

Krassimir Markov, Vitalii Velychko, Oleksy Voloshin
(editors)

**Information Models
of
Knowledge**

ITHEA[®]
KIEV – SOFIA
2010

Krassimir Markov, Vitalii Velychko, Oleksy Voloshin (ed.)

Information Models of Knowledge

ITHEA®

Kiev, Ukraine – Sofia, Bulgaria, 2010

ISBN 978-954-16-0048-1

First edition

Recommended for publication by The Scientific Council of the Institute of Information Theories and Applications FOI ITHEA
ITHEA IBS ISC: 19.

This book maintains articles on actual problems of research and application of information technologies, especially the new approaches, models, algorithms and methods for information modeling of knowledge in: Intelligence metasynthesis and knowledge processing in intelligent systems; Formalisms and methods of knowledge representation; Connectionism and neural nets; System analysis and synthesis; Modelling of the complex artificial systems; Image Processing and Computer Vision; Computer virtual reality; Virtual laboratories for computer-aided design; Decision support systems; Information models of knowledge of and for education; Open social info-educational platforms; Web-based educational information systems; Semantic Web Technologies; Mathematical foundations for information modeling of knowledge; Discrete mathematics; Mathematical methods for research of complex systems.

It is represented that book articles will be interesting for experts in the field of information technologies as well as for practical users.

General Sponsor: Consortium FOI Bulgaria (www.foibg.com).

Printed in Ukraine

Copyright © 2010 All rights reserved

© 2010 ITHEA® – Publisher; Sofia, 1000, P.O.B. 775, Bulgaria. www.ithea.org ; e-mail: info@foibg.com

© 2010 Krassimir Markov, Vitalii Velychko, Oleksy Voloshin – Editors

© 2010 Ina Markova – Technical editor

© 2010 For all authors in the book.

® ITHEA is a registered trade mark of FOI-COMMERCE Co., Bulgaria

ISBN 978-954-16-0048-1

C/o Jusautor, Sofia, 2010

Neural Networks and Pattern Recognition

СРАВНИТЕЛЬНЫЙ АНАЛИЗ МЕТОДОВ МНОГОКРИТЕРИАЛЬНОЙ ОПТИМИЗАЦИИ АРХИТЕКТУРЫ НЕЙРОСЕТЕВЫХ КЛАССИФИКАТОРОВ

Альберт Воронин, Юрий Зиятдинов, Анна Антонюк

Аннотация. Рассматривается метод оптимизации архитектуры нейросетевого классификатора: поисковый метод симплекс-планирования. С помощью программы обучения нейронной сети и программы оптимизации проводится исследование корректности выбранного поискового метода для выявления наилучшей архитектуры нейросетевого классификатора. В качестве целевой функции предложена скалярная свертка критериев по нелинейной схеме компромиссов

Ключевые слова: многокритериальная оптимизация, архитектура, нейронные сети, классификатор.

ACM Classification Keywords: H.1 Models and Principles – H.1.1 – Systems and Information Theory; H.4.2 – Types of Systems; C.1.3 Other Architecture Styles – Neural nets

Введение

Настоящая работа является развитием результатов, полученных в работе [1]. Как и в [1], будем рассматривать структуру q -слойного нейросетевого классификатора с прямыми связями, представленную на рис.1.

