

Для доказательства адекватности модели следует сравнить критерий Фишера с табличным значением [6]. Табличное значение критерия Фишера при доверительной вероятности 0,95, $\nu_1 = k = 6$, $\nu_2 = n - k - 1 = 70$, составляет 2,23. Поскольку $F_{\text{расч}} =$

$= 14,72 > F_{\text{табл}} = 2,23$, то уравнение регрессии можно считать адекватным. Сравнение экспериментальных и смоделированных значений концентраций геосмина (с 2003 по 2014 г.) представлено на рис. 3.

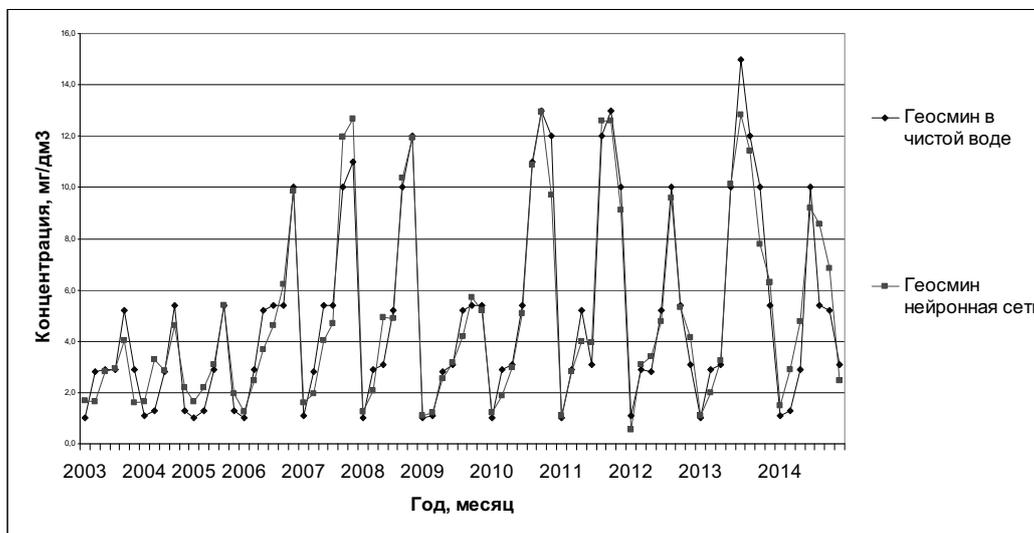


Рис. 3. Графики концентраций геосмина в питьевой воде (в период с 2003 по 2014 г.) для экспериментальных и смоделированных значений

Как видно из графика, результаты нейромоделирования хорошо аппроксимируют фактические данные. При этом общая квадратичная ошибка составляет всего 2,5 %. Значения на основе уравнения регрессии для смоделированных данных при помощи нейронной сети довольно точно описывают изменение концентрации геосмина в питьевой воде в период с 2003 по 2014 г. и могут быть использованы для разработки прогноза концентраций данного вещества в будущем, а также для подбора оптимальных параметров при дезодорации питьевой воды. Достоверность разработанной модели концентрации геосмина в питьевой воде и полученные результаты были подтверждены при помощи критерия Фишера.

Получено 18.12.15

Библиографические ссылки

1. Жолдакова З. И. Выявление причин образования запаха питьевой воды в г. Ижевске и разработка рекомендаций по его предотвращению на основании полученных данных : отчет о НИР (Заключ.) / ГУ НИИ ЭЧ и ГОС им. А. Н. Сысина, 2006. – С. 21–94.
2. Технологические отчеты водопроводного узла № 2 за 2003–2014 гг. // ВКХ СПВ «Пруд-Ижевск» ; Ижевск, 2003–2014. – С. 34–51.
3. Федотов В. Х. Нейронные сети в MS Excel / Чувашский гос. ун-т имени И. Н. Ульянова. – Чебоксары, 2004. – С. 35–61.
4. Там же.
5. Там же.
6. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике. – М. : Высш. шк., 2003. – С. 210–235.

