

УДК 004.942

АВТОМАТИЗИРОВАННАЯ СИСТЕМА ИДЕНТИФИКАЦИИ ЗАКОНОВ РАСПРЕДЕЛЕНИЯ ДВУМЕРНОЙ НЕЙРОННОЙ СЕТЬЮ КОХОНЕНА

© 2014 И.В. Лёзина, Н.Ю. Яшин

Самарский государственный аэрокосмический университет имени академика С.П. Королёва
(национальный исследовательский университет)

Поступила в редакцию 17.12.2013

В статье описана разработанная авторами автоматизированная система для идентификации законов распределения. В системе используется двумерная нейронная сеть (карта) Кохонена, обучаемая алгоритмами WTA, Кохонена и нейронного газа. В статье проводится анализ результатов идентификации и исследование зависимости качества идентификации от параметров сети.

Ключевые слова: автоматизированная система, нейронная сеть, сеть Кохонена, идентификация, закон распределения, алгоритм WTA, алгоритм Кохонена, алгоритм нейронного газа.

Распознавание (идентификация) законов распределения по исходным числовым данным является достаточно распространённой задачей, решение которой может быть применено в различных областях приборостроения для оценки брака на производстве, для анализа экономических процессов, в банковском деле - для расчетов времени обслуживания заявок, а также в учебных целях. Для таких задач как нельзя лучше подходит использование интеллектуальных систем на основе нейронных сетей. В данной статье рассматривается использование нейронной сети Кохонена для идентификации законов распределения.

Сеть Кохонена относится к классу самоорганизующихся нейронных сетей и представляет собой одно- или двумерную решетку нейронов, каждый из которых путем изменения своих весовых коэффициентов настраивается на определенный класс входных векторов. Общая формула адаптации весов нейронов сети может быть представлена в виде [1] (1):

$$w_j(n+1) = w_j(n) + \eta(n)h_{j,i(x)}(n)(x - w_j(n)), \quad (1)$$

где $\eta(n)$ – параметр скорости обучения; $h_{j,i(x)}(n)$ – функция окрестности с центром в победившем нейроне $i(x)$.

В исследовании использовался набор выборок случайных величин с заданным законом распределения. Генерация выборок осуществлялась при помощи метода обратной функции. Входными данными для нейронов сети послужили значения высот столбцов гистограммы.

На этапе, предвещающем обучение, произво-

дится инициализация весовых коэффициентов нейронов сети псевдослучайными числами, а также производится нормализация значений входных векторов.

В качестве алгоритмов обучения были использованы алгоритмы WTA, Кохонена и нейронного газа [2]. Обучение по алгоритму WTA отличается тем, что на каждой итерации обучения изменению весов подвергается лишь один нейрон сети (нейрон-победитель), т.е. значение функции окрестности равно единице. Веса этого нейрона в наибольшей степени соответствуют весам предъявляемой на данной итерации обучения входной выборки. Мерой соответствия выбрано евклидово расстояние весов входного нейрона и нейрона-победителя, определяемое по формуле [2] (2):

$$d(x, w_i) = \|x - w_i\| = \sqrt{\sum_{j=1}^N (x_j - w_{ij})^2}, \quad (2)$$

где x_j и w_{ij} - веса входного вектора и нейрона сети соответственно.

Обучение по алгоритму Кохонена подразумевает адаптацию весовых коэффициентов как нейрона-победителя, так и нейронов-соседей. Существуют два типа соседства: прямоугольное соседство или соседство по Гауссу. При прямоугольном соседстве функция соседства (окрестности) определяется в виде [2] (3):

$$S(i, x) = \begin{cases} 1, & \text{для } -K < d(i, w) < K \\ 0, & \text{иначе} \end{cases}, \quad (3)$$

Функция окрестности по Гауссу может быть представлена в виде [1] (4):

$$h_{j,i(x)} = \exp\left(-\frac{d_{j,i}^2}{2\sigma^2}\right), \quad (4)$$

Лёзина Ирина Викторовна, кандидат технических наук, доцент кафедры "Информационные системы и технологии". E-mail: chuchyck@yandex.ru
Яшин Никита Юрьевич, студент первого курса магистратуры факультета информатики.
E-mail: nikita_888@list.ru

где $d_{j,i}$ – расстояние между победившим и соседним нейронами, σ – эффективная ширина топологической окрестности (уровень соседства).