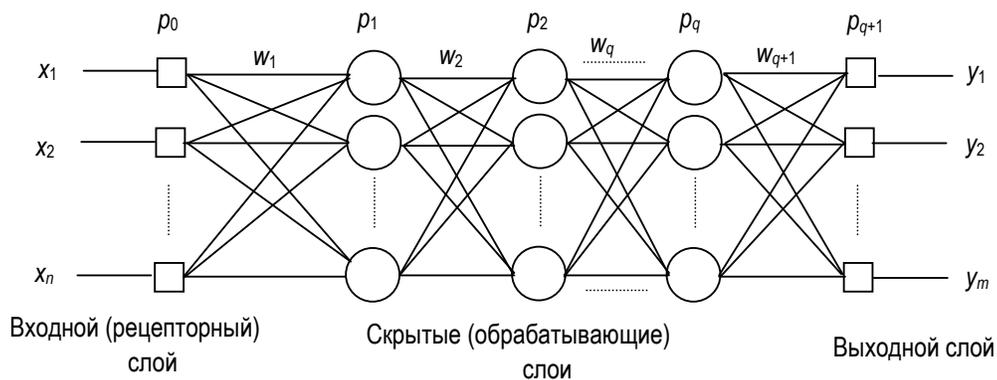


Рис.1

Здесь x_1, x_2, \dots, x_n – признаки объекта классификации, составляющие входной вектор $x = \{x_i\}_{i=1}^n$; $p_0 = n$ – число нейронных элементов в рецепторном слое; p_1, p_2, \dots, p_q – число нейронов в каждом из скрытых слоев; $p_{q+1} = m$ – число нейронов в выходном слое (число классов); $y = \{y_k\}_{k=1}^m$ – выходной вектор нейронной сети, определяющий принадлежность объекта

классификации одному из m классов; $w_1, w_2, \dots, w_q, w_{q+1}$ – векторы синаптических весов нейронной сети.

Необходимые сведения из теории нейронных сетей содержатся в [2-4]. На рис.2 представлена структурная схема нейрона.

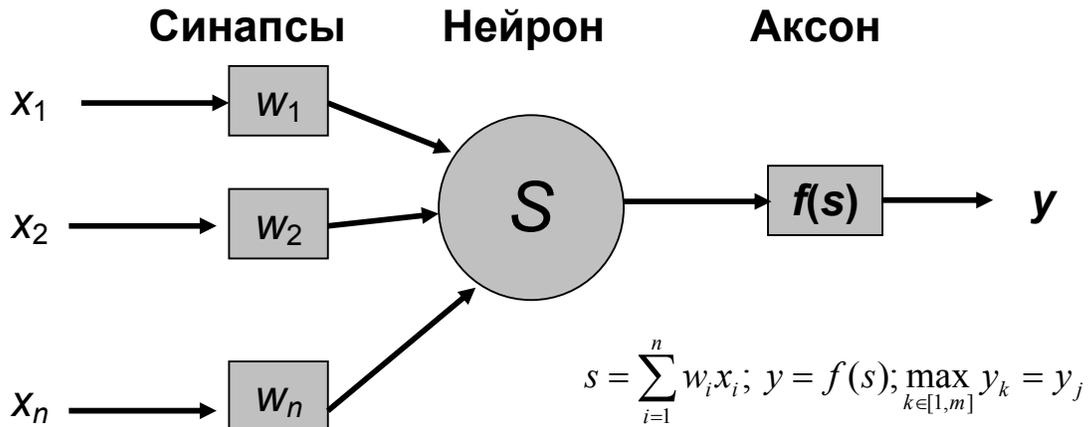


Рис. 2

Количество нейронов входного слоя $p_0 = n$ определяется размерностью входного вектора признаков и не подлежит изменениям. Аналогично, количество нейронов выходного слоя $p_{q+1} = m$ определяется числом областей (классов), на которые делится пространство признаков и тоже является постоянным. Количество же обрабатывающих (скрытых) слоев q и число нейронов в каждом из них p_1, p_2, \dots, p_q составляют понятие **архитектуры** [2] нейронной сети и могут служить аргументами (независимыми переменными) при ее оптимизации.

В настоящей работе ограничим исследование случаем, когда число q является фиксированным и заданным. Тогда аргументами оптимизации архитектуры нейронного классификатора являются количества нейронов в каждом из обрабатывающих слоев, составляющие вектор независимых переменных $p = \{p_j\}_{j=1}^q$. От выбора архитектуры p зависит качество функционирования нейронного классификатора.

Проблема заключается в таком выборе архитектуры, при котором нейронный классификатор в заданных условиях функционирования характеризуется наилучшими свойствами.

