

УДК 66:681.3(075.8)

В. Л. Колесников, доктор технических наук, профессор (БГТУ)**КЛАССИФИКАЦИЯ ДАННЫХ МЕТОДОМ НЕЙРОННЫХ СЕТЕЙ**

В данной статье описан метод обнаружения в отчетных данных ранее неизвестных, нетривиальных знаний и закономерностей, необходимых для принятия решений в различных сферах человеческой деятельности. Сбор информации осуществлен на основе виртуального производственного комплекса. Собранные данные проанализированы методом искусственных нейронных сетей. Определено наполнение классов показателей качества продукции и классов загрязнения окружающей среды.

This article describes a method of detection in the accounting data of unknown, non-trivial knowledge, and patterns needed to make decisions in various spheres of human activity. Collection of information carried by the virtual production complex. The collected data were analyzed using artificial neural networks. Defined content classes of quality of products and classes of environmental pollution.

Введение. В информационной среде активно работающего предприятия накапливаются большие объемы разнообразных сведений: отчетов, сводных таблиц, диаграмм. Обычные отчеты не рассчитаны на поиск неочевидных правил и на извлечение неизвестных закономерностей, т. е. на генерацию новых знаний. Для решения подобных задач предназначена подсистема *DataMining* (интеллектуальный анализ данных). Эта подсистема включает в себя много методов классификации, кластеризации факторного анализа для обнаружения в сырых данных ранее неизвестных, нетривиальных, практически полезных, доступных для интерпретации знаний (закономерностей), необходимых для при-

нятия решений в различных сферах человеческой деятельности.

Основная часть. Добыча данных осуществлялась на разработанной системной математической модели производственного комплекса утилизации волокнистых отходов [1].

Значения внешних воздействий (независимых переменных) можно задавать с помощью генератора псевдослучайных чисел в заданных диапазонах. Значения зависимых переменных определяются синхронно по математической модели производственного комплекса в соответствии с условиями каждой строки.

Таким образом, получена главная таблица для обработки (11 независимых переменных, 5 – зависимых).

Таблица наблюдений

№ п/п	Дебет реки	Электролиты в реке	Цена энергии	Цена речной воды	Цена отходов	Степень помола	Концентрация массы	Скорость	Расход речной воды	Расход полимера	Расход волокна	Прочность	Пластичность	Влагопрочность	Загрязнение водоема	Загрязнение атмосферы
1	6	0,0015	0,07	0,65	50	16,02	0,29	164,26	73	97,88	1279	3,03	329	48,41	572	0,8
2	17	0,0008	0,1	1,2	25	15,62	0,55	111,47	63,57	66,54	1111	4,16	337	33,73	247	0,3
3	23	0,0027	0,09	0,9	25	27,17	0,35	131,88	66,83	50,75	1019	4,11	382	22,5	557	0,7
4	12	0,0006	0,12	1,1	35	19,81	0,4	115,28	40,78	87,68	143	3,92	364	9,2	326	0,7
5	15	0,0005	0,06	1,5	30	36,64	0,4	140,36	38,03	95,82	1267	9,29	274	16,76	271	0,7
6	13	0,0008	0,08	0,35	45	39,98	0,6	126,18	72,29	52,34	1361	9,05	272	27,74	116	0,7
7	11	0,0042	0,15	1,9	55	26,72	0,9	164,43	67,76	54,17	1001	5,04	107	13,61	185	0,7
8	17	0,0025	0,55	1,8	29,17	34,29	1,49	173,97	79,82	51,1	1166	12,99	194	46,4	294	0,5
9	4	0,0017	0,1	0,42	66,87	21,84	1,7	190,06	32,31	91,46	1302	4,97	146	51,69	443	0,8
10	3	0,0042	1,42	1,31	67,4	26,14	1,61	126,03	34,24	70,19	1293	4,6	187	27,51	315	0,7
11	14	0,0048	1,84	0,96	62,16	23,89	0,44	133,78	93,49	84,43	1427	13,26	25	25,35	488	0,4
12	7	0,0002	0,68	0,54	30,39	15,65	1,04	176,01	86,28	81,72	1301	7,83	291	35,57	41	0,6
13	3	0,0023	2,03	1,6	58,51	29,53	1,23	116,17	35,75	79,57	125	7,21	233	34,32	493	0,5
...
50	14	0,0049	1,28	0,12	58,74	35,15	0,37	121,98	78,25	83,89	146,9	3,56	207	42,13	124	0,5

