

УДК 681.3

Н.В. Мурашкин, В.Н. Кострова  
**ПРОБЛЕМЫ ИСПОЛЬЗОВАНИЯ ИСКУССТВЕННЫХ  
НЕЙРОННЫХ СЕТЕЙ ДЛЯ РЕШЕНИЯ ЗАДАЧ БИНАРНОЙ  
КЛАССИФИКАЦИИ**

*Воронежский институт высоких технологий  
Воронежский государственный технический университет*

*Статья направлена на выявление проблем использования искусственных нейронных сетей для решения задач бинарной классификации. Для решения задач бинарной классификации предполагается отнесение образцов, которые уже имеются, к определенным классам. Ведущим подходом к исследованию данной проблемы является алгоритм Левенберга-Марквардта, позволяющий проводить оптимизацию параметров нелинейных регрессионных моделей. В качестве критерия оптимизации принимается среднеквадратичная ошибка модели на обучающей выборке. Предлагается для ускорения вычислений в несколько раз применять метод эластичного распространения, в работе указаны основные шаги алгоритма, построенные на основе этого метода. Материалы статьи представляют практическую ценность для специалистов, использующих искусственные нейронные сети для задач классификации.*

**Ключевые слова:** бинарные задачи, искусственные нейронные сети, алгоритм Левенберга-Марквардта, алгоритм Гаусса-Ньютона, метод М. Ридмиллера и Г. Брауна.

**Введение.** Классификация представляет собой наиболее распространенную и часто решаемую задачу в разных практических приложениях. При решении задачи классификации подразумевают, что рассматриваемые предметы или явления распределяются по категориям, основываясь на некоторых признаках, которые демонстрируют природу анализируемых явлений, и их располагают в некотором порядке, который отражает степень сходства между ними [1-3].

В процессе решения задач классификации одним из возможных подходов может быть использование нейронных сетей [4, 5]. В результате достигается линейная разделимость множества классифицируемых объектов, при этом значительно упрощается построение классификатора. Но для решения задач с малым количеством образцов в обучающей выборке, не всегда представляется возможность для обработки исходных данных таким образом, чтобы была достигнута линейная разделимость объектов, подлежащих классификации.

В рамках аппарата искусственных нейронных сетей есть возможность для аппроксимации функций, что позволяет выстраивать разделяющие поверхности большей сложности и способствует эффективно проведению классификации. При этом нейронной сетью генерируется большое количество регрессионных моделей, базирующихся, при непосредственном решении задачи, на статистических методах

стандартных средств классификации. В данной работе даются предложения по построению алгоритма, позволяющего ускорить обработку данных.

**Анализ существующих подходов.** Для решения задач бинарной классификации предполагается отнесение образцов, которые уже имеются, к определенным классам. Некоторому образцу проводится сопоставление определенного признакового описания, то есть, вектор, компонентами которого являются разнообразные количественные и качественные характеристики, о которых было сказано ранее. Совокупность компонентов данного вектора определяется размерностью своего рода «пространства входов», которое уже в свою очередь разбивается на классы. К задаче алгоритма классификации можно отнести произвольный объект, который принадлежит к одному из классов. Именно поэтому искусственные нейронные сети получили широкое распространение в данной области.

Одним из главных недостатков указанных подходов можно считать очень долгий процесс обучения, что не позволяет его использовать для ряда задач, на которые нужно быстро реагировать. На сегодняшний день используют алгоритмы, позволяющие существенно ускорить обучающий процесс, к ним относятся: метод сопряженных градиентов, алгоритм Левенберга-Марквардта и другие.

Алгоритм Левенберга-Марквардта [6] предназначается для оптимизации параметров нелинейных регрессионных моделей. Предположим, в качестве критерия оптимизации принимается среднеквадратичная ошибка модели на обучающей выборке. Алгоритм будет заключаться в последовательном приближении начальных заданных значений параметров к локальному оптимуму, который является искомым.