Постановка задачи

В общем виде проблема может быть формально представлена задачей

$$p^* = \arg \underset{p \in P}{extr} Y(p) \quad (1)$$

где $Y(p)$ – целевая функция; $\underset{p \in P}{extr}$ – оператор экстремизации целевой функции по аргументам p ; P

– допустимая область независимых переменных.

Для конструктивного решения задачи сделаем дополнительные частные предположения. Каждому свойству нейронного классификатора поставим в соответствие количественную характеристику $f(p)$, имеющую смысл критерия качества его функционирования. Одним из таких критериев является вероятность ошибки классификации. Будем определять этот критерий экспериментально и приближенно представим его как количество ошибок классификации $e(p)$, отнесенное к общему, достаточно большому количеству испытаний N :

$$f_1(p) = \frac{e(p)}{N} \quad (2)$$

Предполагается, что с ростом в некоторых разумных пределах числа нейронов в обрабатывающих слоях точность классификации повышается, и величина этого критерия уменьшается. Предельно допустимое значение ошибки сети должно быть известно из физических соображений и задано как ограничение

$$f_1(p) \leq A_1.$$

Второй критерий характеризует время, необходимое для обучения нейронной сети при данной архитектуре p . Наблюдается тесная корреляция между таким временем и суммарным количеством нейронов в скрытых слоях классификатора. Поэтому представим этот критерий в виде

$$f_2(p) = \sum_{k=1}^q p_k \quad (3)$$

Отметим, что данным критерием характеризуется и время прохождения сигнала через нейронную сеть от входа к выходу. С ростом числа нейронов значение критерия увеличивается. Предельно допустимое значение второго критерия определяется допустимым временем обучения нейронной сети и задается как ограничение $f_2(p) \leq A_2$.

Существуют и другие критерии для характеристики различных свойств нейронного классификатора. В данной работе мы ограничимся только приведенными двумя основными критериями, имея в виду, что излагаемая методика допускает включение в рассмотрение и других свойств классификатора.

Допустимая область аргументов оптимизации задается параллелепипедным ограничением $P = \{p | 0 \leq p_k \leq P_u, k \in [1, P_u], u \in [1, q]\}$, где P_u – максимальное число нейронов в u -м слое.

Поскольку оба включенных в рассмотрение критерия подлежат минимизации (чем критерий меньше, тем лучше соответствующее свойство классификатора), то оператор экстремизации целевой функции приобретает вид: $extr = \min_{p \in P}$.

Таким образом, оба критерия являются противоречивыми, неотрицательными, минимизируемыми и ограниченными. Налицо все предпосылки для использования в качестве целевой функции скалярной свертки критериев по нелинейной схеме компромиссов [4]. Такая свертка в унифицированной версии выражается формулой

$$Y(p) = Y[f(p)] = \frac{A_1}{A_1 - f_1(p)} + \frac{A_2}{A_2 - f_2(p)} \quad (4)$$

где $f(p) = \{f_r(p)\}_{r=1}^{r=2}$ – двумерный вектор частных критериев. Учитывая (2), (3) и (4), выражение (1) для задачи оптимизации архитектуры нейронного классификатора преобразуется к виду

$$p^* = \arg \min_{p \in P} \left[\frac{A_1}{A_1 - e(p)/N} + \frac{A_2}{A_2 - \sum_{k=1}^q p_k} \right] \quad (5)$$

Нетрудно видеть, что в формуле (5) зависимость $e(p)$ априори является неизвестной и подлежащей экспериментальному определению.

Методы оптимизации целевой функции

В работе [1] показано, что задача (5) является задачей нелинейного дискретного (целочисленного) программирования. Указаны предположения, при которых эта задача может решаться хорошо разработанными непрерывными методами.

В том случае, когда компоненты возможных решений многокритериальных задач могут принимать только дискретные значения $p_k^{(P_u)}$, $k \in [1, P_u]$, $u \in [1, q]$, скалярная свертка по нелинейной схеме компромиссов $Y(p)$ является *решетчатой* функцией, заданной на дискретном множестве P . Оптимизация решетчатой целевой функции, построенной по нелинейной схеме компромиссов, сводится к задаче нелинейного программирования с дискретными (целочисленными) аргументами, решение которой, как отмечено выше, достаточно сложно.

