

2. Коршунов, А. И. Комплексное решение проблемы нормирования машиностроительного изделия с использованием теории конструктивно-технологической сложности / А. И. Коршунов, Б. А. Якимович // Интеллектуальные системы в производстве : науч.-практ. журн. – 2003. – № 2. – С. 86–105.

3. Суслов, А. Г. Технология машиностроения : учеб. для студентов машиностроит. специальностей вузов / А. Г. Суслов. – М. : Машиностроение, 2004. – 400 с.

УДК 502.55:66+004.032.26

М. Г. Кургузкин, кандидат технических наук, доцент;  
М. А. Корепанов, кандидат технических наук, доцент;  
В. А. Тенев, доктор физико-математических наук, профессор  
Ижевский государственный технический университет

### АНАЛИЗ БИОЛОГИЧЕСКОГО ТЕСТИРОВАНИЯ ВЛИЯНИЯ ХИМИЧЕСКИХ ВЕЩЕСТВ ИЗ ПОЧВЫ В ЗОНЕ ДЕЙСТВИЯ ОБЪЕКТА УХО Г. КАМБАРКИ

*Рассмотрено влияние содержания химических веществ на биологические показатели. Применение методов извлечения знаний из данных (нейронных сетей, деревьев решений) позволяет установить связи между содержанием химических веществ и параметров биотестов.*

Для контроля влияния промышленной деятельности на окружающую среду применяется биологическое тестирование. В зоне действия объекта по уничтожению химического оружия в г. Камбарке организован мониторинг содержания химических веществ в почве. Химический анализ веществ, содержащихся в почве, проводится для 84 точек. Для этих же точек осуществляются биологические тесты. Виды биотестов соответствуют табл. 1.

Таблица 1. Виды биотестов

1	2	3	4	5
Дафнии, БКР96-10	Цериодафнии, БКР96-10	Хлорелла	Инфузории	Бактерии

Перечень химических веществ содержится табл. 2.

Таблица 2. Анализируемые химические вещества в почве

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
Люизит	Оксид люизита	2-ХВА	pH	Cl – ион	Sr	Pb	As	Zn	Cu	Ni	Co	Fe <sub>2</sub> O <sub>3</sub>	MnO	Cr	V	TiO <sub>2</sub>

Связь между показателями биотестов и содержанием веществ в почве в контрольных точках определялась на основе методов извлечения знаний из данных – нейронных сетей и деревьев решений. Эти методы лучше приспособлены для обработки многопараметрических зависимостей, чем традиционные статистические методы, так как не связаны с обоснованием законов распределения случайных величин при наблюдениях.

#### 1. Нейросетевые модели

Для извлечения знаний из системы данных и для решения задач аппроксимации сложных функций многих переменных широкое применение нашли однонаправленные многослойные нейронные сети [1]. Важным свойством нейронных сетей является способность к обучению и обобщению полученных знаний. Обученная на ограниченном множестве обучающих выборок, сеть обобщает накопленную информацию и выдает реакцию на данные, не применявшиеся при обучении.

Нейронная сеть осуществляет нелинейное преобразование вектора  $\mathbf{x}$  в вектор  $\mathbf{y}$ :

$$\mathbf{y} = \Phi(\mathbf{W}, \mathbf{x}),$$

где  $\mathbf{W}$  – матрица коэффициентов преобразования, определяемая в процессе обучения сети.

Многослойная нейронная сеть состоит из входного и выходного слоев, а также из нескольких внутренних (скрытых) слоев. Входной слой имеет размерность входного вектора  $\mathbf{x} = [x_1, \dots, x_n]$ . Обычно размерность вектора  $\mathbf{x}$  увеличивают еще на единицу, добавляя  $x_0 = 1$ . Это делается для включения величины смещения функции активации в множество весовых коэффициентов. Каждый нейрон первого скрытого слоя ( $k = 1$ ) осуществляет суммирование входящих сигналов

$$u_i^1 = \sum_{j=0}^n w_{ij}^1 x_j, \quad i = \overline{1, N_1}.$$

Выходной сигнал нейрона преобразуется с помощью функции активации

$$z_i^k = G(u_i^k), \quad i = \overline{1, N_k}; \quad k = \overline{1, Kc},$$

где  $N_k$  – число нейронов в  $k$ -м слое;  $Kc$  – число слоев.