УДК 004.85

Ю. В. Николаева, аспирант, ИжГТУ имени М. Т. Калашникова

МАТЕМАТИЧЕСКАЯ ПОСТАНОВКА ЗАДАЧИ ОБУЧЕНИЯ МНОГОСЛОЙНОГО ПЕРЦЕПТРОНА С ТОЧКИ ЗРЕНИЯ КЛАССИЧЕСКОЙ ОПТИМИЗАЦИИ

Парадигма нейронных сетей широко используется в решении практических задач в самых различных предметных областях – от

распознавания графических объектов в поисковых системах и системах автоматического проектирования до интеллектуального моделирования финансо-

вых стратегий на финансовых рынках. Подавляющее большинство информационно-управляющих систем, основанных на нейронных сетях, используют распределенные методы обучения нейросетей, такие как прямой и обратный алгоритмы распространения ошибки, не углубляясь в сам процесс обучения. Системы подобного рода используют нейронную сеть в качестве готового инструмента моделирования, оставляя в стороне математическую постановку задачи обучения нейронной сети, то есть фактически не учитывая применимость алгоритмов подбора оптимальных весов нейронов для конкретной архитектуры сети.

Автор статьи, не оспаривая высокие результаты применения широко используемого алгоритма распространения ошибок [1], ставит задачу рассмотре-

ния альтернативного варианта отыскания весов нейронов одной из популярных архитектур нейросетей для распознавания образов и классификации – многослойного перцептрона [2]. Также в статье опустим этапы подготовки тестовых данных, такие как консолидация и нормирование данных (подробнее см. 3, 4)), и будем считать, что данные для обучения сети уже прошли предобработку.

Для этого вернемся к математическим основам нейросетевой теории. Математическая модель нейронной сети основана на теореме Колмогорова (Колмогоров, 1957) которая утверждает, что любая непрерывная функция n аргументов на единичном кубе $[0,1]^n$ представима в виде суперпозиции непрерывных функций одного аргумента и операции сложения.

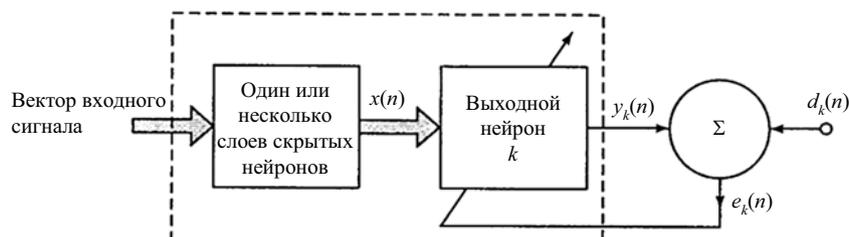


Рис. 1. Многослойная нейронная сеть прямого распространения

Выходы двухслойной нейросети (рис. 1) представляют собой суперпозицию функций одного аргумента с операциями сложения:

$$f(x^1, x^2, \dots, x^n) = \sum_{r=1}^{2n+1} h_k \left(\sum_{i=1}^n \varphi_{ik} (x^i) \right), \quad (1)$$

где h_k – функция активации нейронов скрытого слоя; x^i – входные нейроны с весами φ_{ik} .

Цель функционирования сети – найти оптимальные веса, с которыми функция наилучшим образом была бы приближена к реальным выходам нейронной сети из обучающей выборки:

$$e_k(n) = d_k(n) - y_k(n); \quad (2)$$

$$E(n) = \frac{1}{2} e_k^2(n) \rightarrow \min, \quad (3)$$

то есть в минимизации суммарной ошибки $E(n)$ выходов сети (где e_k – ошибка одного выхода сети).

Таким образом, функция выходов нейронной сети является универсальным механизмом моделирования широкого спектра задач [5].

Для нейросетей с одним выходным нейроном задача сводится к экстраполяции функции выхода сети. Рассмотрим простую структуру многослойного перцептрона с одним выходом.