Для решения этой проблемы мы предполагаем, что при дискретном множестве P существует вспомогательная область непрерывных аргументов $p_c \in P_c$, которая содержит все дискретные точки $p_k^{(P_u)}$ и всё непрерывное пространство между ними. В области P_c определена непрерывная функция $Y(p_c)$, которая в точках $p_k^{(P_u)}$ совпадает с решетчатой функцией $Y(p)$.

Это предположение позволяет получить аналитическое решение, если в выражении (5) зависимость $e(p)$ задана, например, в виде регрессионной модели. Тогда можно воспользоваться необходимым условием минимума функции: $\frac{\partial Y(p_c)}{\partial p_c} = 0$. Решение этой системы уравнений дает компромиссно-

оптимальную непрерывную точку p_c^* . Последний шаг алгоритма – поиск на P ближайшей к p_c^* возможной *дискретной* точки, которая и будет искомым дискретным решением p^* . В нашем случае, к сожалению, задание аналитических зависимостей весьма затруднительно или вообще невозможно.

Мы рассматриваем как основной случай, когда функции $e(p)$ и, следовательно $Y(p)$, неизвестны, но есть возможность определять значения функции $Y(p)$ в точках $p_k^{(P_u)}$ измерением или вычислением. Тогда можно организовать натурный или вычислительный эксперимент, в результате которого осуществляется поисковое движение к искомой дискретной компромиссно-оптимальной точке p^* .

Возможны различные подходы к организации поисковой процедуры, которая должна давать последовательность улучшающихся решений. Один из них – дискретный аналог метода симплекс-планирования Нелдера-Мида (метод деформируемого многогранника) [5]. Это – разновидность градиентных методов, весьма часто и успешно применяющихся на практике.

Проведем исследование для $q=2$. На плоскости аргументов $p_1 - p_2$ в некоторой стартовой области строится исходный регулярный симплекс, который в двумерном случае представляет собой равнобедренный треугольник с вершинами $p^{(1)}, p^{(2)}, p^{(3)}$. Для архитектур стартового симплекса вычисляются значения скалярных сверток $Y^{(1)}, Y^{(2)}, Y^{(3)}$. Сравнивая между собой эти значения, определяются худшее. С большой вероятностью можно утверждать, что архитектура, полученная зеркальным отображением худшей в исходном симплексе точки относительно центра противоположной грани, окажется лучше. Исходя из этого, образуется новый симплекс. Процедура построения симплексов продолжается до тех пор, пока мы получим архитектуру p^* , соответствующую минимуму целевой функции [1].

Рассмотренный метод предусматривает старт поисковой процедуры от архитектуры, которая, по мнению разработчика, находится достаточно близко к оптимальной точке. Если в процессе поиска имеет место возрастание числа нейронов в обрабатываемых слоях, то теория нейронных сетей [2] характеризует данный подход как *конструктивный*. При избыточном стартовом количестве нейронов подход именуется *деструктивным* (принцип Родена: чтобы изваять скульптуру, нужно взять целую глыбу мрамора и удалить из нее лишнее).

Осуществление изложенных в работе этапов векторной оптимизации позволяет получить архитектуру нейросетевого классификатора, при которой системно увязываются противоречивые критерии эффективности его функционирования.

Обучение нейросетевого классификатора текстов

В качестве примера рассмотрим в общих чертах задачу многокритериальной оптимизации архитектуры нейросетевого классификатора текстов. Система текстовой классификации [4] состоит из двух основных частей: частотный анализатор с системным словарем и собственно нейросетевой классификатор. Структурная схема, а также этапы работы нейросетевого классификатора приведены в [1].

Обучение нейросетевого классификатора текстов было реализовано при помощи программы, написанной на языке программирования Delphi. (Рис. 4)

В данной программе использовался метод обучения с обратным распространением ошибки. Основная идея этого метода состоит в распространении сигналов ошибки от выходов сети к её входам, в направлении, обратном прямому распространению сигналов в обычном режиме работы.

