



### Литература

1. Костин В.А., Снегуренко А.П. Теория и практика решения обратных задач прочности летательных аппаратов. Казань: Издательство Казанского государственного технического университета, 2004.
2. Осовский С. Нейронные сети для обработки информации. / Пер. с польского И.Д. Рудинского: М. Финансы и статистика, - 2002, - 344с. с ил.
3. Норман Дрейпер, Гарри Смит Прикладной регрессионный анализ, 3-е издание. М: Вильямс, 2007, - 912 с.
4. Мурга О.К., Еремеева А.А. Методы оптимизации: учебное пособие / О.К. Мурга, А.А.Еремеева. – М.: Изд-во Казан. гос. техн. ун-та, 2013. – 189 с.
5. В.Медведев, В.Потемкин. Нейронные сети. MATLAB 6. М.: ДИАЛОГ-МИФИ, 2001. - 630 с.

З.З. Мингалиев

## РАСПОЗНАВАНИЕ БИНАРНЫХ СИГНАЛОВ НА ОСНОВЕ НЕЙРОННОЙ СЕТИ ХЭММИНГА

(Казанский национальный исследовательский технический университет им. А.Н. Туполева - КАИ)

За последние годы в различных отраслях науки и техники отмечается тенденция к возрастанию интереса к искусственным нейронным сетям. Их популярность объясняется тем, что они позволяют эффективно решать задачи, которые плохо решаются классическими «аналитическими» методами.

По истечении многих лет забвения интерес к интеллектуальным системам быстро вырос за последнее время. Специалисты из несмежных областей (логистика, экология, физиология, архитектура, психология и другие) заинтригованы возможностями, предоставляемыми этой технологией, и ищут приложения им внутри своих дисциплин [1].

Искусственные нейронные сети сегодня находят применение в различных предметных областях: экономике, бизнесе, медицине, связи, интернете, вводе и обработке информации, безопасности и охраняемые системы, авионике и других.

Необходимо разработать программный комплекс, который будет выполнять задачу распознавания состояния окружающей среды. Входные данные: значения вероятностных рисков отдельных составляющих окружающей среды (воздуха, снега, почвы и биосреды). Выходные данные – номер одного из запомненных ранее состояний окружающей среды.

Распознаватель должен основываться на парадигме нейронной сети Хэмминга.

Алгоритм работы сети базируется на определении Хэммингово расстояния. Хэммингово расстояния – это количество отличающихся позиций в бинарных векторах. Результатом работы сети является нахождение образа с наименьшим расстоянием.



Отсутствие сигнала кодируется как (-1), наличие (1). Сеть состоит всего из 2-х слоев. Структура нейронной сети представлена на Рис. 1.

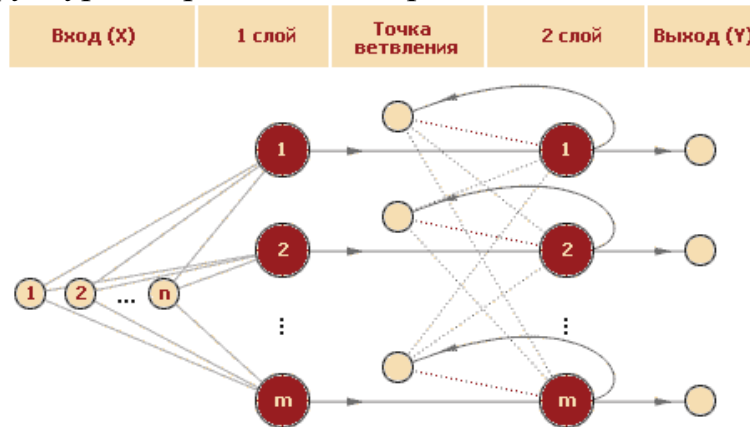


Рис. 1. Структурная схема сети Хэмминга

Нейронная сеть Хэмминга имеет следующий алгоритм работы:

1. На входы сети подается неизвестный вектор

$$X = \{x_i | i = 0, \dots, n\}, \quad (1)$$

исходя из которого рассчитываются состояния нейронов первого слоя (верхний индекс в скобках указывает номер слоя):

$$y_j^{(1)} = s_j^{(1)} = \sum_{i=0}^{n-1} w_{ij} x_i + T_j, \quad j = 0, \dots, m-1. \quad (2)$$

После этого полученными значениями инициализируются значения аксонов второго слоя:

$$y_j^{(2)} = y_j^{(1)}, \quad j = 0, \dots, m-1. \quad (3)$$

2. Вычислить новые состояния нейронов второго слоя:

$$s_j^{(2)}(p+1) = y_j^{(2)}(p) - \varepsilon \sum_{k=0}^{m-1} y_k^{(2)}(p), \quad k \neq j, \quad j = 0, \dots, m-1 \quad (4)$$

и значения их аксонов:

$$y_j^{(2)}(p+1) = f[s_j^{(2)}(p+1)], \quad j = 0, \dots, m-1. \quad (5)$$

Активационная функция  $f$  имеет вид порога, причем величина  $F$  должна быть достаточно большой, чтобы любые возможные значения аргумента не приводили к насыщению. На практике  $F$  обычно берется равным количеству примеров.

