

РАЗРАБОТКА ИНТЕЛЛЕКТУАЛЬНОЙ ИНФОРМАЦИОННОЙ СИСТЕМЫ ДЛЯ КАТЕГОРИЗАЦИИ ДЕТЕЙ, ОСТАВШИХСЯ БЕЗ ПОПЕЧЕНИЯ РОДИТЕЛЕЙ

Л. В. Хливненко, В. В. Васильев, А. Е. Васильев

Воронежский государственный университет

Поступила в редакцию 01.03.2011 г.

Аннотация. Рассматривается модель самоорганизующейся системы – категоризатора с архитектурой искусственной нейронной сети Кохонена и процесс извлечения на ее основе метазнаний из поля знаний для поддержки документооборота в отделе опеки и попечительства. Приводится описание компьютерной реализации модели в виде разработанного Windows-приложения, созданного в свободно распространяемой среде программирования Lazarus.

Ключевые слова: нейросетевые алгоритмы, самоорганизующиеся интеллектуальные системы.

Abstract. The article discusses a model of the self-organizing system with architecture Kohonen's neural network and a process of the extraction knowledge from database for support of the document processing in organization of the trusteeship. Computer realization of the models is executed in the manner of Windows-application, created in free integrated environment of the development Lazarus.

Keywords: neural network algorithms, self-organizing intellectual systems.

ВВЕДЕНИЕ

Задача категоризации детей, оставшихся без попечения родителей, заключается в разделении общей массы детей на группы, позволяющие судить о благоприятности условий жизни ребенка и своевременном оказании помощи детям, которые в ней нуждаются. Прогноз попадания в ту или иную социальную группу в будущем, проведенный по имеющейся неполной информации в настоящем, позволит вовремя сконцентрировать работу психологов и социальных педагогов с детьми, вошедшими в группу риска.

В настоящее время действует Постановление Правительства РФ от 25 декабря 2009 г. N 1088 о создании единой вертикально интегрированной государственной автоматизированной информационной системы «Управление» [1]. В рамках выполнения данного Постановления в органы опеки и попечительства внедряется программное обеспечение «АИСТ» с целью формирования региональных банков данных о детях, нуждающихся в устройстве в замещающую семью [2], www.opeka39.ru.

При анализе всей имеющейся в отделе опеки и попечительства информации необходимо выделить значимые, ключевые в описании те-

кущего момента жизни ребенка данные. В поле знаний, построенном для поддержки документооборота в отделе опеки и попечительства, нужно ограничить имеющийся объем информации с последующим выявлением закономерностей и построением выводов [3].

При решении задач категоризации набор формальных признаков, характерных для категории, неизвестен. Поэтому построить базу знаний посредством правил типа «если... – то...» в данном случае затруднительно. Сама категория понимается не через формальное определение, а только при сравнении с другими категориями [4]. Границы различных категорий могут быть расплывчатыми.

Для решения задачи категоризации детей, оставшихся без попечения родителей, был выбран подход, базирующийся на принципах синергетики, на построении самоорганизующихся систем.

Задачей системы – категоризатора является формирование обобщающих признаков в совокупности примеров. При увеличении числа примеров несущественные, случайные признаки сглаживаются, а часто встречающиеся – усиливаются, при этом происходит постепенное уточнение границ категорий.

Хорошо обученная самоорганизующаяся система способна извлекать признаки из новых

примеров, ранее неизвестных системе, и принимать на их основе приемлемые решения [5].

Основной целью представленного в работе исследования являлось построение системы, предназначенной для упорядочивания имеющейся информации о детях и выделения (формирования) категорий, извлечения признаков, по которым можно строить прогнозы о дальнейшем развитии ситуации с ребенком. В качестве инструментария были выбраны самоорганизующиеся искусственные нейронные сети, успешно решающие поставленные задачи.

Чтобы оперативно реагировать на ситуацию с ребенком, правильно выбрать форму его устройства, необходим своевременный анализ. Время реакции системы должно быть допустимо малым. Обученные нейросетевые системы оказываются более выигрышными в плане временных затрат на обработку эмпирических данных [6].