На основе имеющегося множества учебных текстов (легальных писем и спама) создается системный словарь, который содержит в себе множество слов, встречающихся во всех учебных текстах, а также их частотную характеристику. При помощи кнопки «Загрузка матриц» формируется файл с весовыми коэффициентами нейронной сети, после чего можно приступить непосредственно к этапу обучения, в процессе которого весовые коэффициенты будут регулироваться программой, таким образом настраивая нейронную сеть для дальнейшей работы. Настроенную нейронную сеть можно использовать для классификации различных текстов. Также данная программа позволяет регулировать размерность нейронной сети и другие аргументы, использующиеся при вычислениях.

Многокритериальная оптимизация нейросетевого классификатора текстов

Для оптимизации архитектуры нейросетевого классификатора использовалось специальное программное обеспечение (Рис. 5)

В данной программе реализована возможность проведения оптимизации с использованием разных методов и в разных режимах. Режим «Эксперимент» позволяет провести оптимизацию в том случае, когда зависимости критериев от параметров не известны и для определения значения критериев необходим эксперимент. Режим «Аналитика» используется в том случае, когда критерии выражаются через

параметры явно, то есть, когда известны функциональные зависимости критериев от параметров. В данной работе будут использованы оба режима.

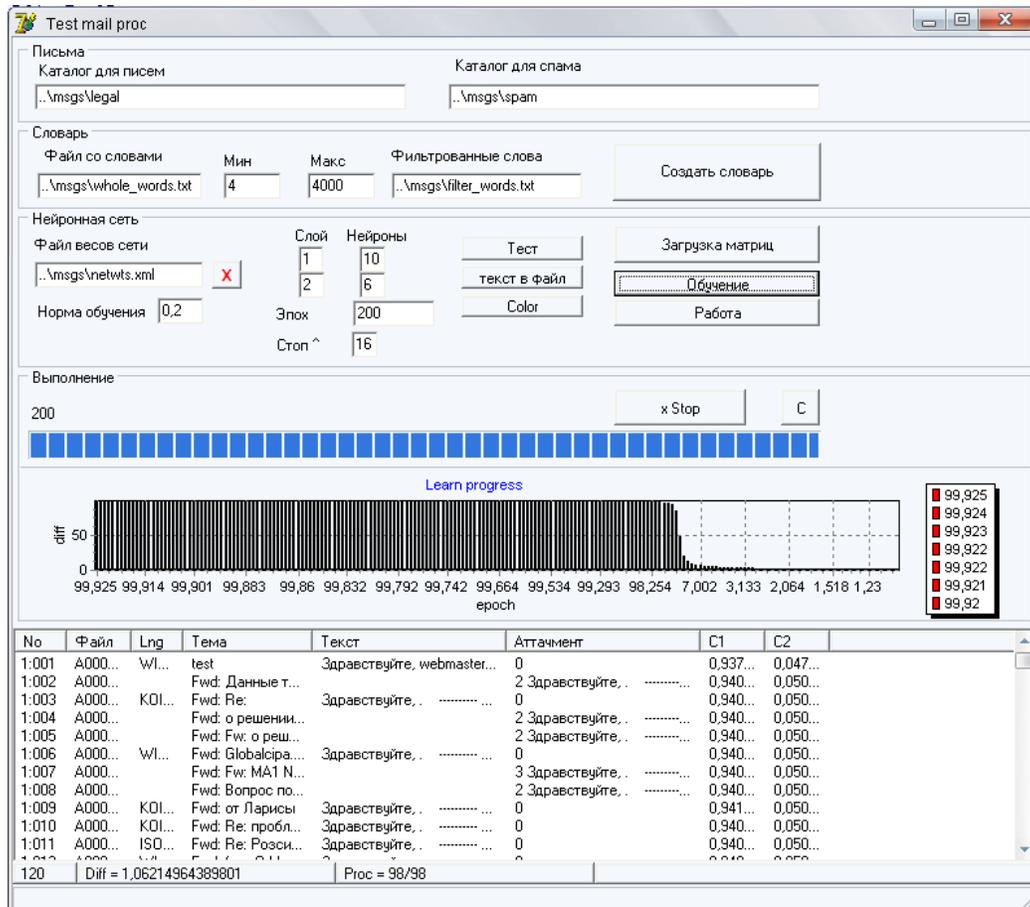


Рис. 4

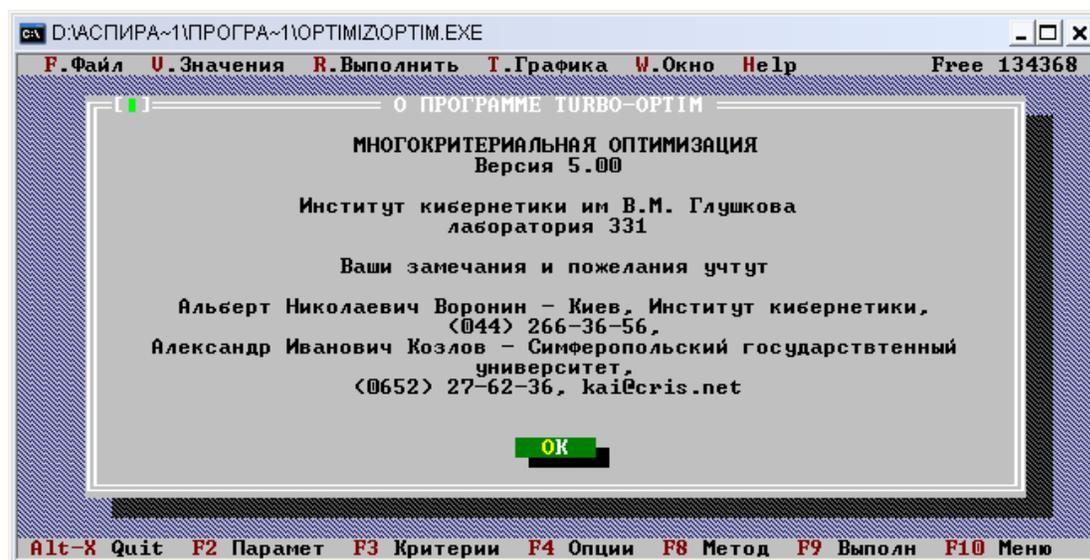


Рис. 5

Как уже упоминалось выше, в качестве критериев оптимизации были выбраны:

- Время обучения нейронной сети, пропорциональное суммарному количеству нейронов в сети, аналитическое выражение критерия выглядит следующим образом: $f_2(p) = \sum_{k=1}^q p_k$.
- Вероятность возникновения ошибки $f_1(p)$. Аналитическое выражение на начальном этапе проведения процедуры оптимизации является неизвестным.

Поскольку аналитические выражения известны не для всех критериев, на начальной стадии проводится оптимизация архитектуры нейросетевого классификатора в режиме «Эксперимент». В данном режиме процедура оптимизации проводится пошагово и требует знание значений критерия, определяющего вероятность возникновения ошибки, для каждой итерации. Значения данного критерия в каждом шаге итерации вычисляются в программе обучения нейросетевого классификатора.

Как уже упоминалось ранее, для проведения поисковой процедуры нахождения оптимальной архитектуры нейросетевого классификатора был выбран поисковый метод симплекс-планирования.

Используя программу обучения нейронной сети и программу оптимизации для метода симплекс-планирования, получим следующую таблицу значений:

Таблица 1

<i>и</i> - номер испытания	<i>p</i> ₁ - количество нейронов в 1-ом слое	<i>p</i> ₂ - количество нейронов в 2-ом слое	<i>e</i> (%) - значение ошибки классификации	<i>t</i> (с) - время обучения, $f_2(p) = \sum_{k=1}^q p_k$
1	6	6	99.9	12
2	9	7	3.87	16
3	7	9	99.0	16
4	10	10	0.376	20
5	13	8	0.274	21
6	11	8	0.533	19
7	11	5	1.6	16
8	9	3	99.8	12
9	11	7	0.67	18
10	9	5	99.6	14
11	10	6	1.19	16
12	11	4	99.7	15
13	10	6	1.19	16
14	11	5	1.6	16
15	11	7	0.67	18
16	11	5	1.6	16
17	11	6	0.95	17
18	11	5	1.6	16
19	11	6	0.95	17
20	11	6	0.95	17