Функция значения выходного нейрона перцептрона с одним скрытым слоем (рис. 2) принимает вид

$$y = h \sum_{j=1}^2 g_j \left(\sum_{i=1}^3 \varphi_{ij} \cdot x_i \right), \quad (4)$$

где h – функция активации выходного нейрона; g_j – функции активации нейронов скрытого слоя.

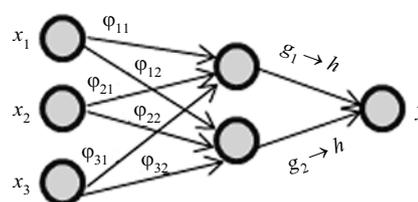


Рис. 2. Перцептрон с одним скрытым слоем

Процесс обучения нейронной сети с одним выходом сводится к нахождению таких весов, которые бы максимально приближали функцию значений выходного нейрона к реальному значению выхода из обучающей выборки:

$$E(n) = e^2 = (d - y)^2 = \left(d - h \sum_{j=1}^2 g_j \left(\sum_{i=1}^3 \varphi_{ij} \cdot x_i \right) \right)^2 \rightarrow \min. \quad (5)$$

В качестве обучающей выборки возьмем выборку для задачи классификации по 3 классам:

$$d = \begin{cases} 0, \\ 0,5, \\ 1. \end{cases}$$

Рассмотрим процесс обучения для первого класса ($d = 0$):

$$\left(h \sum_{j=1}^2 g_j \left(\sum_{i=1}^3 \varphi_{ij} \cdot x_i \right) \right)^2 \rightarrow \min.$$

В данном случае задача обучения сводится к минимизации функции выхода нейросети. Значения x_i – нормированные в интервал $[0,1]$ величины входов из обучающей выборки. Для рассмотрения задачи минимизации ошибки с точки зрения классической оптимизации необходимо определиться с функциями активации нейронов скрытого и выходного слоев.

В большинстве случаев функциями активации нейронов рекомендуется выбирать функции, которые нормируют значения суммы произведений весов предыдущего слоя в интервалы $[0,1]$ или $[-1,1]$. По сравнению с пороговыми, знаковыми или функциями насыщения (подробнее о функциях активации нейронов см. [6, 7]) непрерывные функции типа логистической (сигмоиды), гиперболического тангенса являются более гибкими для задач оптимизации, так как они равномерно сглаживают средневзвешенное значение нейронов предыдущего слоя. Чтобы исключить различия влияний на значение нейрона выходного слоя отдельных нейронов, функции активации всех нейронов должны совпадать.

Анализ начнем с выходного нейрона. Для этого введем обозначения: $t_j = \sum_{i=1}^3 \varphi_{ij} \cdot x_i$, тогда

$$y^2 = \left(h \sum_{j=1}^2 g_j \left(\sum_{i=1}^3 \varphi_{ij} \cdot x_i \right) \right)^2 = \left(h \sum_{j=1}^2 g_j (t_j) \right)^2, \text{ и если } K = \sum_{j=1}^2 g_j (t_j), \text{ то } y^2 = h^2 (K).$$

Возьмем в качестве функции активации выходного нейрона логистическую функцию $h(K) = \frac{1}{1 + e^{-K}}$. Тогда задача оптимизации сводится к отысканию минимума функции $\frac{1}{\left(1 + e^{-\sum_{j=1}^2 g_j(t_j)} \right)^2} \rightarrow \min.$

Функция вида $\frac{1}{(1+L)^2}$ принимает наименьшие

значения при максимизации аргумента L (рис. 3) при положительных значениях аргумента (так как в нашем случае аргументом является экспонента, которая принимает значения только больше 0).

