки выборочных констант Липшица.

На рис.3,а приведены результаты обучения нейронной сети на данных, сформированных без применения алгоритма (графики на рис.3 получены в программе Matlab). Среднеквадратичная ошибка обучения для данного случая составила 0,214. Энтропия обучающего множества, вычисленная по формуле Шеннона, равна 6,05 бит.

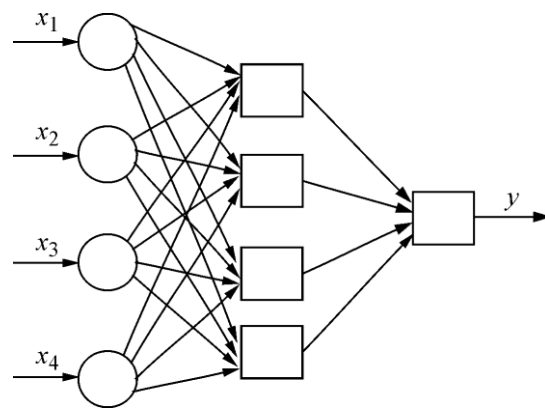


Рис.2. Архитектура нейронной сети
Fig.2. Neural network architecture

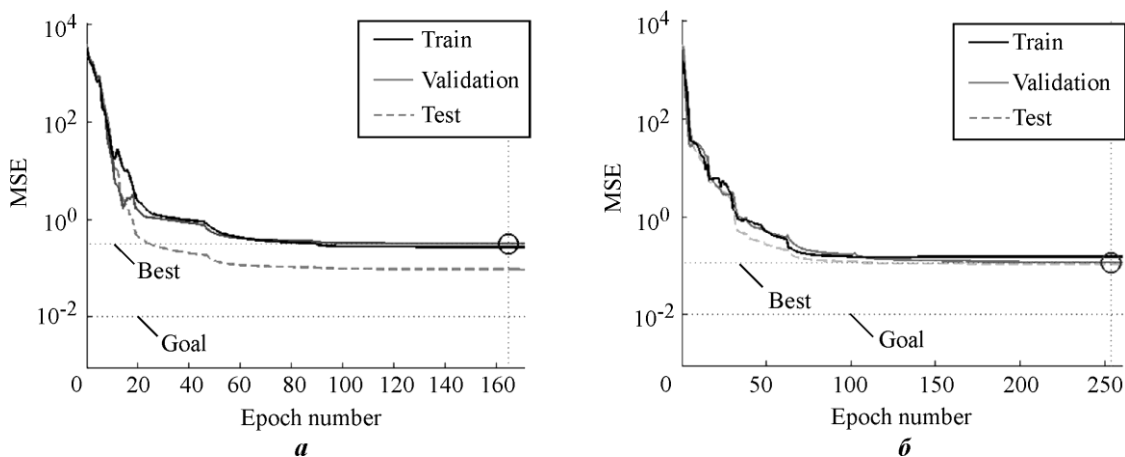


Рис.3. Результаты обучения нейронной сети типа MLP на данных, сформированных без использования кластеризации (а) и с применением алгоритма SOM (б): MSE – среднеквадратичная ошибка обучения; Epoch number – текущая эпоха; Train, Validation, Test – поведение ошибки для обучающего, проверочного и тестового множеств соответственно; Goal, Best – целевое и наилучшее значения ошибки (последнее достигнуто для проверочного множества)

Fig.3. Results for training the MLP neural network on the data generated without using of algorithm (a) and with using of SOM algorithm (b)

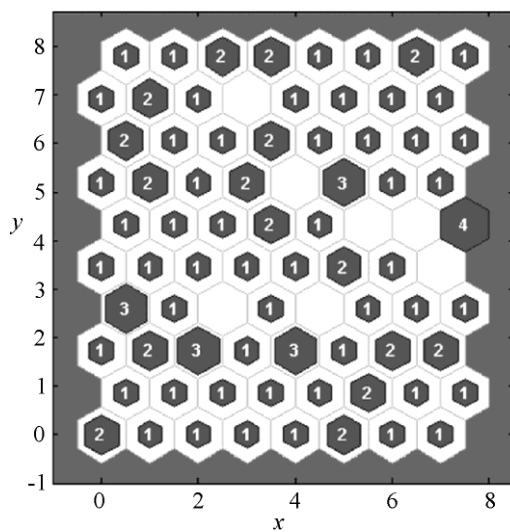


Рис.4. Разбиение факторного пространства с применением самоорганизующихся карт Кохонена

Fig.4. Distribution of the factor space, constructed by Kohonen's network

Обучение второй нейронной сети проводилось с применением рассматриваемого алгоритма. Результаты приведены на рис.3,б.