Обучение по алгоритму нейронного газа включает в себя проводимую на каждой итерации обучения сортировку всех нейронов зависимости от их расстояния до входного вектора x . Отсортированные нейроны размещаются в последовательности, соответствующей увеличению удаленности

$$d_0 < d_1 < d_2 < \dots < d_{K-1}, \quad (5)$$

где $d_i = \|x - w_{m(i)}\|$ обозначает удаленность i -го нейрона, занимающего в результате сортировки m -ю позицию в последовательности [2]. Значение функции соседства для i -го нейрона $S(i, x)$ определяется по формуле

$$S(i, x) = \exp\left(-\frac{m(i)}{\lambda}\right), \quad (6)$$

в которой $m(i)$ обозначает очередность, полученную в результате сортировки, а λ – параметр, аналогичный уровню соседства в алгоритме Кохонена, который уменьшается с течением времени [2].

Значения параметров $\eta(n)$ и σ для реализации эффективного обучения с течением времени должны уменьшать свои значения. В системе величины этих параметров экспоненциально убывают в зависимости от числа пройденных итераций (эпох) обучения.

Таким образом, обучение сети можно разделить на три этапа: конкуренция нейронов, кооперация нейронов и синаптическая адаптация нейронов сети.

Для решения поставленных задач было разработано программное обеспечение, моделирующее работу нейронной сети [3]. Функции системы наиболее полно можно изобразить с помощью диаграммы вариантов использования, составленной по методологии UML, представленной на рис. 1.

В системе реализована работа с выборками обучающих и тестовых данных (генерация по заданному закону распределения, сохранение в файл). На рис. 2 представлен внешний вид программы во время генерации выборок.

Как уже упоминалось ранее, в качестве обучающей информации в системе выступают значения высот столбцов гистограммы. Пример гистограммы, формируемой системой для закона Симпсона, представлен на рис. 3.

Перед обучением пользователь имеет возможность выбрать алгоритм обучения, задать стартовый коэффициент обучения и количество итераций обучения, тем самым регулируя степень и качество обучения сети.

Окно программы во время идентификации представлено на рис. 4. Пользователь имеет возможность наблюдать обученную карту Кохонена, веса тестовой выборки, веса нейрона-победителя и их отклонение от весов тестовой выборки, название закона, а также увидеть какой именно нейрон в сети сработал при подаче исходной тестовой выборки.

С помощью системы [4] было произведено исследование идентификационных возможностей сети Кохонена в зависимости от алгоритма обучения, параметров обучения и прочих пара-