Нервная система и мозг человека состоят из нейронов, соединенных между собой нервными волокнами. Нервные волокна способны передавать электрические импульсы между нейронами. Все процессы передачи раздражений от нашей кожи, ушей и глаз к мозгу, процессы мышления и управления действиями – все это реализовано в живом организме как передача электрических импульсов между нейронами.

Каждый нейрон имеет отростки нервных волокон двух типов – дендриты, по которым принимаются импульсы, и единственный аксон, по которому нейрон может передавать импульс. Аксон контактирует с дендритами других нейронов через специальные образования – синапсы, которые влияют на силу импульса.

Можно считать, что при прохождении синапса сила импульса меняется в определенное число раз, которая называется весом синапса. Импульсы, поступившие к нейрону одновременно по нескольким дендритам, суммируются. Если суммарный импульс превышает некоторый порог, нейрон возбуждается, формирует собственный импульс и передает его далее по аксону. Важно отметить, что веса синапсов могут изменяться со временем, а значит, меняется и поведение соответствующего нейрона.

Передаточные функции всех нейронов в нейронной сети фиксированы, а веса являются параметрами нейронной сети и могут изменяться. Некоторые входы нейронов могут являться входами нейронной сети, а некоторые выходы – внешними выходами нейронной сети. Подавая любые числа на входы нейронной сети, мы получаем набор чисел на выходах нейронной сети. Таким образом, работа нейронной сети состоит в преобразовании входного вектора в выходной вектор, причем это преобразование задается весами нейронной сети.

Обучить нейронную сеть – значит, сообщить ей то, чего мы от нее добиваемся. Этот процесс очень похож на обучение ребенка. Показав ребенку конкретное изображение, мы просим отыскать его на множестве других картинок. Если ответ неверен, мы сообщаем ребенку тот ответ, который мы хотели бы от него получить. Мы будем повторять процесс предъявления изображений снова и снова до тех пор, когда все будут твердо запомнены. Такой процесс называют «обучение с учителем».

Нейронная сеть, грамотным образом обученная, может с большой вероятностью правильно реагировать на новые, не предъявленные ей ранее данные.

Теоретическое обоснование нейросетевого моделирования базируется на теореме А. Н. Колмогорова, доказавшего, что любую непрерывную многомерную функцию на единичном от-

резке $[0; 1]$ можно представить в виде конечно-го числа одномерных.

Многие современные нейронные сети сконструированы из формальных нейронов, отдалённо напоминающих свой биологический прототип.

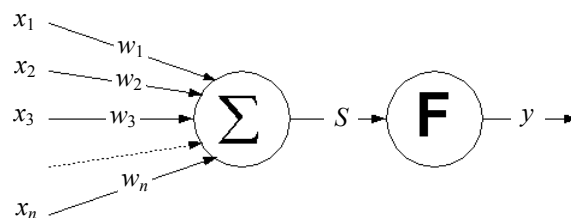


Рис. 1. Структура искусственного нейрона

Структура нейрона имеет вид, представленный на рис. 1, со следующими обозначениями:

- 1) x_1, \dots, x_n – значения, поступающие на входы (синапсы) нейрона;
- 2) w_1, \dots, w_n – веса синапсов, которые могут быть как тормозящими, так и усиливающими;
- 3) S – взвешенная сумма входных сигналов:

$$S = \sum_{i=1}^n w_i x_i - T; \quad (1)$$

- 4) T – порог нейрона (во многих моделях можно обойтись и без него);