Задача обучающей выборки состоит из множества пар свободных переменных  $x \in X^M$ , которые являются входами сети, и зависимых переменных  $y \in Y^M$ . Так же будет задана функциональная зависимость, которая представляет собой регрессионную модель  $y = f(w, x_n)$ , которая, в свою очередь, непрерывно дифференцируется в области  $W \times X$ . Параметр  $w$  является вектором весовых коэффициентов. Необходимо найти определенное значение вектора  $w$ , который бы был связан с локальным минимумом функции ошибки (1)

$$E_D = \sum_{n=1}^N (y_n - f(w, x_n))^2 \quad (1)$$

При старте алгоритма необходимо задать начальный весовой вектор коэффициентов  $w$ . Для каждого шага итерации идет замена вектора на вектор  $w + \Delta w$ .

Для оценивания прироста  $\Delta w$  необходимо использовать подход, связанный с линейным приближением функции (2).

$$f(w + \Delta w, x) \approx f(w, x) + J \Delta w, \quad (2)$$

где  $J$  – является якобианом функции  $f(w, x_n)$  по точке  $w$ . Наглядное представление матрицы  $J$  можно сделать в виде (3).

$$J = \begin{bmatrix} \frac{\partial f(w, x_1)}{\partial w_1} & \dots & \frac{\partial f(w, x_1)}{\partial w_R} \\ \vdots & \ddots & \vdots \\ \frac{\partial f(w, x_N)}{\partial w_1} & \dots & \frac{\partial f(w, x_N)}{\partial w_R} \end{bmatrix}. \quad (3)$$

В (3) весовой вектор коэффициентов  $w = [w_1, \dots, w_R]^T$ .

Приращение  $\Delta w$  в точке  $w$ , ведет к минимуму  $E_D$ , равному 0. Следовательно, для того чтобы найти последующее приращение  $\Delta w$  требуется приравнять к нулю вектор частных производных  $E_D$  по  $w$  (4).

$$E_D = |y - f(w + \Delta w)|^2, \quad (4)$$

где  $f(w + \Delta w) = [f(w + \Delta w, x_1), \dots, f(w + \Delta w, x_N)]^T$ .

Преобразовывая и дифференцируя это выражение получаем (5)

$$\begin{aligned} |y - f(w + \Delta w)|^2 &= (y - f(w + \Delta w))^T (y - f(w + \Delta w)) = \\ &= f^T(w + \Delta w) f(w) - 2y^T f(w + \Delta w) + y^T y \end{aligned} \quad (5)$$

В результате получим

$$\frac{\partial E_D}{\partial w} = (J^T J) \Delta w - J^T (y - f(w)) = 0 \quad (6)$$

Поэтому для того чтобы возможно было найти значение  $\Delta w$  следует провести решение системы линейных уравнений (7).

$$\Delta w = (J^T J)^{-1} J^T (y - f(w)). \quad (7)$$

В связи с тем, что число обусловленности матрицы  $J^T J$  является квадратом числа обусловленности матрицы  $J$ , то сама матрица  $J^T J$  оказывается вырожденной достаточно существенно. Поэтому параметр регуляризации  $\lambda \geq 0$  ввел Марквардт:

$$\Delta w = (J^T J + \lambda I)^{-1} J^T (y - f(w)), \quad (8)$$

где  $I$  – единичная матрица. Данный параметр получается из каждой последующей итерации алгоритма. Если быстро убывает значение ошибки  $E_D$ , тогда малое значение  $\lambda$  сводится к тому, что данный алгоритм приравнивается к алгоритму Гаусса-Ньютона.

Если приращение  $\Delta w$  в каждой последующей итерации меньше заданного значения, либо весовой вектор коэффициентов доставляет ошибку  $E_D$ , которая меньше заданной величины, тогда происходит остановка алгоритма, или если количество циклов обучения искусственной нейронной сети закончилось, также наступает остановка алгоритма. Значение величины вектора  $w$  считается искомым на последней итерации.

**Предложения по разработке алгоритма.** Укажем алгоритм, способный дать преимущество в 5 раз, если сравнивать со способом обратного распространения ошибки с точки зрения того, какое время сходимости. Таким алгоритмом является эластичное распространение (resilient propagation). Этот подход был предложен М. Ридмиллером и Г. Брауном [7].