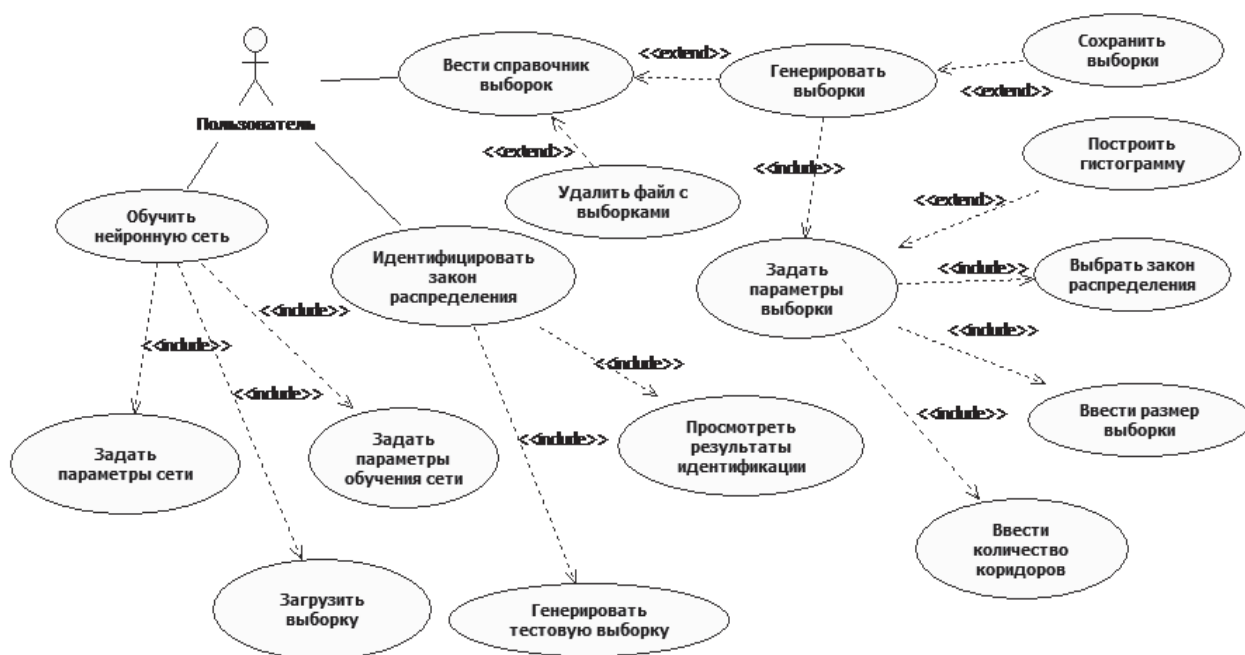


Рис. 1. Диаграмма вариантов использования системы

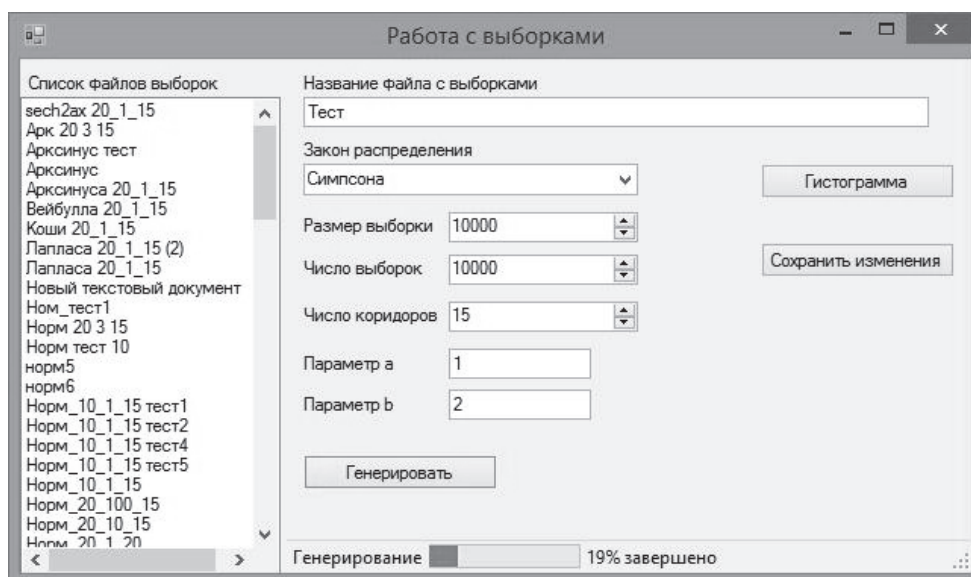


Рис. 2. Окно программы во время генерации выборок

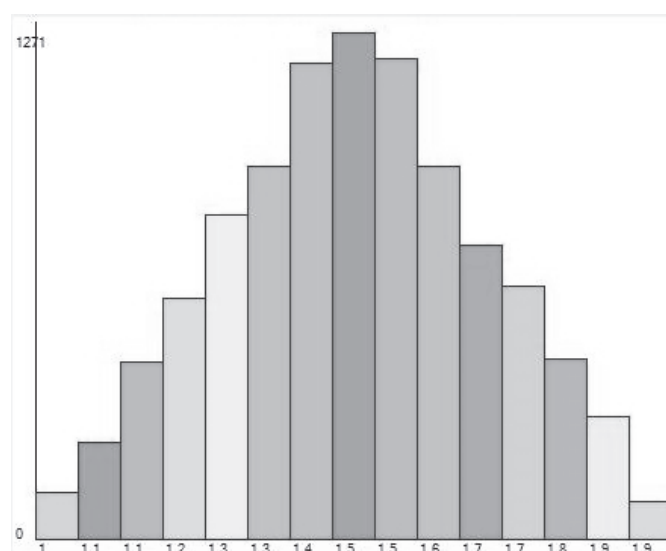


Рис. 3. Пример гистограммы, формируемой системой для закона Симпсона

метров. В качестве показателя качества идентификации при разных параметрах использовался процент верно распознанных законов распределения. В исследовании использовались различные законы распределения случайной величины, ниже рассмотрены результаты для закона арксинуса, нормального и равномерного законов.