На этапе 1 – вычислены выборочные константы Липшица элементов факторного пространства по формуле (4). На этапе 2 кластеризация разбила факторное пространство на 73 кластера, 20 из которых содержат более одного элемента (рис.4). Очевидно, что для 53 единичных кластеров нет необходимости в вычислении минимальной константы Липшица и элементы из них должны быть включены в обучающее множество «как есть». На этапе 3 в соответствии с формулой (5) сформировано обучающее множество, в которое включены 73 элемента, имеющие минимальную константу Липшица в пределах каждого кластера. На этапе 4 по формуле (6) вычислена опорная точка

алгоритма. Для сформированного на этапе 3 обучающего множества ее значение равно 44,66. На этапе 5 проведена оценка характеристики нейрона по формуле (8). Для нейронной сети выбранной архитектуры $m = 2$, $n_0 = 4$, $n_1 = 4$, $n_2 = 1$, $L_{op} = 44,66$.

Таким образом, для нейронов сети выбранной архитектуры, а также сформированного с учетом оценки выборочных констант Липшица обучающего множества оценка характеристики нейрона $c \geq 2,362$. Данная оценка является оценкой снизу, поэтому ее значение можно округлить в большую сторону, например до первого знака (до 2,37), и взять за минимально допустимое значение характеристики нейрона.

В случае формирования обучающего множества из представителей каждого кластера (по одному) имеем равновероятные события выбора, т.е. вероятность такого события равна $\frac{1}{73}$. Используя формулу Шеннона, получаем энтропию, равную 6,19. Значение среднеквадратичной ошибки обучения составляет 0,116.

Заключение. Описанный алгоритм позволяет сформировать обучающее множество со значением энтропии больше, чем при формировании его случайным образом, а также дать оценку характеристики нейрона снизу и уточнить его вид. Применение алгоритма дает возможность добиться прироста энтропии обучающего множества на 0,14 бит и уменьшить среднеквадратичную ошибку обучения на 0,098 по сравнению с обучением аналогичной нейронной сети, для которой алгоритм не применялся.

За рамками настоящей работы остается вопрос оценки характеристики нейрона сверху, что позволило бы определить более точное значение. Вычисление оценки сверху является важной задачей и в настоящее время осуществляется опытным путем с учетом вида активационной функции. Тем не менее оценка характеристики нейрона снизу сужает множество возможных значений, которое необходимо перебрать опытным путем для решения задачи – определения вида нейрона многослойного персептрона.

Литература

1. **LeCun Y.** Efficient learning and secondary methods. – MA: MIT Press, 1993. – 71 p.
2. **Хайкин С.** Нейронные сети: пер. с англ. – М.: ИД «Вильямс», 2008. – 1104 с.
3. **Пастухов А.А.** Применение нейронных сетей для прогнозирования параметров энергетических установок с лазерным зажиганием // Научно-технические ведомости СПбГПУ. Физико-математические науки. – 2015. – № 2(218). – С. 19–29. DOI: 10.5862/JPM.218.2
4. **Зароченцев С.Г., Ковалев В.И., Пастухов А.А., Прокофьев А.А.** Использование нейронных сетей для построения прогностических моделей процессов в энергетических установках и их агрегатах // Изв. вузов. Электроника. – 2016. – Т. 21. – №3. – С. 247–253.
5. **Пастухов А.А., Прокофьев А.А.** Применение алгоритмов кластеризации к формированию представительской выборки для обучения многослойного персептрона // Научно-технические ведомости СПбГПУ. Физико-математические науки. – 2017. – Т. 10. – № 2. – С. 58–68.
6. **Пастухов А.А., Прокофьев А.А.** Применение самоорганизующихся карт Кохонена для формирования представительской выборки при обучении многослойного персептрона // Научно-технические ведомости СПбГПУ. Физико-математические науки. – 2016. – № 2 (242). – С. 95–107.
7. **Кохонен Т.** Самоорганизующиеся карты. – М.: Бином; Лаборатория знаний, 2008. – 655 с.
8. **Федоренко Ю.С., Гапанюк Ю.Е.** Кластеризация данных на основе самоорганизующихся растущих нейронных сетей и марковского алгоритма кластеризации // Нейрокомпьютеры: разработка, применение. – 2016. – № 4. – С. 3–13.
9. **Царегородцев В.Г.** Предобработка обучающей выборки, выборочная константа Липшица и свойства обученных нейронных сетей: материалы X Всерос. семинара «Нейроинформатика и ее приложения». – Красноярск: КГТУ, 2002. – С. 146–150.
10. **Асеев Г.Г.** Методы интеллектуального анализа данных в электронных хранилищах // Бионика интеллекта. – 2008. – № 1(70). – С. 28–33.