В качестве функции активации используется сигмоида  $G(s) = \frac{1}{1 + \exp(-\beta s)}$ .

Производная от этой функции выражается через значения самой функции  $\frac{dG}{ds} = \beta G(s)(1 - G(s))$ .

Выходные преобразованные сигналы суммируются на последующем слое и так далее, до последнего выходного слоя.

$$u_i^k = \sum_{j=0}^{N_{k-1}} w_{ij}^k z_j^{k-1}, \quad z_i^k = G(u_i^k), \quad i = \overline{1, N_k}, \quad k = 1, Kc,$$

так что  $\mathbf{z}^0 = \mathbf{x}$ ,  $\mathbf{y} = \mathbf{z}^{Kc}$ .

Построенная таким образом нейронная сеть содержит весовые коэффициенты  $w_{ij}^k$ ,  $i = \overline{1, N_k}$ ,  $j = \overline{0, N_{k-1}}$ ,  $k = 1, Kc$ , требующие определения в процессе обучения.

Для обучения используется система данных, представляющая собой набор наблюдаемых точек  $(\mathbf{x}^j, \mathbf{f}^j)$ ,  $j = \overline{1, p}$ , где  $\mathbf{x}$ ,  $\mathbf{f}$  – входной вектор и вектор функции соответственно. Система данных из  $p$  точек делится на две выборки: обучающую  $(\mathbf{x}^j, \mathbf{f}^j)$ ,  $j = \overline{1, h}$  и проверочную  $(\mathbf{x}^j, \mathbf{f}^j)$ ,  $j = \overline{h+1, p}$ . Весовые коэффициенты нужно подобрать таким образом, чтобы они обеспечили минимальное отклонение расчи-

тываемых в сети значений  $y$  от имеющихся  $f$ , т. е. давали бы минимум целевой функции

$$F(\mathbf{W}) = \frac{1}{2} \sum_{i=1}^m (y_i - f_i^q)^2 \Rightarrow \min.$$

Здесь  $\mathbf{W}$  – матрица коэффициентов  $w_{ij}^k, i = \overline{1, N_k}, j = \overline{0, N_{k-1}}, k = \overline{1, Kc}, q$  – номер предъявляемой для обучения пары из выборки  $(\mathbf{x}^q, \mathbf{f}^q), q = \overline{1, h}$ .

Для обучения нейронной сети (настройки коэффициентов  $\mathbf{W}$ ) применялся либо алгоритм обратного распространения, либо генетический алгоритм.

## 2. Метод деревьев решений

К группе методов, получивших в последнее время широкое применение в самых разных областях, относятся алгоритмы деревьев решений. Эти алгоритмы применяются для решения задач классификации. Метод деревьев решений может применяться для целевой переменной, имеющей булев или категориальный тип. Такие переменные содержат значения, принадлежащие некоторому конечному множеству без определенного отношения порядка на нем.

Деревья решений – это способ представления правил в иерархической, последовательной структуре, где каждому объекту соответствует единственный узел, дающий решение. Под правилом понимается логическая конструкция, представленная в виде *if A then B* ( $A \rightarrow B$ ).

Пусть задано некоторое обучающее множество  $X$ , содержащее объекты, каждый из которых характеризуется  $m$  атрибутами и один из них указывает на принадлежность объекта к определенному классу. Это множество обозначим  $X = \{\mathbf{x}^j, C_k^j\}, j = \overline{1, N}; k = \overline{1, K}$ , где каждый элемент этого множества описывается атрибутами  $\mathbf{x} = (x_i), i = \overline{1, m-1}$  и принадлежит одному из классов  $C_k$ . Количество примеров в множестве равно  $N$  является мощностью этого множества  $|X|$ . Через  $\{C_k\}$  обозначается множество классов.

