

Сорокин А. С., Бондарев В. Ю., Кротова Е. Л.

СОЗДАНИЕ И ОБУЧЕНИЕ ИСКУССТВЕННОЙ НЕЙРОННОЙ СЕТИ ДЛЯ СТАТИСТИЧЕСКОГО ОЦЕНИВАНИЯ ДАННЫХ

В данной работе представлено использование искусственной нейронной сети для оценки предоставляемых услуг пользователям. Предлагаемая нами искусственная вычислительная модель решает задачи статистической классификации. Данная сеть позволяет сравнить полученные данные с поставленной целью, т. е. строятся границы для значений параметров. В работе описывается подготовка данных для ввода их в сеть, создание самой сети и ее обучение, а также проверка на правильность обучения и нахождение искажений. Также говорится о применении сети к другим независимым параметрам. Венчает нашу статью знакомство с искусственной нейронной сетью.

Ключевые слова: искусственная нейронная сеть, статистическая классификация, статистическое оценивание, применение нейронной сети.

Sorokin A. S., Bondarev V. Yu., Krotova E. L.

CREATING AND TRAINING ARTIFICIAL NEURAL NETWORK FOR STATISTICAL DATA EVALUATION

This paper presents the use of artificial neural networks to assess the services provided to users. Our proposed artificial computational model solves the problems of statistical classification. This network allows you to compare the data with the intended purpose, i.e. constructed boundary parameter values. The paper describes how to prepare data for input into the network, the establishment of the network itself and its training, as well as checking for proper training and finding distortion. Also it refers to the use of the network to the other independent parameters. Crowned our article familiarity with the artificial neural network.

Keywords: artificial neural network, statistical classification, statistical estimation, the use of a neural network.

Искусственные нейронные сети стали часто использовать для решения актуальных задач, таких как прогнозирование, классификация, оценивание и управление. В нашей статье мы используем созданную нами нейронную сеть для оценки данных, после чего

проанализируем полученный результат. Что же такое нейронные сети? Нейронные сети – это исключительно мощный метод имитации процессов и явлений, позволяющий воспроизводить чрезвычайно сложные зависимости [1]. Они представляют собой распределен-

ные и параллельные системы, способные к адаптивному обучению путем анализа положительных и отрицательных воздействий. Элементарным преобразователем в данных сетях является искусственный нейрон [2]. Проще говоря, имитируя модели сети нейронов, мы можем найти связь между входным вектором данных и выходным. Для нахождения этой зависимости мы не будем программировать сеть, а обучим её, после чего можно выявить искаженность других данных к обученным данным по нейронной сети.

Разработка искусственной нейронной сети в пакете MATLAB

Наша задача заключается в создании и обучении нейронной сети для оценки параметров пользователей. Сначала об источнике данных: есть данные shareman, где отображено количество визитов посетителей, самих посетителей будем трактовать как одно число, слепленное из вектора бинарных чисел, т. е. 0 и 1. Получается отображение $f(x) \rightarrow R$, где x – эта «сигнатура» посетителя, R – закодированный результат его посещения: 0 – не задержится больше чем на 15 секунд, 1 – задержится дольше.

И главная цель – это натренировать/выявить то самое $f()$, чтобы при новом x было с «хорошей» вероятностью и попадало в правильное R : «произведет целевое действие» (1) или «не произведет» (0). Т. е. в совсем редуцированном виде R тоже сопоставляем 0 или 1. На этих данных мы и будем обучать сеть.

```
001010111111101011011001 -> 1
101101010101010010010010 -> 0
001010100001000011010001 -> 0
.....
```

Возьмем часть данных shareman, где их количество составляет около 28–29 тысяч значений, в нашем случае 10000. Но так как «сигнатура» посетителя (входные данные) огромные, нужно их подготовить для нейрон-

ной сети. Берем первые 30 значений двоичного числа, и находим корреляцию, т. е. статистическую взаимосвязь между входными значениями и выходными. Для этого нам хорошо подойдет коэффициент Пирсона, где значения коэффициента больше, те значения мы и возьмем на вход.