1. МОДЕЛЬ САМООРГАНИЗУЮЩЕЙСЯ СИСТЕМЫ-КАТЕГОРИЗАТОРА

Искусственная нейронная сеть с архитектурой Кохонена состоит из входного слоя и выходного слоя нейронов (рис.1). В такой сети, обученной по алгоритму self – organizing map, нейроны в ходе конкурентного процесса избирательно настраиваются на представление кластеров во входной информации, что позволяет использовать сети данной архитектуры для решения задач категоризации [7].

На вход сети подается вектор $X(x_1, \dots, x_n) \in R^n$. Входной вектор является закодированным кортежем с информацией о ребенке.

Закодированные данные кортежа переводятся в безразмерную форму вычитанием вы-

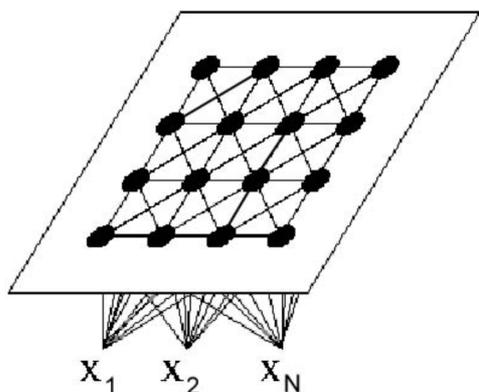


Рис.1. Структура искусственной нейронной сети Кохонена

борочного среднего и нормированием на корень из исправленной выборочной дисперсии.

Количество нейронов выходного слоя совпадает с количеством требуемых категорий или превышает их для случая представления одной категории нейронным ансамблем.

Нейрон выходного слоя характеризуется весовым вектором $Y_j(w_{1j}, \dots, w_{mj}) \in R^m$, где $j = 1, m$, m – количество нейронов выходного слоя.

В сети Кохонена используется конкурентное обучение без учителя на основе самоорганизации. В процессе обучения веса нейронов становятся прототипами классов – групп векторов обучающей выборки.

Алгоритм обучения включает три основных шага: подвыборка, поиск максимального соответствия и корректировка весов [8].

При инициализации сети весовым коэффициентам сети w_{ij} присваиваются малые случайные значения. Задаются значения α_0 – начальный темп обучения и D_0 – максимальное расстояние между весовыми векторами.

Шаг 1. Сети предъявляется пример x^N из обучающего множества. Вычисляется квадрат расстояния от входа x^N до всех нейронов сети:

$$d_j^2 = \sum_{i=1}^n (x_i - w_{ij}^N)^2, \quad j = \overline{1, m}.$$

Шаг 2. Выбирается нейрон k , $1 \leq k \leq m$, расстояние d_k от которого до x^N минимально. Нейрон объявляется «победителем», т.к. он выиграл соревнование за право представлять в сети входной пример x^N .

Шаг 3. Происходит настройка весов нейрона k и всех нейронов, находящихся от него на расстоянии не превосходящим D_N по формуле: $w_{ij}^{N+1} = w_{ij}^N + \alpha_N (x_i - w_{ij}^N)$.

В процессе обучения значения α_N и D_N уменьшаются. Шаги 1–3 повторяются до тех пор, пока суммарное изменение всех весов не станет допустимо малым.

Во время обучения каждый вектор из входного набора данных связывается с ближайшим к нему нейроном. По завершению обучения весовые векторы нейронов становятся центрами тяжести сгустков близких входных сигналов (рис.2).

Результатом обучения самоорганизующейся нейронной сети с архитектурой self – organizing map, является карта упорядоченности входной информации, по которой можно формировать категории, состоящие из наборов расположенных рядом кластеров [9].