Каждое множество  $X$  на любом этапе разбиения характеризуется следующими показателями:

- 1) множество  $X$  содержит несколько объектов, относящихся к одному классу  $C_k$ . В этом случае множество  $X$  является листом, определяющим класс  $C_k$ ;
- 2) множество  $X$  не содержит ни одного объекта ( $X = \emptyset$ ). В данной ситуации необходимо возвратиться к предыдущему этапу разбиения;
- 3) множество  $X$  содержит объекты, относящиеся к разным классам. Такое множество является пригодным для разбиения на некоторые подмножества. Для этого выбирается одна из переменных и в соответствии с правилами  $x \leq w, x > w$  множество  $X$  разбивается на два подмножества. Этот процесс рекурсивно продолжается до тех пор, пока конечное множество не будет состоять из примеров, относящихся к одному и тому же классу. Данная процедура лежит в основе многих алгоритмов построения деревьев решений (метод деления и захвата) [2]. Построение дерева решений происходит сверху вниз. Сначала создается корень дерева, затем потомки корня и т. д.

Для построения дерева на каждом внутреннем узле необходимо найти такое условие, которое бы разбивало множество, ассоциированное с этим узлом на подмножества. Для такой проверки  $x \leq w$ ,  $x > w$  должен быть выбран один из атрибутов. Выбранный атрибут должен разбить множество так, чтобы получаемые в итоге подмножества состояли из объектов, принадлежащих к одному классу или были максимально приближены к этому, т. е. количество объектов из других классов в каждом из этих множеств было как можно меньше.

Одним из способов выбора наиболее подходящего атрибута является применение теоретико-информационного критерия.

Задача заключается в построении иерархической классификационной модели в виде дерева из множества объектов  $X = \{x^j, C_k^j\}$ ,  $j = \overline{1, N}$ ;  $k = \overline{1, K}$ . На первом шаге имеется только корень и исходное множество, ассоциированное с корнем.

После проверки в результате разбиения получаются два (по числу условий ( $x \leq w$ ,  $x > w$ )) подмножества и соответственно создаются два потомка корня, каждому из которых поставлено в соответствие свое подмножество, полученное при разбиении множества  $X = \{x^j, C_k^j\}$ ,  $j = \overline{1, N}$ ;  $k = \overline{1, K}$ . Затем эта процедура рекурсивно применяется ко всем подмножествам (потомкам корня) и т. д. Любой из атрибутов можно использовать неограниченное количество раз при построении дерева.

Определим в качестве проверки  $t$  какую-либо переменную, принимающую значения  $x_i \begin{matrix} > \\ < \end{matrix} (w_{ij})$ ,  $i = \overline{1, m-1}$ ;  $j = \overline{1, n}$ . Тогда разбиение  $X$  по проверке  $t$  дает соответствующие подмножества  $X_{ij}$ ,  $i = \overline{1, m-1}$ ;  $j = \overline{1, n}$ . Критерий выбора определяется информацией о том, каким образом классы распределены в множестве  $X$  и его подмножествах, получаемых при разбиении по  $t$ .

Обозначим  $P_{liq}^k$ ,  $k = \overline{1, K}$ ;  $i = \overline{1, m-1}$ ;  $q = \overline{1, n}$ ;  $l = 1, 2$  – вероятность принадлежности классу  $k$  по атрибуту  $i$  и  $q$ -му пороговому значению  $x_i \begin{matrix} > \\ < \end{matrix} (w_{ij})$ ,  $i = \overline{1, m-1}$ ;  $j = \overline{1, n}$ , а  $P_l^k$  – вероятность попадания в класс  $k$ . В качестве меры математического ожидания информации, необходимого для определения класса объекта из множества  $X$ , рассматривается энтропия Шеннона

$$H_l = - \sum_{k=1}^K P_l^k \log_2 P_l^k.$$

Величина энтропии характеризует степень нечеткости системы данных [3].

Разбиению множества  $X$  по проверке  $t$  соответствует выражение для энтропии

$$H_{iq} = - \sum_l \sum_{k=1}^K P_{liq}^k \log_2 P_{liq}^k.$$

Критерием выбора является выражение, соответствующее максимальному упорядочиванию данных по классам

$$H_{iq} \rightarrow \min.$$

Минимальное значение энтропии  $H_{iq}$  отвечает максимуму вероятности появления одного из классов. Выбранный номер переменной  $x_i$  и пороговое значение  $w_{iq}$ , минимизирующие  $H_{iq}$ ,  $(i, w_{iq}) = \text{ArgMin } H_{iq}$  используются для проверки значения переменной  $x_i$  по условию  $x_i \begin{matrix} > \\ < \end{matrix} (w_{iq})$ , и дальнейшее движение по дереву производится в зависимости от полученного результата.