Переходим к задаче максимизации экспоненциальной функции от аргумента функции активации скрытых нейронов (рис. 4):

$$e^{-\sum_{j=1}^2 g_j(t_j)} \rightarrow \max.$$

Функция монотонно возрастающая. Переходим к задаче максимизации ее аргумента:

$$-\sum_{j=1}^2 g_j(t_j) \rightarrow \max;$$

$$-g_1(t_1) - g_2(t_2) \rightarrow \max \Rightarrow g_1(t_1) + g_2(t_2) \rightarrow \min.$$

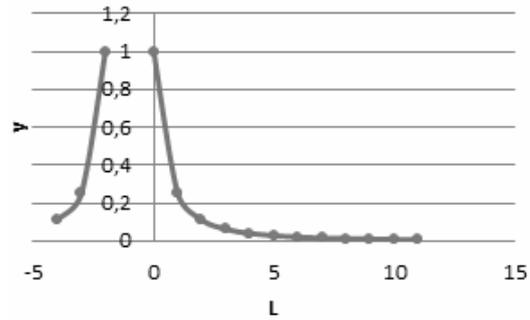


Рис. 3. График логистической функции активации

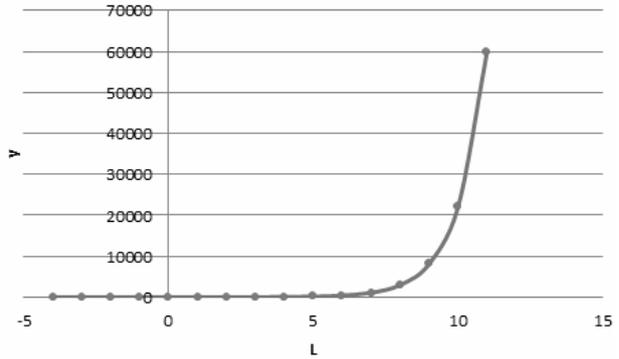


Рис. 4. График экспоненциальной функции активации

Для скрытого слоя берем такую же функцию активации, что и для выходного нейрона – логистическую, тогда

$$g_1(t_1) + g_2(t_2) \rightarrow \min \Rightarrow \frac{1}{1 + e^{-t_1}} + \frac{1}{1 + e^{-t_2}} \rightarrow \min.$$

За t_1 было взято средневзвешенное значение, попадающее на вход первого нейрона скрытого слоя, за t_2 – средневзвешенное, попадающее на вход второго скрытого нейрона:

$$t_1 = \varphi_{11}x_1 + \varphi_{21}x_2 + \varphi_{31}x_3; \quad t_2 = \varphi_{12}x_1 + \varphi_{22}x_2 + \varphi_{32}x_3.$$

Задача обучения трехслойного перцептрона с тремя входными, двумя промежуточными и одним выходным слоем сводится к задаче безусловной минимизации непрерывной функции 9 переменных (6 весов и 3 значений входных нейронов):

$$\frac{1}{1 + e^{-\varphi_{11}x_1 + \varphi_{21}x_2 + \varphi_{31}x_3}} + \frac{1}{1 + e^{\varphi_{12}x_1 + \varphi_{22}x_2 + \varphi_{32}x_3}} \rightarrow \min.$$

Задача безусловной оптимизации состоит в отыскании оптимального плана $(\varphi_{ij}, x_i)^*$ и может быть решена с помощью классических методов, таких как различные модификации метода градиентного спуска. Но если включать переменные x_i в базис, то будет найден один из допустимых планов, в котором значения входных нейронов будут фиксированными. А процесс обучения предусматривает нахождение таких весов, которые будут подстраиваться под значения входов нейронов. То есть значения входов сети являются управлениями над системой весов.

Таким образом, можно перейти к задаче оптимального управления, где вектор входных значений нейронной сети является вектором управления, вектор весов является вектором состояния, а минимизируемое выражение – критерием качества. На каждом последующем промежутке времени значения входов нейронов меняются, и полученные на предыдущем шаге веса влияют на веса нового шага.

$$\varphi(\tau+1) = A\varphi(\tau) + Bx(\tau), \quad (6)$$

где τ – переменная, характеризующая динамику процесса обучения, момент, за который подаются значения входов сети; A – матрица влияния весов предыдущего шага; B – вектор корректировки предыдущего управления;

$$\varphi(\tau) \text{ – вектор состояния;} \quad (7)$$

$$x(\tau) \text{ – вектор управления,}$$

$$x(\tau) \in X : \{x_i \in [0, 1], i = 1 \dots 3\}, \quad (8)$$

где X – множество допустимых управлений (множество отнормированных значений входов нейронной сети).