3. Проверить, изменились ли выходы нейронов второго слоя за последнюю итерацию. Если да – перейти к шагу 2. Иначе – конец [2].

Исходными данными задачи классификации является набор вероятностных рисков возникновения отрицательных изменений в отдельных природных сферах и в окружающей среде в целом. Часть таблицы вероятностных рисков, использованная при моделировании сети, представлена в таблице 1. Вероятностный риск представляет собой число от 0 до 1.



Таблица 3. Исходные данные задачи классификации

$P_{\text{возд}}$	$P_{\text{снег}}$	$P_{\text{почв}}$	$P_{\text{биоср}}$	$P_{\text{общ}}$	Класс состояния окружающей среды
0,3880	0,0680	0,2440	0,0550	0,3810	хорошее
0,3880	0,0680	0,2440	0,3000	0,3810	хорошее
...	...	...	...	...	...
0,7550	0,0680	0,2440	0,7500	0,8500	критическое
0,5230	0,0680	0,3670	0,0550	0,6460	плохое

Каждый набор можно представить в виде бинарных сигналов и обрабатывать по аналогии с обработкой изображений. Результатом обработки должно являться наименование класса состояния окружающей среды.

При распознавании изображений каждый пиксель представляет собой бинарный сигнал, где отсутствие сигнала (белый пиксель в черно-белом изображении) кодируется как  $-1$ , наличие сигнала (черный пиксель) как  $1$ .

Для перевода вероятностного риска в двоичное представление будет использована функция хеширования. Значение вероятностного риска, умноженное на  $10000$ , в виде числовой строки передаем на вход функции хеширования, а на выходе получаем хеш, содержащий  $128$  бит. Функция хеширования работает по алгоритму MD5.

В программном комплексе для реализации хеширования используется специальная криптографическая библиотека, интегрированная в среду разработки.

Был реализован программный комплекс на высокоуровневом языке программирования C#, реализующий сети Хэмминга. Были проведены вычислительные эксперименты с различными вероятностными рисками. Результаты работы программы представлены в таблице 2.

Таблица 2. Сравнительный анализ полученных данных

№ экс-та	$P_{\text{воздуха}}$	$P_{\text{снега}}$	$P_{\text{почвы}}$	$P_{\text{био-среды}}$	Полученный ответ сети	Ожидаемый результат	Отметка о совпадении
1	0.9495	0.8654	0.7891	0.8765	Критическое	Критическое	+
2	0.4800	0.1040	0.2510	0.4790	Среднее	Среднее	+
3	0.1200	0.3300	0.4220	0.4830	Плохое	Плохое	+
4	0.0700	0.3800	0.1020	0.0300	Среднее	Хорошее	-
5	0.7700	0.1800	0.3560	0.4090	Критическое	Критическое	+

При классификации состояний окружающей среды нейронная сеть Хэмминга правильно распознала состояние среды только в  $80\%$  случаев, причем доля успешных распознаваний сильно зависит от способа представления экологических показателей в виде бинарных сигналов.

Нейронная сеть Хэмминга требует дополнительных исследований для более успешного решения задачи классификации состояний окружающей среды.



## Литература

1. Каширина, И.Л. Нейросетевые технологии: учебно-методическое пособие для вузов / И.Л. Каширина. – Воронеж: ИПЦВГУ, 2008. – 72 с.
2. Короткий, С. Нейронные сети Хопфилда и Хэмминга / С. Короткий. – 69 с.

U.S. Otajonov

## DATA MINING METHODS

(Tashkent university of information technologies named after Muhammad al-Khwarizmi)

Data mining is information processing and identifying trends in it, which helps to make decisions. There are many different methods of data mining, modeling requests for processing and collecting information.

Intelligent analysis + data is one of the most relevant topics in the modern world. And this is not an exaggeration: business analysis + is only a small part of the scope of this powerful tool.

The principles of data mining have been known for a long time, but with the advent of big data, they have become even more widespread. When working with large data sets, relatively simple and straightforward statistics are no longer enough. Having about a million detailed records, you need to know not only the information about the location of several thousand records, but you also need to understand whether this category of records belongs to a specific group.

These requirements have created a complex data mining process. To solve problems, an analysis + data is required, in which a model is built to describe the information, and also a final report is created.

Big data combines three dimensions: volume, speed, and diversity.

Volume: when enterprises are overflowed with data of all types, the volume of which is growing and can easily reach terabytes and even petabytes.

Speed: sometimes 2 minutes is too long. In time-sensitive processes such as fraud detection, big data should be used as they enter the organization in order to maximize their value.

Diversity: big data can be of any structured or unstructured type, consisting of text data from sensors, sound, video recording, lists of visited web pages, protocol files and other data. When all these data of different types are analyzed, new ideas are born.

Intelligent analysis + data can be performed with relatively modest database systems and simple tools including creating your own or using ready-made software packages. Sophisticated data mining + is based on past experience and algorithms defined using existing software and packages, and various specialized tools are associated with various methods.