Данный алгоритм применяется к полученным подмножествам и позволяет продолжить рекурсивно процесс построения дерева, до тех пор, пока в узле не окажутся примеры из одного класса. Если в процессе работы алгоритма получен узел, ассоциированный с пустым множеством (т. е. ни один пример не попал в данный узел), то он помечается как лист, и в качестве решения листа выбирается наиболее часто встречающийся класс у непосредственного предка данного листа.

Пороговые величины для переменной  $x_i^s, i = \overline{1, m-1}; s = \overline{1, N_i}$  определяются выражением

$$w_{ij} = x_i^{\min} + (x_i^{\max} - x_i^{\min}) \frac{j}{M_i}, i = \overline{1, m}; j = \overline{1, M_i - 1},$$

где  $x_i^{\max}, x_i^{\min}$  – максимальные и минимальные значения переменной  $x_i^s, i = \overline{1, m-1}; s = \overline{1, N_i}; N_i$  – количество элементов разбиваемого подмножества.

Вместо вероятности определим возможность принадлежности классу  $k$   $F_{liq}^k, k = \overline{1, K}; i = \overline{1, m}; q = \overline{1, n}; l = 1, 2$  – по атрибуту  $i$  и  $q$ -му пороговому значению  $x_i \begin{matrix} > \\ < \end{matrix} (w_{ij}), i = \overline{1, m}; j = \overline{1, n}$ , а  $F_l^k$  – возможность попадания в класс  $k$ . Мера возможности определяется выражением  $F = \frac{f_k}{\max_k f_k}$ .

Возможностная мера нечеткости системы данных определяется по формулам [3]:

$$U(F) = \frac{1}{L_F} \int_0^{L_f} \log_2 |c(F, L)| dL \text{ или}$$

$$U(F) = \frac{1}{L_F} \sum_{k=1}^{K-1} (L_{k+1} - L_k) \log_2 |c(F, L_{k+1})|.$$

Функция  $c(F, L) = \{F_k \geq L\}$  называется функцией уровня  $L$ . Возможностный подход для определения критерия разбиения бывает предпочтительнее в случае ограниченного количества точек в обучающей выборке.

### 3. Снижение размерности системы данных (метод главных компонент)

При большом количестве влияющих факторов затруднительным является установление зависимости между ними и зависимыми показателями. Нами рассматривается влияние концентрации 17 химических веществ в почве на 5 биологических

показателей по результатам отборов проб в 84 точках. При наличии корреляции между влияющими факторами хороший эффект дают методы снижения размерности данных.

Методы снижения размерности предназначены для представления системы, характеризующейся переменными  $\mathbf{x} \in R^n$ , в координатном пространстве меньшей размерности  $\mathbf{y} \in R^m$ , причем  $m \ll n$ . Для этого можно использовать нелинейные методы (многомерное шкалирование) либо линейные (метод главных компонент). В методе главных компонент каждая координатная ось является линейной комбинацией исходных переменных системы  $\mathbf{x} \in R^n$ . Преобразование по методу главных компонент имеет вид

$$\mathbf{y} = \mathbf{W}\mathbf{x},$$

где  $\mathbf{W} \in R^{m \times n}$  – матрица преобразования. Матрица  $\mathbf{W}$  строится таким образом, что

первая главная компонента  $y_1(\mathbf{x}) = \sum_{j=1}^n w_{1j}x_j$  обладает наибольшей дисперсией.

Вторая главная компонента имеет наибольшую дисперсию из оставшихся и т. д. Размерность  $m$  выбирается таким образом, чтобы выходное пространство  $\mathbf{y} \in R^m$  сохранило наиболее важную информацию об исходной системе. Преобразование по методу главных компонент заменяет большое количество взаимно коррелирующих данных статистически независимыми компонентами с определенным вкладом.