Для аналитического выражения критерия, определяющего вероятность возникновения ошибки, была выбрана содержательная модель, имеющая вид: $f_1(p_1, p_2) = a_1 / p_1 + a_2 / p_2$, где a_1, a_2 – коэффициенты регрессии, p_1, p_2 – количество нейронов соответственно в 1-ом и 2-ом слоях.

Для нахождения коэффициентов регрессии составляется функция суммы квадратов невязок:

$$E(a) = \sum_{u=1}^N \left(\frac{a_1}{p_1^{(u)}} + \frac{a_2}{p_2^{(u)}} - f_1^{(u)} \right)^2,$$

минимизация которой, в соответствии с методом наименьших квадратов, дает неизвестные коэффициенты регрессии.

Таким образом, для метода симплекс-планирования было получено аналитическое выражение критерия возникновения ошибки:

$$y_1 = \frac{9.97}{p_1} + \frac{3.7}{p_2}.$$

Подставив аналитические выражения для критериев в режиме «аналитика», получим следующие значения аргументов, определяющие компромиссно-оптимальную архитектуру нейросетевого классификатора, найденные методом симплекс-планирования:

p_1 – количество нейронов в 1-ом слое, $p_1^* = 9$;

p_2 – количество нейронов во 2-ом слое, $p_2^* = 6$.

Таким образом, видно, что полученный результат подтверждает экспериментальные данные, представленные в Таблице 1. То есть использование метода симплекс-планирования для нахождения оптимальной архитектуры нейросетевого классификатора дает компромиссно-оптимальную модель, при которой системно увязываются противоречивые критерии эффективности его функционирования: время обучения и вероятность возникновения ошибки.

Благодарности

Статья частично финансирована из проекта ITHEA XXI Института Информационных Теорий и Приложений FOI ITHEA и Консорциума FOI Bulgaria (www.ithea.org, www.foibg.com).

Библиография

1. Альберт Воронин, Юрий Зиатдинов, Анна Антонюк, Многокритериальная оптимизация архитектуры нейросетевых классификаторов // International Book Series Number 8, P.32-39, ITHEA Sofia, 2009.
2. Бодянский Е.В., Руденко О.Г. Искусственные нейронные сети: архитектуры, обучение и применение. – Харьков: ТЕЛЕТЕХ, 2004. – 372 с.
3. Головкин В.А. Нейронные сети: обучение, организация и применение. – М.: ИПРЖР, 2001. – 256 с.
4. Борисов В.С. Самообучающийся классификатор текстов на естественном языке // Кибернетика и системный анализ – 2007. – №3. – С.169-176.
5. Воронин А.Н., Зиатдинов Ю.К., Козлов А.И. Векторная оптимизация динамических систем. – Киев: Техніка, 1999. – 284 с.
6. Воронин А.Н., Мосорин П.Д., Ясинский А.Г. Многокритериальные задачи оптимизации с дискретными аргументами // Автоматика-2000. Міжнародна конференція з автоматичного управління: Праці. –Т.1. – Львів: ДНДІІ, 2000. – С. 75-78.

Сведения об авторах

Воронин Альберт Николаевич – профессор, доктор технических наук, профессор кафедры компьютерных информационных технологий Национального авиационного университета, проспект Комарова, 1, Киев-58, 03058 Украина; e-mail: alnv@voliacable.com

Зиатдинов Юрий Кашафович – профессор, доктор технических наук, заведующий кафедрой компьютерных информационных технологий Национального авиационного университета, проспект Комарова, 1, Киев-58, 03058 Украина; e-mail: oberst@nau.edu.ua

Антонюк Анна Александровна – аспирант Национального авиационного университета, проспект Комарова, 1, Киев-58, 03058 Украина; e-mail: niuriel@mail.ru