- 5) F – функция активации нейрона, преобразующая взвешенную сумму в выходной сигнал:

$$y = F(S). \quad (2)$$

Вид функции активации может иметь различное математическое выражение, выбор которого определяется характером решаемых задач. Например, функция активации может быть:

– *линейная*

$$y = kS; \quad (3)$$

– *пороговая бинарная*

$$y = \text{sign}(S) = \begin{cases} 1, & S > 0; \\ 0, & S \leq 0, \end{cases} \quad (4)$$

или *биполярная*

$$y = \begin{cases} 1, & S > 0; \\ -1, & S \leq 0; \end{cases}$$

– *линейная ограниченная*

$$y = \begin{cases} p, & S > \alpha; \\ -p, & S < -\alpha; \\ S, & -\alpha \leq S \leq \alpha; \end{cases} \quad (5)$$

– сигмоидная

$$y = \frac{1}{1 + e^{-cS}}, \quad (6)$$

где $C > 0$ – коэффициент ширины сигмоиды по оси абсцисс;

– логистическая

$$y = e^{-e^{-(b_0 + b_1 S)}}, \quad (7)$$

где b_0 и b_1 – эмпирические коэффициенты, определяющие крутизну экспоненты;

– функция гиперболического тангенса

$$y = \text{th}(cS) = \frac{e^{cS} - e^{-cS}}{e^{cS} + e^{-cS}}; \quad (8)$$

– логарифмическая

$$y = \ln(S + \sqrt{S^2 + 1}). \quad (9)$$

Функция усиливает очень слабые сигналы и ослабляет очень сильные;

– радиально-базисная, имеющая вид функции Гаусса

$$y = e^{-\frac{S^2}{2\sigma^2}}, \quad (10)$$

где σ – среднеквадратичное отклонение нормального распределения, характеризующее ширину функции; S в этом случае определяется как расстояние между входным и весовым вектором: $S^2 = \sum_i (x_i - w_i)^2$.

Выбор количества слоев и типа активационной функции влияет на способность сети решать те или иные задачи [2].

В многослойной сети подобные функции отклика комбинируются друг с другом с помощью последовательных линейных комбинаций и применения нелинейных функций активации.

После того как определено число слоев и число элементов в каждом из них, нужно найти значения весовых коэффициентов и порогов для каждого составляющего нейрона, которые бы минимизировали ошибку прогноза, выдаваемого сетью. Для того чтобы оценить, насколько верное решение сделала ИНС, определяется среднеквадратичная ошибка сети как разница между эталонным и выходным вектором.

Обучение многослойного персептрона производится с помощью алгоритма обратного распространения, являющегося модификацией градиентного спуска.

Здесь оказывается очень полезным понятие поверхности ошибок. Поверхность ошибок в нейронных сетях имеет сложное строение и

имеет локальные минимумы, плоские участки, седловые точки и длинные узкие овраги.

Аналитическими средствами в таких условиях найти глобальный минимум невозможно, поэтому процесс обучения ИНС, по сути дела, заключается в исследовании этой поверхности. Отталкиваясь от случайной начальной конфигурации весов и порогов, алгоритм оптимизации начинает передвижение к минимуму, используя на каждом шаге вектор градиента.

Для выходного слоя определение ошибки нейрона j тривиально:

$$\delta_j = (y_j - y_j^*) F'(S_j) \quad (11)$$

и напоминает систему поощрений-наказаний, используемую при обучении однослойных сетей. Для каждого предыдущего слоя ошибка определяется рекурсивно через ошибку следующего слоя:

$$\delta_j = \left(\sum_{i=1}^m \delta_i w_{ij} \right) F'(S_j), \quad (12)$$

т. е. для каждого j -го нейрона ошибки следующего слоя как бы распространяются к нему обратно сквозь соответствующие веса. Этот механизм обратного распространения дополнен традиционными для многих градиентных методов оптимизации процедурами оценки вектора кратчайшего спуска, изменения величины шага пропорционально крутизне склона и пр.