Представим систему данных  $\mathbf{x}_k, k = \overline{1, p}$  последовательностью векторов  $\mathbf{z}_k, k = \overline{1, p}$ , образующих матрицу  $\mathbf{Z} = [\mathbf{z}_k]$ ,  $p$  – число наблюдений. Векторы  $\mathbf{z}_k, k = \overline{1, p}$  обозначают случайные векторы с нулевым средним значением

$$z_{ik} = \frac{x_{ik} - \bar{x}_i}{\sigma_i}, i = \overline{1, n}, k = \overline{1, p},$$

$$\text{где } \bar{x}_i = \frac{\sum_{k=1}^p x_{ik}}{p}, \sigma_i = \sqrt{\frac{\sum_{k=1}^p (x_{ik} - \bar{x}_i)^2}{p}}.$$

По значениям векторов вычисляется матрица корреляции

$$\mathbf{S} = \frac{1}{p} \mathbf{Z}\mathbf{Z}^T.$$

Матрицу преобразования  $\mathbf{W}$  образуют собственные векторы матрицы  $\mathbf{S}$ . Собственные векторы и собственные значения  $\lambda_i$  связаны соотношением

$$\mathbf{S}\mathbf{w}_i = \lambda_i \mathbf{w}_i, i = \overline{1, n}.$$

Так как матрица  $\mathbf{S}$  симметричная и положительно определенная, то ее собственные числа действительные и положительные. Последовательность собственных чисел можно упорядочить в порядке убывания

$$\lambda_1 > \lambda_2 > \dots > \lambda_n \geq 0.$$

Соответствующие собственные векторы упорядочиваются в той же последовательности и образуют матрицу преобразования

$$\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m]^T, m \leq n.$$

Вектор  $\mathbf{y} = \mathbf{W}\mathbf{x} = [y_1, \dots, y_m]^T$  является вектором главных компонент. Погрешность реконструкции вектора данных  $\mathbf{x} = \mathbf{W}^T \mathbf{y}$  определяется как сумма отброшенных собственных чисел  $\varepsilon = \sum_{i=m+1}^n \lambda_i$ .

Преобразование по методу главных компонент определяет корреляцию между переменными, образующими входное множество. Если переменные коррелируют между собой, то для определения всех данных достаточно взять меньшее число переменных.

Для нахождения матрицы преобразования можно применять стандартные методы нахождения собственных векторов, например продолжение декомпозиции  $QR$ .

#### 4. Анализ данных

Рассмотрены данные, содержащие показатели биотестов (табл. 1) и содержание загрязняющих веществ в почве (табл. 2) в 84 контрольных точках за 2005 г., первое и второе полугодия 2006 г. Обработка данных дала спектр собственных значений  $\lambda$  корреляционной матрицы, приведенный на рис. 1.

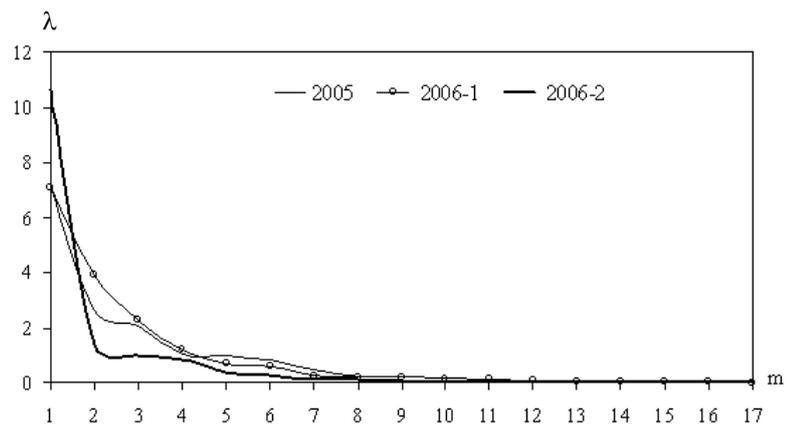


Рис.1. Спектр собственных значений

Как следует из рис. 1, имеется значимая корреляция между влияющими факторами. Погрешность реконструкции вектора данных представлена на рис. 2.