Результаты исследования зависимости процента верно распознанных законов распределения от числа нейронов в сети представлены в табл. 1. Обучающие выборки, построенные по 20000 отсчетов (число дифференциальных коридоров $M = 15$), 50 тестовых выборок с аналогичными характеристиками, коэффициент обучения $\alpha = 0,5$, максимальное число итераций обучения 3000. Возьмем за a – размерность карты Кохонена по вертикали, за b – размерность карты по горизонтали.

По данным табл. 1 видно, что при маленьких размерах карты Кохонена наиболее целесообраз-

ным для обучения является использование алгоритмы WTA или нейронного газа. Сеть, обученная по алгоритмам Кохонена с различными типами соседства, при небольших значениях размерностей показывает худшие результаты. Из таблицы видно, что не имеет значения, какого типа карта – квадратная или прямоугольная, основным фактором является количество нейронов в сети. Наилучшие результаты показаны при количестве нейронов не меньше 225 и размерах решетки 15 на 15.

Результаты исследования зависимости процента верно распознанных законов от числа дифференциальных коридоров выборки представлены в таб. 2. Обучающие выборки, построенные по 20000 отсчетов, 50 тестовых выборок с аналогичными характеристиками, коэффициент обучения $\eta = 0,5$, максимальное число итераций обучения 3000, количество нейронов сети – 225.

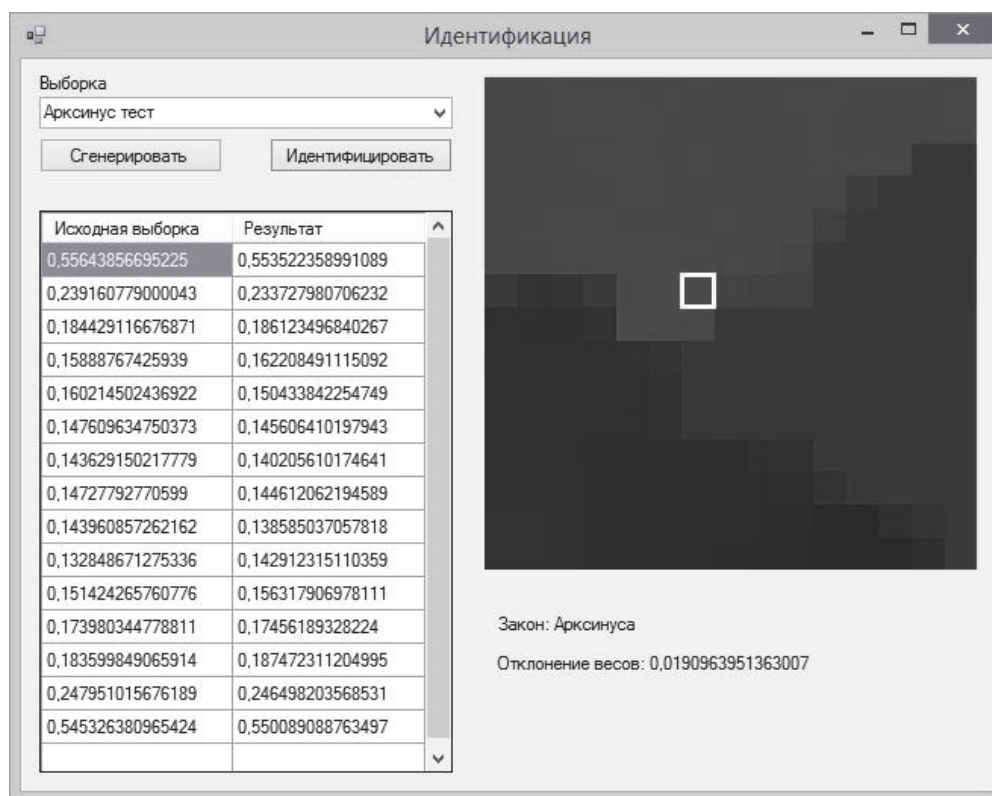


Рис. 4. Форма программы во время идентификации

Таблица 1. Зависимость процента верно распознанных законов распределения идентификации от числа нейронов в сети