Отличие указанного алгоритма от того, какие характеристики в алгоритме обратного распространения состоит в том, что данный алгоритм может использовать знаки частных производных для модификации весовых коэффициентов. При этом обучение проводится по так называемым «эрам», каждой из которых ставится в соответствие представление всех объектов из сети обучающей выборки.

Величину коррекции весов необходимо рассчитывать на основе такой формулы (9):

$$\Delta_{ij}^k = \begin{cases} \eta^+ \Delta_{ij}^k \frac{\partial E^k}{\partial w_{ij}} \frac{\partial E^{k-1}}{\partial w_{ij}} > 0 \\ \eta^- \Delta_{ij}^k \frac{\partial E^k}{\partial w_{ij}} \frac{\partial E^{k-1}}{\partial w_{ij}} < 0 \end{cases} \quad (9)$$

$$0 < \eta^- < 1 < \eta^+ \quad (9)$$

То, что меняется знак частной производной по этому шагу для соответствующего веса, указывает на то, что локальный минимум был упущен и величину изменения необходимо уменьшить на  $\eta$ , а значение веса сделать равным предыдущему (10):

$$\Delta w_{ij}(k) = \Delta w_{ij}(k) - \Delta_{ij}^{k-1} \quad (10)$$

Но если неизменным остался знак частной производной, то требуется прибавить коррекцию на  $\eta+$ , с тем, чтобы была увеличена скорость сходимости. В качестве достоинства этого алгоритма, если сравнивать с характеристиками алгоритма обратного распространения ошибки мы можем считать то, что когда фиксируется  $\eta+$  и  $\eta-$  есть возможность отказаться от проведения настроек нейронной сети в глобальном смысле. Указаниями значений для  $\eta+$  и  $\eta-$  будут являться 1,2 и 0,5 соответственно, но, так же можно допустить использование разных значений.

Проведение вычисления значений коррекции весов идет таким образом (11):

$$\Delta w_{ij}(k) = \begin{cases} +\Delta_{ij}^k, \frac{\partial E^k}{\partial w_{ij}} > 0 \\ -\Delta_{ij}^k, \frac{\partial E^k}{\partial w_{ij}} < 0 \\ 0, \frac{\partial E^k}{\partial w_{ij}} = 0 \end{cases} \quad (11)$$

В том случае, если у производной есть положительный знак, считается что происходит возрастание ошибки, и при этом идет уменьшение весового коэффициента на значение коррекции или, для противоположного случая, осуществляют увеличение.

Подстройку весов осуществляем на основе формулы (12):

$$w_{ij}(k+1) = w_{ij}(k) + \Delta w_{ij}(k) \quad (12)$$

Если данный алгоритм рассматривать по шагам, то шаги будут следующими:

- 1) Устанавливаются начальные значения  $\Delta_{ij}$ ;
- 2) Предъявляются сети все образцы и вычисляются частные производные;

- 3) Подсчитываются новые значения  $\Delta w_{ij}$ ;
- 4) Корректируются веса;
- 5) Повторяется процедура, начиная с п.2

Важно указать присущие для количественных методов специфические достоинства и недостатки [8-11]. Проведение подбора по всем параметрам искусственных нейронных сетей происходит экспериментальным образом, так как возникновение проблем при процессах обучения трудно предсказать заранее. Это представляет собой основной недостаток их применения, помимо отсутствия возможностей проведения интерпретирования результатов классификации, то есть когда устанавливается, для каких именно фактором мы можем говорить о решающем значении при определении классов, к которым принадлежат исследуемые объекты.

Но идет компенсация такого недостатка вследствие того, что искусственные нейронные сети исключительно пригодны для того, чтобы аппроксимировать нелинейные функции, которые характеризуются тем, что в них большое количество переменных, они рассматриваются как превосходный инструмент для того, чтобы решать задач классификации.