Метод Ньютона – Гаусса, в общих чертах, использует матрицу Якоби J производных первого порядка функции F для нахождения вектора x значений параметра, который минимизирует остаточные суммы квадратов (сумму квадратов невязок предсказанных значений от наблюдаемых).

Функция $f(\vec{x})$, подлежащая минимизации, представляет собой сумму квадратов:

$$f(\vec{x}) = \left\| \vec{F}(\vec{x}) \right\|^2 = \sum_{k=1}^n (g_k(\vec{x}) - a_k)^2. \quad (13)$$

Под $F(\vec{x})$ имеется в виду вектор-столбец $[g_k(\vec{x}) - a_k]_{k=1}^n$.

Подобного типа задачи широко распространены и имеют ряд практических применений, особенно при подборе модельной функции. Пусть $J(\vec{x})$ – матрица Якоби для функции $F(\vec{x})$, т. е.

$$J(\vec{x}) = \left[\frac{\partial g_i(\vec{x})}{\partial g_j} \right]_{i=1, j=1}^{n, m}, \quad (14)$$

где \vec{x} из R^m .

Выбирая некоторое начальное значение \bar{x}_0 , последовательные приближения находят следующим образом:

$$\bar{x}_{j+1} = \bar{x}_j - (J^T(\bar{x})J(\bar{x}))^{-1} J^T(\bar{x})F(\bar{x}). \quad (15)$$

Пусть необходимо определить экстремум функции многих переменных $f(\bar{x}): R^m \rightarrow R$. Это равносильно нахождению точки $x^* : \text{grad}[f](x^*) = 0$. Если для решения этой задачи использовать итерационный метод Ньютона (метод касательных), то формула обновления для \bar{x}_j выглядит следующим образом:

$$\bar{x}_{j+1} = \bar{x}_j - H^{-1}(\bar{x}_j)\text{grad}[f](\bar{x}_j). \quad (16)$$

Здесь под $H(\bar{x})$ имеется в виду матрица Гесса функции $F(\bar{x})$, т. е. матрица вторых производных:

$$H(\bar{x}) = \left[\frac{\partial^2 f(\bar{x})}{\partial x_i \partial x_j} \right]_{i=1, j=1}^m. \quad (17)$$

Когда рассматривается задача наименьших квадратов

$$f(\bar{x}) = \frac{1}{2} \sum_{i=1}^m (\varphi_i(\bar{x}) - F_i)^2 \rightarrow \min, \quad (18)$$

градиент и матрица Гесса для функции $F(\bar{x})$ принимают особый вид

$$\begin{aligned} \text{grad}[f](\bar{x}) &= J^T(\bar{x})F(\bar{x}), \\ H(\bar{x}) &= J^T(\bar{x})J(\bar{x}) + Q(\bar{x}). \end{aligned} \quad (19)$$

Если использовать итерационный процесс Ньютона для минимизации $F(\bar{x})$, то формула для обновления \bar{x}_j будет следующая:

$$\bar{x}_{j+1} = \bar{x}_j - \left[J^T(\bar{x})J(\bar{x}) + \sum_{i=1}^m \bar{F}_i(\bar{x})H_i(\bar{x}) \right]^{-1} \times J^T(\bar{x})\bar{F}(\bar{x}). \quad (20)$$

Метод Ньютона – Гаусса строится на предположении о том, что $\|J^T(\bar{x})J(\bar{x})\| \gg \|Q(\bar{x})\|$, т. е. слагаемое $J^T(\bar{x})J(\bar{x})$ доминирует над $Q(\bar{x})$. Также такое приближение возможно при условии, что $\|Q(\bar{x})\|$ близко к 0. Это требование не соблюдается, если минимальные невязки велики, т. е. если норма $\|F(\bar{x})\|$ сравнима с максимальным собственным значением матрицы $J^T(\bar{x})J(\bar{x})$. В противном случае пренебрегаем $Q(\bar{x})$ и получаем итерационную процедуру с формулой для обновления \bar{x}_{j+1} :

$$\bar{x}_{j+1} = \bar{x}_j - (J^T(\bar{x})J(\bar{x}))^{-1} J^T(\bar{x})F(\bar{x}). \quad (21)$$