Поступила 26.01.2018 г.; принята к публикации 19.06.2018 г.

Пастухов Алексей Андреевич – аспирант кафедры высшей математики-1 Национального исследовательского университета «МИЭТ» (124498, г. Москва, г. Зеленоград, пл. Шокина, д. 1.), pastuhov1992@gmail.com

Прокофьев Александр Александрович – кандидат физико-математических наук, доктор педагогических наук, заведующий кафедрой высшей математики-1 Национального исследовательского университета «МИЭТ» (124498, г. Москва, г. Зеленоград, пл. Шокина, д. 1.), aaprokof@yandex.ru

References

1. LeCun Y. *Efficient learning and secondary methods*. MA, MIT Press, 1993. 71 p. (In Russian).
2. Haykin S.O. *Neural networks: a comprehensive foundation*. 2nd ed. Moscow, ID Vil'yams Publ., 2008. 1104 p. (In Russian).
3. Pastukhov, A.A. Predicting the parameters of energy installations with laser ignition: neural network models. *Nauchno-tekhnicheskiye vedomosti SPbGPU. Fiziko-matematicheskiye nauki = St. Petersburg State Polytechnical University Journal. Physics and Mathematics*, 2015, no. 2(218), pp. 19–29. DOI 10.5862/JPM.218.2. (In Russian).
4. Zarochentsev S.G., Kovalev V.I., Pastukhov A.A., Prokofiev A.A. Simulation modeling using neural network to create prediction models of processes in liquid-propellant rockets and their aggregates. *Izvestiya vuzov. Elektronika = Proceedings of Universities. Electronics*, 2016, vol. 21, no. 3, pp. 247–253. (In Russian).
5. Pastukhov A.A., Prokofiev A.A. Clustering algorithms application to the representative sample formation in the training of the multilayer perceptron. *Nauchno-tekhnicheskiye vedomosti SPbGPU. Fiziko-matematicheskiye nauki = St. Petersburg State Polytechnical University Journal. Physics and Mathematics*, 2017, vol. 10, no. 2, pp. 58–68. DOI: 10.18721/JPM.102-201706. (In Russian).
6. Pastukhov, A.A., Prokofiev, A.A. Kohonen self-organizing map application to representative sample formation in the training of the multilayer perceptron. *Nauchno-tekhnicheskiye vedomosti SPbGPU. Fiziko-matematicheskiye nauki = St. Petersburg State Polytechnical University Journal. Physics and Mathematics*, 2016, no. 2(242)2016, pp. 95–107. DOI: 10.5862/JPM.242.11. (In Russian).

7. Kohonen T. *Self-organized formation of topologically correct feature maps*. Moscow, Binom. Laboratoriya znaniy Publ., 2008. 655 p. DOI:10.1007/bf00337288 (In Russian).
8. Fedorenko Yu.S., Gapanyuk Yu.E. Clustering of data on the basis of self-organizing growing neural networks and the Markov algorithm of clustering. *Neyrokomп'yutery: razrabotka, primeneniye = Neurocomputers: development and application*, 2016, no. 4, pp. 3–13. (In Russian).
9. Tsaregorodtsev V.G. Preprocessing the training sample, the selective Lipschitz constant, and the properties of the trained neural networks. *Materials of the 10th All-Russian Seminar «Neuroinformatics and its applications»*. Krasnoyarsk, KSTU Publ., 2002, pp. 146–150. (In Russian).
10. Aseev G.G. Methods of data mining in electronic storages. *Bionika intellektal = Bionics of the intellect*, 2008, no. 1 (70), pp. 28–33. (In Russian).

Submitted 26.01.2018; Accepted 19.06.2018.

Information about the authors:

Aleksey A. Pastukhov – PhD student of the Higher Mathematics-1 Department, National Research University of Electronic Technology (Russia, 124498, Moscow, Zelenograd, Shokin sq., 1), pastuhov1992@gmail.com

Aleksander A. Prokofiev – Cand. Sci. (Phys.-Math.), Dr. Sci. (Ped.), Head of the Higher Mathematics-1 Department, National Research University of Electronic Technology (Russia, 124498, Moscow, Zelenograd, Shokin sq., 1), aaprokof@yandex.ru