Для идеального случая вероятность, которая предсказывается моделью бинарного выбора, является близкой или к 0, или к 1. С точки зрения практики существуют случаи, когда неуверенная оценка, которая получается при использовании бинарной регрессии.

**Выводы.** Бинарная модель предоставляет возможности для того, чтобы определять, как влияет каждый фактор на результат. Недостаток модели состоит в том, что для некоторых случаев она может привести к неустойчивым и неуверенным оценкам. Такого недостатка нет в искусственных нейронных сетях, но их коэффициенты не всегда могут быть интерпретированы.

## ЛИТЕРАТУРА

1. Чопоров О.Н. Методы анализа значимости показателей при классификационном и прогностическом моделировании / О.Н.Чопоров, А.Н.Чупеев, С.Ю.Брегеда // Вестник Воронежского государственного технического университета. 2008. Т. 4. № 9. С. 92-94.
2. Преображенский Ю.П. Оценка эффективности применения системы интеллектуальной поддержки принятия решений / Ю.П. Преображенский // Вестник Воронежского института высоких технологий. 2009. № 5. С. 116-119.
3. Чопоров О.Н. Методика преобразования качественных характеристик в численные оценки при обработке результатов медико-социального

- исследования / О.Н.Чопоров, А.И.Агарков, Л.А.Куташова, Е.Ю. Коновалова // Вестник Воронежского института высоких технологий. 2012. № 9. С. 96-98.
4. Круглов В. В. Искусственные нейронные сети. Теория и практика. - 2-е изд., стереотип / В. В. Круглов, В. В. Борисов. - М. : Горячая линия-Телеком, 2002. - 382 с.
  5. Бодянский Е. В. Искусственные нейронные сети: архитектуры, обучение, применения: монография / Е. В. Бодянский, О.Г. Руденко. - Харьков: Телетех, 2004.-369 с.
  6. Gavin, H. P. The Levenberg-Marquardt method for nonlinear least squares curve-fitting problems : : [electronic resource] / H. P. Gavin // Department of Civil and Environmental Engineering Duke University. - Access mode: <http://people.duke.edu/~hpgavin/lm.pdf> (date of request: 20.05.2017).
  7. Riedmiller M. A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm : [electronic resource] / M. Riedmiller, H. Braun // University of Karlsruhe. - Access mode: <http://deeplearning.cs.cmu.edu/pdfs/Rprop.pdf> (Дата доступа: 20.05.2017).
  8. Чопоров О.Н. Методика формирования информационной базы данных для проведения многоуровневого мониторинга и классификационно-прогностического моделирования / О.Н. Чопоров, О.В. Золотухин, И.И. Манакин, С.В. Болгов // Вестник Воронежского института высоких технологий. - 2015. - № 14. - С. 19-24.
  9. Чернов А.В. Разработка классификационно-прогностических моделей развития гнойно-септических осложнений у родильниц / А.В. Чернов, В.Ю. Бригадирова, О.Н. Чопоров, В.И. Чернов // Системный анализ и управление в биомедицинских системах. - 2012. - Т. 11. - № 1. - С. 261-266.
  10. Бабкин А.П. Разработка процедур оценки выраженности диабетической ретинопатии у больных сахарным диабетом на основе дискриминантного анализа / А.П. Бабкин, В.Г. Мединцев, А.Г. Черноусенко, О.Н. Чопоров // Вестник Воронежского государственного технического университета. - 2005. - Т. 1. - № 7. - С. 97-99.
  11. Choporov O. Technique of information database formation for carrying out multilevel monitoring and classificatory-and-forecasting modeling /O. Choporov, A. Kurotova, I. Manakin // Information Technology Applications. - 2015. - № 1. - С. 111-123.