В стандартном итерационном методе Ньютона на каждой итерации требуется вычисление и обращение матрицы Гесса, что зачастую является достаточно сложным процессом. В методе Ньютона – Гаусса подобная необходимость отпадает, причем скорость сходимости также может достигать квадратичной, хотя вторые производные и не учитываются.

Поскольку в наших данных отсутствуют отрицательные значения элементов, то в качестве функции активации мы выбрали сигмоид.

На рис. 2 показано диалоговое окно приложения JMPv8 с выведенной архитектурой полученной нейронной сети – однослойный персептрон с четырьмя скрытыми нейронами.

На рис. 3, 4 приведены только некоторые выборочные иллюстрации анализа данных таблицы. В действительности получено и обработано более 100 подобных картинок, которые не поместились в статью, но позволили классифицировать мощности связей независимых переменных для всех пяти оценочных признаков.

Оказалось, что пластичность образует только один класс, который включает дебет реки, цену отходов и расход полимера с доминирующим влиянием последнего.

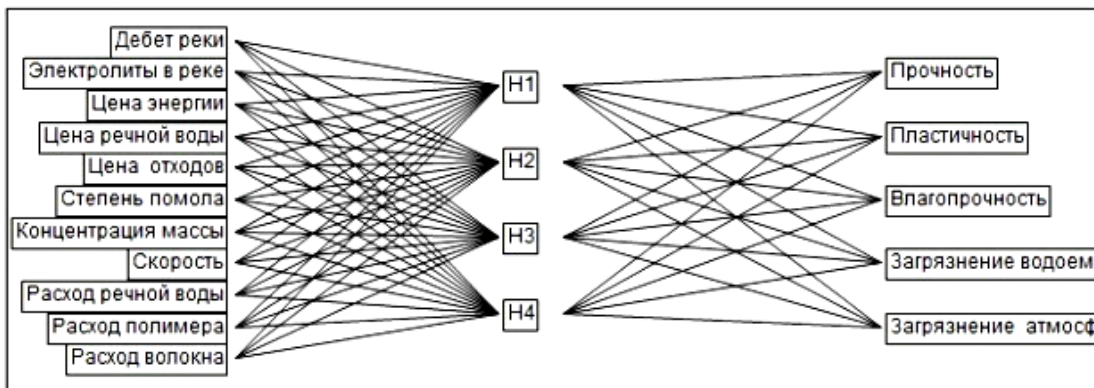


Рис. 2. Архитектура анализируемой нейронной сети

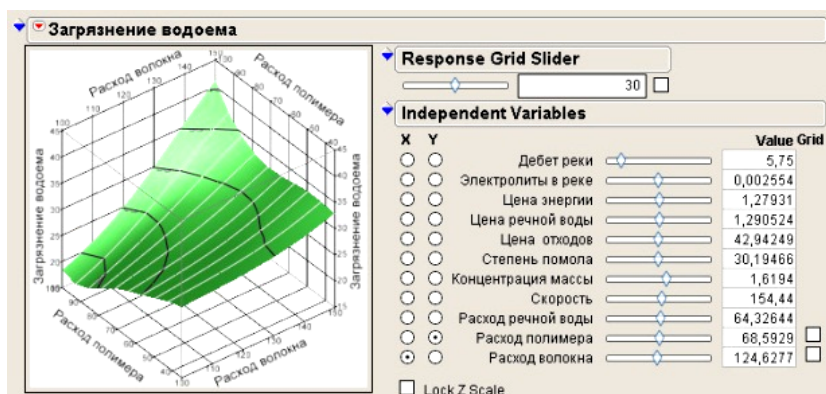


Рис. 3. График поверхности загрязнения водоема

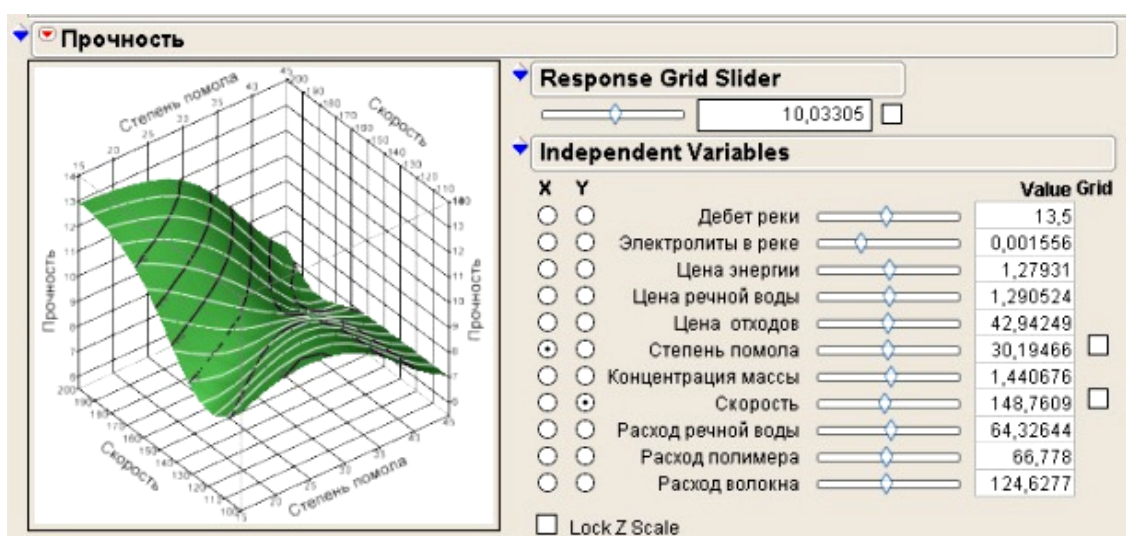


Рис. 4. График поверхности прочности продукции

Класс прочности более многочислен и включает в себя расход полимера, расход волокна, степень помола, скорость и концентрацию массы, т. е. все технологические параметры, а из экономических – только цену отходов и цену речной воды. Наиболее сильные связи со скоростью и концентрацией массы.