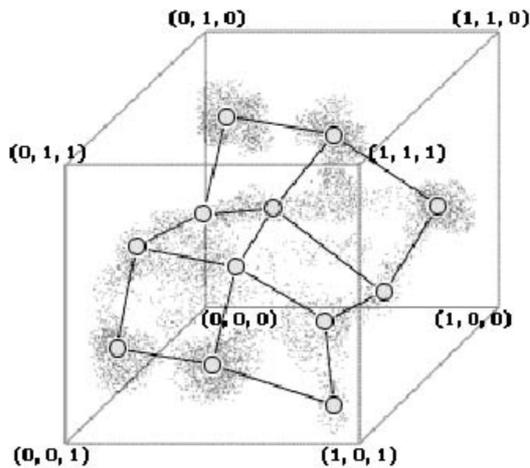


Рис.2. Распределение выходных сигналов сети после обучения

2. ПРОГРАММНАЯ РЕАЛИЗАЦИЯ МОДЕЛИ

Демонстрационный прототип интеллектуальной системы был выполнен в свободной среде программирования Lazarus, ориентированной на взаимодействие со свободным программным обеспечением. Нейросетевой модуль интегрирован с клиент-серверной базой данных

MySQL, реляционная модель которой разработана авторами для поддержки документооборота в отделе опеки и попечительства.

На учете в районном отделе опеки и попечительства состоит от 60 до 80 детей. С учетом информации о родственниках и опекунах получается более 150 человек. Тестирование демонстрационного прототипа системы выполнялось на 100 обучающих примерах.

Схема демонстрационного прототипа интеллектуальной информационной системы представлена на рис.3. Работа программного модуля оформляется в виде отчета на языке PHP, формируемого на стороне сервера.

При кодировании информации о ребенке были учтены такие данные, как: пол, год рождения, год постановки на учет в отдел, год снятия с учета, вид формы устройства, состояние здоровья, занятость, статус матери и отца, год рождения опекуна, родственные отношения с опекуном, занятость опекуна и др.

При запуске программы в оперативной памяти формируется матрица входных данных по выделенной информации из базы данных.

Ниже приведен фрагмент кода программы, в котором проводится нормализация входных данных.

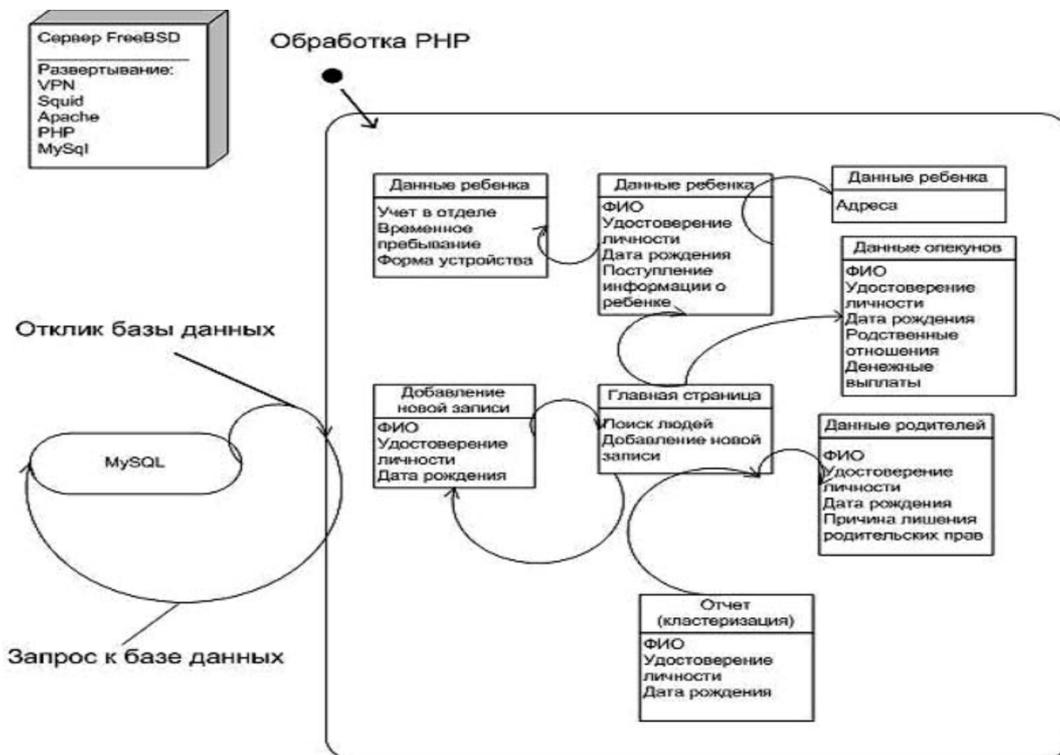


Рис.3. Крупноблочная схема работы интеллектуальной информационной системы

```
for i := 1 to count do begin
  average := 0; for j := 1 to n
do average := average + x[i,j];
  average := average / n;
  sumsq := 0; for j := 1 to n
do sumsq := sumsq + sqr(x[i,j]);
  variance := sumsq / n -
sqr(average); variance := n *
variance / (n-1);
  for j := 1 to n do x[i,j] :=
(x[i,j] - average) / sqrt(variance)
end;
```