В программе MATLAB вызовем графический интерфейс пользователей для управления сетями и данными (nntool). Подаем входные и выходные данные в нашу сеть, и после чего наблюдаем, что сеть мы обучили за 2 секунды и 11 итераций.

Ниже на рисунке показаны на оси абсцисс номер итерации, а на оси ординат – степень средней квадратичной ошибки.

На графике мы наблюдаем, что степень средней квадратичной ошибки маленькая, и наилучшая проверка представлена 0,09 на восьмой итерации.

Проверка созданной искусственной нейронной сети для пригодности к анализу других параметров. Выявление искаженности

Теперь проверим, правильно ли обучилась наша сеть, для этого мы возьмем оставшуюся часть данных shareman, это около 18–19 тысяч, и просимулируем их. Она должна нам показать, что получится на выходе, после этого мы сравниваем их с целевыми значениями и выявляем количество ошибок.

Нейронная сеть будет пригодна для анализа данных, если коэффициент искажения будет не более 25%. Чтобы найти его, найдем отношение количества ошибок на количество всего значений. После обработки полученных данных коэффициент искажений составляет 21,39%. Значит, сеть обучилась правильно.

Воспользуемся нашей созданной сетью для анализа других данных. У нас есть данные посещения сайта steklodom.com. Сделаем все те же операции подготовки данных для нейронной сети, а именно: возьмем 30 первых