Класс влаготочности обладает самой сложной структурой, так как один и тот же параметр входит в несколько подклассов с максимальными и минимальными значениями и сам образует несколько подклассов. Главными компонентами класса являются расход полимера и степень помола, а также содержание электролитов в речной воде и цена энергии.

Загрязнение водоема наиболее тесную связь имеет с расходами упрочняющих добавок и дебетом реки.

Самыми мощными связями в классе загрязнения атмосферы обладают расход полимера и содержание электролитов в речной воде.

Заключение. Возможности программного обеспечения системного анализа позволяют

получать трехмерные поверхности зависимых переменных, выстраивать иерархию взаимоотношений между признаками. Применяя метод сечений, устанавливая флажки на соответствующих осях и перемещая ползунки на слайтерах значений независимых переменных, можно классифицировать многомерные данные производственных процессов.

Отсюда можно выявить слабые места технологии, необходимые объемы инвестиций и составить график реализации программы модернизации.

Литература

1. Колесников, В. Л. Математические основы компьютерного моделирования химико-технологических систем: учеб. пособие для вузов / В. Л. Колесников. – Минск: БГТУ, 2004. – 312 с.

2. Колесников, В. Л. Системный анализ производственных процессов в полиграфии / В. Л. Колесников. – Минск: БГТУ, 2011. – 352 с.

Поступила 05.03.2012