После обучения сети по алгоритму self – organizing map весовые векторы нейронов получают в нормализованном виде. Для формирования прототипов категорий необходимо выполнить денормализацию весовых векторов по усредненной выборочной средней и исправленной выборочной дисперсии группы.

Примеры сформированных карт, полученных при варьировании числом итераций (1000–10000) и параметром скорости обучения (0,001–0,1), приведены на рис. 4. При обучении были также построены карты, в которых количество нейронов совпадало с количеством примеров. В этом случае каждый из нейронов настроился на представление конкретного примера и карта упорядоченности входной информации получилась полной.

Результатом обучения самоорганизующейся нейросети являются также сформированные в процессе обучения весовые векторы нейронов, которые служат координатами центров тяжести кластеров обучающей выборки и являются

маркерами для классификации новых входных примеров.

Обученная искусственная нейронная сеть может использоваться для прогнозирования рисков попадания ребенка в неблагополучную категорию. Распознавание категории происходит по нейрону сети, ближайшему к новому входному вектору. Нейрон-победитель в данном случае служит маркером кластера, к которому следует отнести новый входной пример.

Смысловая интерпретация карт упорядоченности примеров из обучающей выборки, полученных усредненных данных о центрах кластеров состава кластеров и уточнение границ категорий должна выполняться специалистами в данной предметной области. Данные исследования запланированы на перспективу.

Представленная работа была выполнена по инициативе работников отдела опеки и попечительства Железнодорожного р-на г. Воронежа. Демонстрационный прототип интеллектуальной информационной системы был представлен в отделе и получил положительные отзывы. Внедрение программного обеспечения в государственные структуры связано с длительными процедурами проведения экспертиз и распределением финансовых потоков структуры.

ЗАКЛЮЧЕНИЕ

В рамках представленной работы рассмотрено решение задачи категоризации детей, оставшихся без попечения родителей, основанное на построении модели самоорганизу-

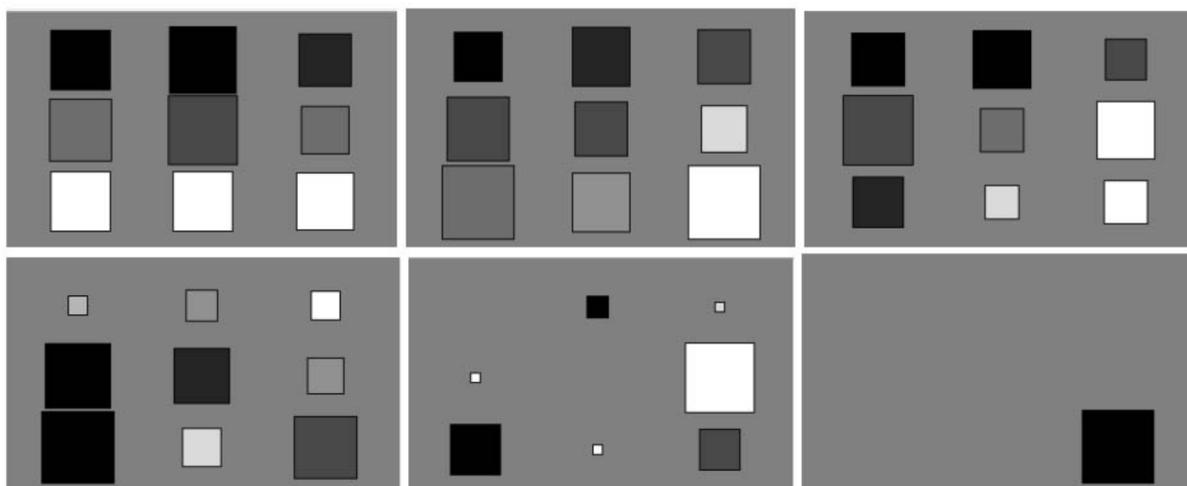


Рис.4. Карты кластеризации входной информации, полученные при разных настройках процесса обучения