N. I. Murashkin, V. N. Kostrova  
**THE SHORTCOMINGS IN THE USE OF ARTIFICIAL NEURAL  
NETWORKS FOR SOLVING PROBLEMS OF BINARY  
CLASSIFICATION**

*Voronezh Institute of high technologies  
Voronezh state technical university*

*The article aims at identifying problems of the use of artificial neural networks for solving problems of binary classification. For solving problems of binary classification it is expected to classify samples that are already available to a certain class. A leading approach to the study of this problem is the algorithm of Levenberg-Marquardt, which allows to optimize the parameters of nonlinear regression models. As optimization criterion is adopted to the root mean square error of the model on the training set. It is proposed to accelerate the computation to apply the method to the elastic distribution. The materials of the article are of practical value to professionals who use artificial neural networks for classification tasks.*

**Keywords:** binary tasks, artificial neural networks, algorithm, algorithm of Levenberg-Marquardt, algorithm of Gauss-Newton, method of M. Riedmiller and G. Brown.

#### REFERENCES

1. Choporov O.N. Metody analiza znachimosti pokazateley pri klassifikatsionnom i prognosticheskom modelirovanii / O.N.Choporov, A.N.Chupeev, S.Yu.Bregeda // Vestnik Voronezhskogo gosudarstvennogo tekhnicheskogo universiteta. 2008. Vol. 4. No. 9. pp. 92-94.
2. Preobrazhenskiy Yu.P. Otsenka effektivnosti primeneniya sistemy intellektual'noy podderzhki prinyatiya resheniy / Yu.P. Preobrazhenskiy // Vestnik Voronezhskogo instituta vysokikh tekhnologiy. 2009. No. 5. pp. 116-119.
3. Choporov O.N. Metodika preobrazovaniya kachestvennykh kharakteristik v chislennye otsenki pri obrabotke rezul'tatov mediko-sotsial'nogo issledovaniya / O.N.Choporov, A.I.Agarkov, L.A.Kutashova, E.Yu. Konovalova // Vestnik Voronezhskogo instituta vysokikh tekhnologiy. 2012. No. 9. pp. 96-98.
4. Kruglov V. V. Iskusstvennye neyronnye seti. Teoriya i praktika. - 2-e izd., stereotip / V. V. Kruglov, V. V. Borisov. - M. : Goryachaya liniya-Telekom, 2002. – p. 382 .
5. Bodyanskiy E. V. Iskusstvennye neyronnye seti: arkhitektury, obuchenie, primeneniya: monografiya / E. V. Bodyanskiy, O.G. Rudenko. - Khar'kov: Teletekh,2004.- p. 369.
6. Gavin, H. P. The Levenberg-Marquardt method for nonlinear least squares curve-fitting problems : : [electronic resource] / H. P. Gavin // Department of Civil and Environmental Engineering Duke University. - Access mode: <http://people.duke.edu/~hpgavin/lm.pdf> (date of request: 20.05.2017).



7. Riedmiller M. A Direct Adaptive Method for Faster Backpropagation Learning: The RPROP Algorithm : [electronic resource] / M. Riedmiller, H. Braun // University of Karlsruhe. - Access mode: <http://deeplearning.cs.cmu.edu/pdfs/Rprop.pdf> (Data dostupa: 20.05.2017).
8. Choporov O. N. The technique of formation of information database for multi-level monitoring and classification-predictive modeling / O. N. Choporov, O. V. Zolotukhin, I. Manakin, Bolgov S. V. // Vestnik of Voronezh Institute of high technologies. - 2015. - No. 14. - S. 19-24.9.
9. Chernov A.V. the Development of a taxonomy of predictive models the development of purulent-septic complications in puerperants / A.V. Chernov, V. Y. Brigadirovka, O. N. Choporov, and V. I. Chernov // System analysis and management in biomedical systems. - 2012. - T. 11. - No. 1. - P. 261-266.10.
10. Babkin A. P. the Development of procedures for evaluating the severity of diabetic retinopathy in diabetic patients based on discriminant analysis / A. P. Babkin, V. G. Medentsev, A. G., Eng, O. N. Choporov // Bulletin of Voronezh state technical University. - 2005. - Vol. 1. - No. 7. - P. 97-99.11.
11. Choporov O. Technique of information database formation for carrying out multilevel monitoring and classificatory-and-forecasting modeling / Choporov O., A. Kurotova, I. Manakin // Information Technology Applications. - 2015. - No. 1. - P. 111-123.