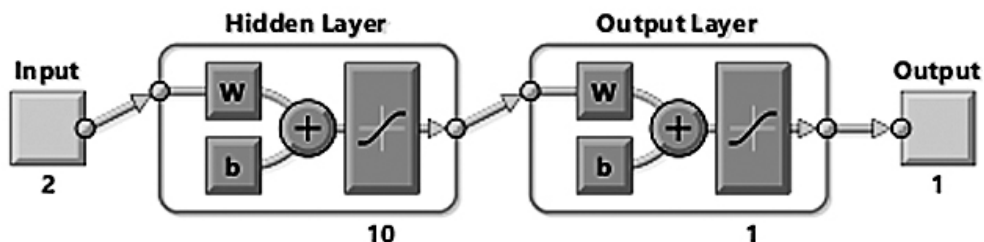


Рис. 1. Искусственная нейронная сеть

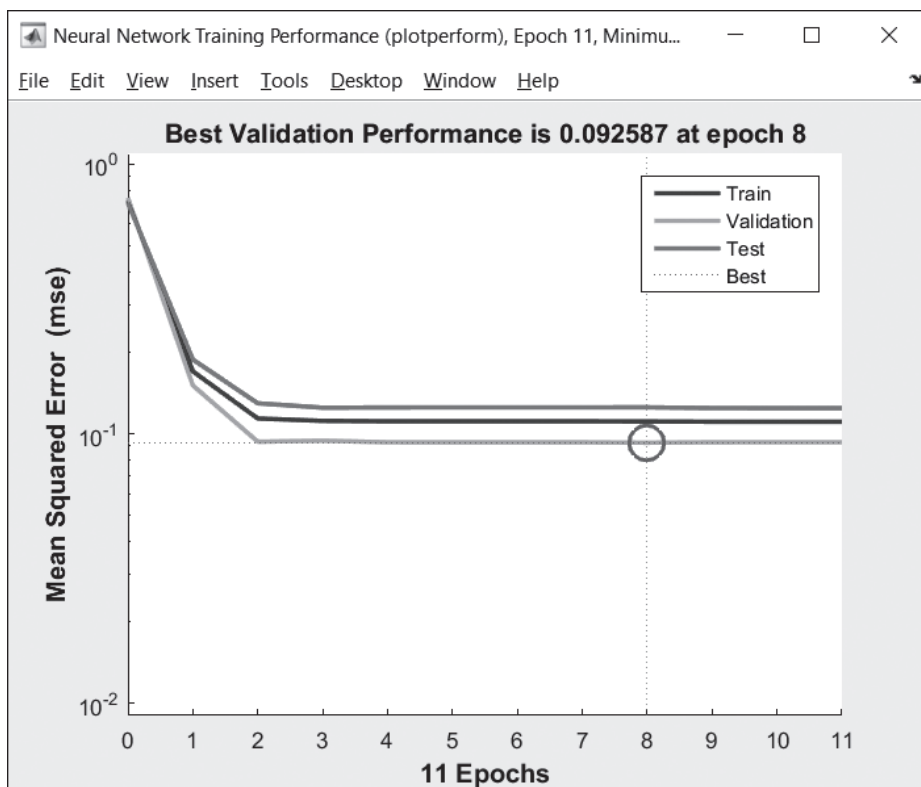


Рис. 2. Представление обучения нейронной сети

значений каждого бинарного числа, найдем наибольший коэффициент Пирсона, чтобы взять найденный вектор на вход. Далее про симулируем их и сравним с целевыми значениями. Нашли коэффициент искажения 36,27%. Значит, для этих данных целевые значения ставятся по-другому, т. е. наша сеть не подойдет для статистической классификации этих данных, так как надо, чтобы он составлял не более 25%.

Протестируем еще один сайт – omsk.dom.ru. Его коэффициент искажения составляет 39,52%. Наглядно видим, что нейронная сеть

не предназначена для классификации всех данных, а только для некоторых из них.

Заключение

В рассмотренной статье мы создали и обучили искусственную нейронную сеть. Также она была рассмотрена как один из способов классификации данных. Как мы уже говорили выше, после симулирования было выявлено, что не все данные подходят для нашей созданной сети. Нужно иметь другую сеть, обученную уже по этим данным, чтобы произвести их статистическое оценивание.

Примечания

1. Медведев В. С., Потемкин В. Г. Нейронные сети. MATLAB 6/ Под общ. ред. к. т. н. В. Г. Потемкина. — М. : ДИАЛОГ-МИФИ, 2002.— С. 496.
 2. Круглов В. В., Борисов В. В. Искусственные нейронные сети. Теория и практика. – 2-е изд., стереотип. — М. :Горячая линия - Телеком, 2002.— С. 382.
 3. Krotov I. N., Krotova E. L., Bogdanov N. V. Identification and counteractions to attacks of malefactors in the automated working system. – 2016.
-

Сорокин Андрей Станиславович, студент по направлению «Информационная безопасность» ПНИПУ. 614990, Пермский край, г. Пермь, ГСП, Комсомольский проспект, д. 29. E-mail: sly-kyper@yandex.ru

Бондарев Владислав Юрьевич, студент по направлению «Информационная безопасность» ПНИПУ. 614990, Пермский край, г. Пермь, ГСП, Комсомольский проспект, д. 29. E-mail: mr.bond1995@mail.ru

Кротова Елена Львовна, кандидат физико-математических наук, доцент кафедры высшей математики ПНИПУ. 614990, Пермский край, г. Пермь, ГСП, Комсомольский проспект, д. 29. E-mail: lenkakrotova@yandex.ru

Sorokin Andrew Stanislavovich, student in the direction of «Information Security» PNIPU, 29, Komsomolsky prospect, Perm, 614990. E-mail: sly-kyper@yandex.ru.

Bondarev Vladislav Yuryevich, student in the direction of «Information Security» PNIPU, 29, Komsomolsky prospect, Perm, 614990. E-mail: mr.bond1995@mail.ru

Krotova Elena Lvovna, Candidate of Physical and Mathematical Sciences, Associate Professor of the Department of Higher Mathematics PNIPU, 29, Komsomolsky prospect, Perm , 614990. E-mail: lenkakrotova@yandex.ru