При рассмотрении процесса управления в динамике критерий качества выражается как сумма минимизируемых выражений:

$$J(\varphi, x) = \sum_{\tau=0}^T \left(\frac{1}{1 + e^{-\sum_{i=1}^3 \varphi_{1i} x_i(t)}} + \frac{1}{1 + e^{-\sum_{i=1}^3 \varphi_{2i} x_i(t)}} \right) \rightarrow \min. \quad (9)$$

Получена линейная дискретная задача оптимального управления с суммарным функционалом [8], которая является хорошо исследованной и имеет специальные методы оптимизации: принцип максимума Понтрягина и методы динамического программирования (метод Беллмана).

Описанным методом может быть проведено обучение многослойного перцептрона для задач классификации. Структура перцептрона и количество классов будут влиять на размерность конечной задачи, но алгоритм проведенных действий останется неизменным. Основным отличием от стандартного метода обучения алгоритма распространения ошибки является отсутствие риска переобучения сети, которое

Получено 22.12.15

часто возникает при обучении в несколько эпох на одном и том же входе сети. Переобучение сети означает, что веса максимально подстраиваются под конкретные значения входных значений сети, и в рабочем режиме такая сеть дает неверные выходы даже при подаче незначительно отличающихся от обучающей выборки входных данных. В методе, основанном на классической оптимизации и сведению процесса обучения к дискретной задаче оптимального управления, сеть не может переобучиться, поскольку на каждом новом шаге обучения подается новое управление (входной набор значений сети), и на каждом шаге предыдущее управление вносит корректировки в веса. Соответственно, состояния (веса) подстраиваются под динамику управлений, а не под конкретные значения входов сети.

Библиографические ссылки

1. Rumelhart D. E., Hinton G. E., Williams R. J. Learning representations by back-propagating errors. – Nature 323 (1986). – P. 533–536.
2. Nikolaeva Yu. V. The neural networks. The multilayer perceptron // Communication of Students, Master Students and Post-Graduates in Academic, Professional and Scientific Fields : материалы межвуз. студ. науч. конф. – 2011.
3. Николаева Ю. В. Нормировка данных для нейронных сетей // Молодые ученые – ускорению научно-технического прогресса в XXI веке : сб. материалов III Всерос. науч.-техн. конф. аспирантов, магистрантов и молодых ученых с междунар. участием (Ижевск, 22–23 апреля 2015 года) / ИжГТУ имени М. Т. Калашникова.
4. Сучкова Е. А. Консолидация данных в системах поддержки принятия решения // Перспективы развития научных исследований в 21 веке : сб. материалов 9-й Междунар. науч.-практ. конф. (г. Махачкала, 31 октября 2015 г.). – Махачкала : Апробация, 2015. – 234 с.
5. Автоматизированные нейронные сети [Электронный ресурс statsoft.ru]. – URL: http://www.statsoft.ru/products/STATISTICA_Neural_Networks/ (дата обращения: 01.12.2015).
6. Там же.
7. Выбор функции активации и обучения нейронной сети [Электронный ресурс rac.ru]. – URL: <http://www.rac.ru/monographs/65-2465> (дата обращения: 01.12.2015).
8. Понтрягин Л. С. Математическая теория оптимальных процессов. – 3-е изд. – М. : Наука, 1976. – 392 с.

УДК 004.89

Е. А. Сучкова, аспирант, ИжГТУ имени М. Т. Калашникова

АНАЛИЗ КОНКУРСНЫХ ЗАКУПОК С ПРИМЕНЕНИЕМ МЕТОДОВ КЛАСТЕРИЗАЦИИ

Одним из способов расширения рынка сбыта продукции, товаров и услуг для предприятий является участие в государственных

закупках. Причем если раньше претендовать на победу могли лишь крупные предприятия, то в настоящее время при участии в конкурсных видах закупок