$\varepsilon, \%$

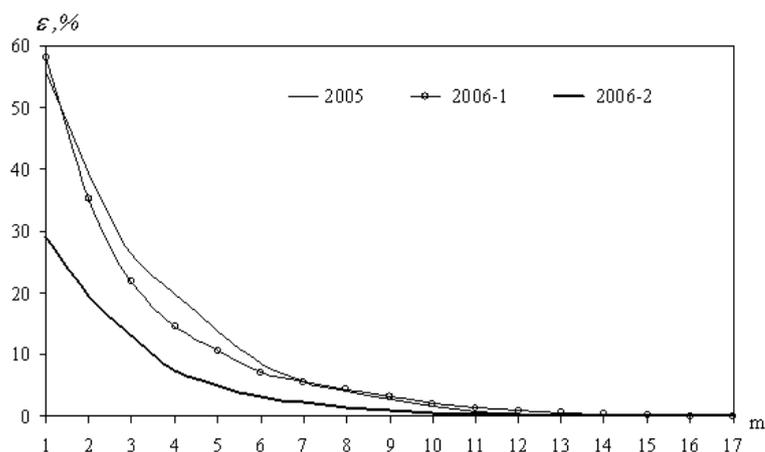


Рис. 2. Ошибка отсечения собственных векторов

Отбрасывание 10 собственных векторов приводит к ошибке порядка 5 %.

Показания биологических индикаторов разделены на два класса: 0 – нет реакции; 1 – есть реакция. Полученная задача классификации решалась с применением нейронной сети и деревьев решений. В табл. 3 показаны результаты классификации для 25 %-ной проверочной выборки. Номера индикаторов соответствуют обозначениям из табл. 1.

Таблица 3. Классификация нейронной сетью

№ индикатора	2005		2006-1		2006-2	
	нет реакции	есть реакция	нет реакции	есть реакция	нет реакции	есть реакция
1	–	–	–	–	–	–
2	–	–	–	–	–	–
3	0	96	0	100	0	100
4	100	0	–	–	100	0
5	93	0	100	0	100	0

В табл. 3 приведен процент правильно классифицированных точек нейросетью. Нейронная сеть не показала правильное разделение по классам.

Таблица 4. Классификация деревьями решений

№ индикатора	2005		2006-1		2006-2	
	нет реакции	есть реакция	нет реакции	есть реакция	нет реакции	есть реакция
1	–	–	–	–	–	–
2	–	–	–	–	–	–
3	12	67	90	67	15	75
4	80	10			83	25
5	78	30	92	10	70	10

Метод деревьев решений дает несколько лучший результат (табл. 4). Методом деревьев получены классификационные правила следующего вида:

2005 г.

3 тест, 34 точки

*if*  $y_1 \geq -0,26$  *then*  $B = 1$

*if*  $y_1 < -0,26$  *then*  $B = 0$

4 тест, 71 точка

*if*  $y_1 \geq 3,43$  *then*  $B = 1$

*if*  $y_1 < 3,43$  *then*  $B = 0$

5 тест, 77 точек

*if*  $y_1 \geq 3,43$  *then*  $B = 1$

*if*  $y_1 < 3,43$  *then*  $B = 0$

2006 -1

3 тест, 28 точек

*if*  $y_1 \geq 2,78$  *then*  $B = 1$

*if*  $y_1 < 2,78$  *then*  $B = 0$

4 тест, 2 точки

5 тест, 64 точки

*if*  $y_1 \geq 3,50$  *then*  $B = 1$

*if*  $y_1 < 3,50$  *then*  $B = 0$

2006 \_2,

3 тест, 19 точек

*if*  $y_1 \geq 4,16$  *then*  $B = 1$

*if*  $y_1 < 4,16$  *then*  $B = 0$

4 тест, 42 точки

*if*  $y_1 \geq 2,09$  *then*  $B = 1$

*if*  $y_1 < 2,09$  *then*  $B = 0$

5 тест, 43 точки

*if*  $y_1 \geq 3,16$  *then*  $B = 1$

*if*  $y_1 < 3,16$  *then*  $B = 0$

Остальные правила являются недостаточно представительными.