Число нейронов			Процент верно распознанных законов			
a	b	Всего	WTA	Кохонена (прямоуг. соседство)	Кохонена (соседство по Гауссу)	Нейронного газа
5	5	25	90	86	82	92
10	5	50	92	94	94	95
10	10	100	96	97	96	96
20	5	100	96	96	97	97
15	10	150	98	97	97	97
15	15	225	98	98	99	99
20	20	400	99	99	99	99

Таблица 2. Зависимость процента верно распознанных законов от числа дифференциальных коридоров

Число диф. коридоров M	Процент верно распознанных векторов			
	WTA	Кохонена (прямоуг. соседство)	Кохонена (соседство по Гауссу)	Нейронного газа
10	97	96	97	96
15	98	98	99	99
20	92	92	98	99

Сеть одинаково хорошо настраивается на обучающие выборки с различным количеством коридоров, однако значение параметра ниже 10 не рекомендуется использовать в силу того, что подобное количество коридоров гистограммы не учитывает индивидуальные особенности законов рас-

пределения и затруднит их идентификацию.

Результаты исследования зависимости процента верно распознанных законов в режиме тестирования от размера выборки представлены в табл. 3. Обучающие выборки, построенные с числом дифференциальных коридоров $M = 15, 50$

Таблица 3. Зависимость процента верно распознанных законов от размера выборки

Объем выборки N	Процент верно распознанных векторов			
	WTA	Кохонена (прямоуг. соседство)	Кохонена (соседство по Гауссу)	Нейронного газа
5000	96	92	94	94
10000	98	94	98	99
20000	98	98	99	99

тестовых выборок с аналогичными характеристиками, коэффициент обучения $\eta = 0,5$, максимальное число итераций обучения 3000, количество нейронов сети – 225.

Исходя из данных табл. 3, можно сделать вывод о том, что чем большее число отсчетов учитывается при построении гистограммы, тем она более точно построена и приближена к реальной функции плотности вероятности. Следовательно, сеть будет учиться на более верных и приближенных к тестовым и контрольным выборкам результатах.

Результаты исследования зависимости процента верно распознанных законов от количества обучающих выборок для каждого закона представлены в таб. 4. Обучающие выборки, построенные по 20000 отсчетов (число дифференциальных коридоров $M = 15$), 50 тестовых выборок с аналогичными характеристиками, коэффициент обучения $\eta = 0,5$, максимальное число итераций обучения 15000 (при числе обучающих выборок для одного закона 5000), количество нейронов сети – 225.

По данным табл. 4 наилучшие результаты дает сеть, обученная на 1000 выборках для каждого закона распределения. При числе обучающих выборок менее 1000 сеть получает слишком мало информации для качественного восстановления информации. Напротив, при слишком большом числе различных обучающих выборок (более 1000 в несколько раз) сеть адаптируется к шумам каждой выборки, то есть “переучивается”.

Результаты исследования зависимости процента верно распознанных законов от стартового коэффициента обучения приведены в табл. 5. Обучающие выборки, построенные по 20000 отсчетов

(число дифференциальных коридоров $M = 15$), 50 тестовых выборок с аналогичными характеристиками, максимальное число итераций обучения 3000, количество нейронов сети – 225.

Учитывая приведенные выше данные, можно сделать вывод о том, что чем меньше стартовый коэффициент обучения, тем медленнее происходит обучение сети, количество “мертвых”, незадействованных в обучении нейронов возрастает. Также, несмотря на сравнимые показатели для малых стартовых значений коэффициентов обучения по алгоритму WTA и реализациям алгоритма Кохонена и нейронного газа, вектора весов, которые были найдены нейронной сетью, обученной по алгоритму WTA, в процессе идентификации имели гораздо более сильные отклонения весов от значений исходной выборки. Рекомендуется использовать значение коэффициента обучения из диапазона $[0,2, 0,8]$.