ющейся системы – категоризатора с архитектурой искусственной нейронной сети Кохонена. Описано построение основного ядра интеллектуальной информационной системы, позволяющего решать задачу прогнозирования на основе выделения закономерностей в большом массиве данных. Рассмотрены аспекты программной реализации решателя интеллектуальной информационной системы, выполняющего разбиение поля знаний на кластеры.

Рекомендации интеллектуальной информационной системы могут быть полезны в процессе своевременной диагностики ситуации с ребенком и в случае необходимости быстрого принятия решений при определении стратегии и тактики устройства ребенка и оказания ему помощи. Качественный своевременный анализ, основанный на накопленной статистической информации, предоставляет возможность построения прогнозов развития событий, связанных с ребенком, оказавшимся в сложной жизненной обстановке.

СПИСОК ЛИТЕРАТУРЫ

1. Положение о единой вертикально интегрированной государственной автоматизированной информационной системе “Управление”. Москва. Правительство РФ. Утверждено Постановлением Правительства РФ от 25 декабря 2009 г. N 1088.

Хливненко Любовь Владимировна – кандидат технических наук, доцент кафедры математического моделирования Воронежского государственного университета. Тел. (473) 2208364. E-mail: hlivnenko_lv@mail.ru

Васильев Валерий Викторович – кандидат физико-математических наук, доцент, зам. председателя профкома Воронежского государственного университета. Тел. (473) 2208732. E-mail: vvv_252v@yandex.ru

Васильев Александр Евгеньевич – выпускник математического факультета Воронежского государственного университета. E-mail: vasilyev_a@bk.ru

2. Целевая программа Калининградской области «Дети-сироты» на 2007–2011 гг. Приложение к постановлению Правительства Калининградской области от 25 августа 2010 года № 700.

3. *Шараевский Г.И.* Алгоритм формирования карт самоорганизации распознающей нейросетевой структуры в условиях априорных ограничений / Г.И. Шараевский // Математичне та комп’ютерне моделювання. Серія: Технічні науки. Випуск 3. – 2010. – С. 195–200.

4. *Гаврилова Т.А.* Базы знаний интеллектуальных систем: учебник / Т.А. Гаврилова, В.Ф. Хоросhevский. – СПб.: Питер, 2001. – 384 с.

5. *Башмаков А.И.* Интеллектуальные информационные технологии: учебное пособие / А.И. Башмаков, И.А. Башмаков. – М.: Изд-во МГТУ им. Н.Э.Баумана, 2005. – 302 с.

6. *Filippi A.* Self-Organizing Map-based Applications in Remote Sensing / A. Filippi, I. Dobрева, A. Klein, J. Jensen // Self-Organizing Maps. – InTech, April 2010.

7. *Teuvo Kohonen.* Self-Organizing Maps: Springer, 2006. – 665 p.

8. *Каширина И.Л.* Искусственные нейронные сети: учебное пособие / И.Л. Каширина. – Воронеж: Издательство ВГУ, 2005. – 51 с.

9. *Садыхов Р.Х.* Новые алгоритмы формирования SOM нейронных сетей в задаче распознавания образов / Р.Х. Садыхов, М.Е. Ваткин // Нейрокомпьютеры: разработка, применение, №1, 2004. – С. 23–31.

Hlivnenko L.V. – Candidate of Technical Sciences, Associate Professor, department of Mathematical Modeling, Voronezh State University. Tel. (473) 2208364. E-mail: hlivnenko_lv@mail.ru

Vasiliev V.V. – Candidate of Physical and Mathematical Sciences, Associate Professor, Voronezh State University. Tel. (473) 2208732. E-mail: vvv_252v@yandex.ru

Vasiliev A.E. – Graduate Student, Mathematical Faculty, Voronezh State University. E-mail: vasilyev_a@bk.ru