Переменная  $y_1$  соответствует первому главному компоненту. В каждой точке  $p$

определяется величина  $y_{1k} = \sum_{j=1}^n w_{j1} z_{jk}$ ,  $k = \overline{1, p}$ , являющаяся интегральным показателем

загрязнения в данной точке. Через измеряемые показатели, содержащиеся в базе данных, интегральный показатель рассчитывается следующим образом:

$$y_{1k} = \sum_{j=1}^n \frac{x_{jk} - \bar{x}_j}{\sigma_j} w_{j1}, \quad j = \overline{1, n}; \quad k = \overline{1, p},$$

где  $j$  – номер загрязняющего вещества;  $\bar{x}_j = \frac{\sum_{k=1}^p x_{jk}}{p}$ ,  $\sigma_j = \sqrt{\frac{\sum_{k=1}^p (x_{jk} - \bar{x}_j)^2}{p}}$  среднее

по точкам и дисперсия. Эти величины вместе с коэффициентами  $w_{ji}$  находятся в табл. 5.

Интегральный показатель показывает вклад каждого из 17 веществ в общее загрязнение. Например, четвертое вещество РН входит с коэффициентом 0,2691, шестое AS с 0,149 и т. д.

Таблица 5. Коэффициенты интегрального показателя

$\bar{x}_j$	$\sigma_j$	$w_{ji1}$
<b>2005</b>		
0,002	0,393	0,0325
0,000	0,000	0
0,036	0,393	0,0325
4,642	0,232	0,2691
10,490	0,125	0,0736
11,070	0,123	0,149
27,267	0,110	0,1128
104,980	0,185	0,3163
39,468	0,233	0,3402
201,905	0,241	0,1702
46,432	0,223	0,321
59,250	0,216	0,325
10,081	0,102	0,1576
53,266	0,248	0,3389
36 256,714	0,154	0,311
1 161,524	0,189	0,2779
3 981,286	0,237	0,3437
<b>2006-1</b>		
0,011	0,337	0,001
0,003	0,350	0,0717
0,000	0,000	0
4,502	0,266	0,2914
7,148	0,118	0,1742
8,173	0,226	0,24
23,298	0,184	0,2723
99,012	0,179	0,2952
38,848	0,253	0,3024
204,048	0,316	0,2242
48,776	0,217	0,2931
47,036	0,216	0,2771
9,968	0,106	0,1268
51,373	0,255	0,2966
32 677,381	0,239	0,3096
919,941	0,207	0,2502
4 065,238	0,227	0,3063

Окончание табл. 5

$\bar{x}_j$	$\sigma_j$	$w_{jil}$
<b>2006-2</b>		
0,000	0,000	0
0,003	0,337	0,118
0,000	0,000	0
2,554	0,351	0,2921
5,506	0,391	0,2923
5,004	0,183	0,2552
14,679	0,147	0,2451
56,798	0,303	0,293
19,674	0,269	0,2854
118,607	0,407	0,2811
27,012	0,262	0,2859
23,262	0,279	0,2819
7,049	0,110	0,1352
26,051	0,270	0,2759
20 723,810	0,133	0,2242
594,631	0,275	0,2838
2 476,667	0,146	0,2397

Обработка стандартными статистическими методами дала следующие коэффициенты корреляции между результатами биотестов и интегральным показателем загрязнения (табл. 6).

Таблица 6. Коэффициенты корреляции

Биотест	2005	2006-1	2006-2
3	0,025 / 34	0,132 / 28	0,179 / 19
4	0,035 / 71	/ 2	-0,038 / 42
5	0,112 / 77	0,177 / 64	-0,022 / 43

Коэффициенты корреляции являются незначимыми. Через дробь в табл. 6 указано количество точек, для которых проводилось биотестирование. Следовательно, с помощью стандартных статистических методов установить влияние загрязнения на биоту не представляется возможным. Деревья решений позволяют получить информацию об этом влиянии.

#### Список литературы

1. *Оссовский, С.* Нейронные сети для обработки информации / С. Оссовский ; пер. с пол. И. Д. Рудинского. – М. : Финансы и статистика, 2002. – 344 с.
2. *Kiselev, M. V.* Regression-Based Classification Methods and Their Comparison with Decision Tree Algorithms / M. V. Kiselev, S. M. Ananyan, S. B. Arseniev // Proceedings of 1st European Symposium on Principles of Data Mining and Knowledge Discovery, Trondheim, Norway, Springer, 1997. – Pp. 134–144.
3. *Клир, Дж.* Системология. Автоматизация решения системных задач / Дж. Клир. – М. : Радио и связь, 1990. – 554 с.