По результатам проведенных исследований можно сделать вывод о достаточно хороших идентификационных возможностях двумерной нейронной сети Кохонена. Процент верно распознанных законов при различных конфигурациях и параметрах сети, обучающих данных и параметрах обучения составляет 74–99%. Однако стоит отметить, что обучение по алгоритму WTA не обеспечивает достаточно полного охвата всей карты Кохонена, некоторые нейроны так и остаются “мертвыми” в процессе обучения.

Важно отметить, что карты Кохонена обладают существенным достоинством – это удобная визуализация процесса обучения сети. Веса нейронов определяют их цвет на карте. По результатам тестирования программы выяснено, что областям, отвечающим за идентификацию закона арксинуса, соответствует фиолетовый цвет,

Таблица 4. Зависимость процента верно распознанных законов от числа обучающих выборок для каждого закона

Число обучающих выборок	Процент верно распознанных векторов			
	WTA	Кохонена (прямоуг. соседство)	Кохонена (соседство по Гауссу)	Нейронного газа
100	74	75	80	88
1000	98	98	99	99
5000	86	88	89	94

Таблица 5. Зависимость процента верно распознанных законов от стартового коэффициента обучения

Стартовый коэффициент обучения	Процент верно распознанных векторов			
	WTA	Кохонена (прямоуг. соседство)	Кохонена (соседство по Гауссу)	Нейронного газа
0,05	91	92	94	96
0,2	98	98	98	98
0,4	97	98	99	98
0,8	95	96	97	99

равномерного – серый, нормального – зеленый, Лапласа – ярко-зеленый, Рэлея - хаки и т.д. Таким образом, пользователь может сделать вывод о принадлежности входной выборки к какому-либо закону распределения, не руководствуясь выводами системы, а просто взглянув на обученную карту Кохонена. Также стоит отметить, что обучение сети по алгоритму Кохонена позволяет визуально разделить карту на четко очерченные области, отвечающие за идентификацию определенного закона, что не позволяют осуществить алгоритмы WTA и нейронного газа, где нейроны, отвечающие за идентификацию одного и того же закона более разрознены.

Наилучшие результаты идентификации получены при использовании сетки нейронов размером 20 на 20, обученной по алгоритмам Кохонена с гауссовским соседством или нейронного газа, при стартовом значении коэффициента обучения, лежащем в диапазоне [0.2, 0.8].

СПИСОК ЛИТЕРАТУРЫ

1. Хайкин С. Нейронные сети: полный курс, 2-е издание [пер. с англ.]. М.: Издательский дом "Вильямс", 2006. 1104 с.
2. Осовский С. Нейронные сети для обработки информации [пер. с польского И.Д. Рудинского]. М.: Финансы и статистика, 2002. 344с.
3. Лёзина И.В., Яшин Н.Ю. Автоматизированная система идентификации законов распределения нейронной сетью Кохонена // Современное общество, образование и наука: сб. науч. тр. по мат-лам Междунар. заоч. науч-практ. конф. 25 июня 2012 г.: в 3 ч. Тамбов: Изд-во ТРОО "Бизнес-Наука-Общество", 2012. Ч. 2. С. 88-90.
4. Лёзина И.В., Яшин Н.Ю. Идентификации законов распределения нейронной сетью Кохонена // Перспективные информационные технологии (ПИТ 2013): труды Международной научно-технической конференции [под ред. С.А. Прохорова]. Самара: Издательство Самарского научного центра РАН, 2013. С. 128-132.

THE AUTOMATED SYSTEM OF IDENTIFICATION OF DISTRIBUTION LAWS THE KOHONEN TWO-DIMENSIONAL NEURAL NETWORK

© 2014 I.V. Lyozina, N.Y.Yashin

Samara State Aerospace University after Academician S.P. Korolev
(National Research University)

This article describes the automated system developed by the authors to identify the distribution laws. The system uses a two-dimensional neural network (map) Kohonen learned by algorithms of WTA, Kohonen and neural gas. The article analyzes the results of identification and researches the dependence of the identification quality from network parameters.

Keywords: automated system, neural network, Kohonen network, identification, distribution law, the algorithm WTA, Kohonen algorithm, the algorithm neural gas.