

СПИСОК ЛИТЕРАТУРЫ

1. *Загоруйко Н. Г.* Методы распознавания образов: состояние и перспективы.- М.:Сов. Радио, 1972.- 120 с.
2. *Васильев В. М.* Распознающие системы.- М.: Киев:Наукова думка,- 180 с.
3. *Ковалевский В. А.* Методы распознавания изображений.- М.: 1975.-380 с.
4. *Батугин С. А., Ковалевская В. А.* Комплекс программ прогноза и управления многофакторными процессами.- Кемерово. КузГТУ, 1980.- 116 с.
5. *Андерсон Т.* Введение в многофакторный статистический анализ. - М.: Наука, 1968.- 500 с.
6. *Ковалевская В. А., Кубрак В. М., Андреев В. Е.* Прогноз успешности обучения курсантов высшей математике по данным вступительных экзаменов и профессионально-технического отбора. / Сборник научно-методических статей вузов МО РФ по математическим и общим естественно-научным дисциплинам.- // Новочеркасск.- 1998.- 85 с.
7. *Ковалевская В.А., Козлова И. Д., Шаламанов В. А.* Оценка перспективности угольных предприятий Кузбасска на примере шахт ОАО "Компании "Кузбассуголь" по графу прогноза и принятия решений" // Вестн. КузГТУ, 2001. №5. С. 96-100.

□ Авторы статьи:

Ковалевская
Валентина Алексеевна,
проф. каф. математики и инженер-
ной графики КФВУС

Кубрак
Валентина Михайловна
- зав. каф. математики и инженер-
ной графики КФВУС

УДК 519.6

В. А. Ковалевская, В. М. Кубрак, Л.Е.Мякишева

**НАУЧНО-МЕТОДИЧЕСКИЕ ОСНОВЫ ПРИМЕНЕНИЯ
РАСПОЗНАВАНИЯ ОБРАЗОВ ДЛЯ МНОГОФАКТОРНОГО ПРОГНОЗА
ПРИ РЕШЕНИИ ПРИКЛАДНЫХ ЗАДАЧ**

Методы распознавания образов позволяют эффективно решать задачи классификации, прогноза и управления многофакторными процессами и принятия в заданных условиях наиболее рационального решения в тех случаях, когда есть опыт прошлого (обучающая выборка). К таким задачам относятся: прогноз безопасности технологического процесса; состояния, надежности и долговечности приборов и систем; технико-экономических показателей ра-

боты предприятия; прогноз качества продукции; распознавания звуковых образов и изображений; задачи социологии, военного дела, теории связи и др.[1, 2, 6, 7, 10, 11].

С непрерывным ростом сложности изучаемых процессов, при исследовании которых необходимо принимать во внимание значительное число взаимосвязанных факторов, методы теории распознавания позволяют автоматизировать процесс обработки информации и

использовать эти методы в исследовательской работе [8].

Этапы оптимизации решения многофакторных задач прикладной направленности. Принимая оптимальное решение (лучшее в том или ином смысле, т. е. по некоторому критерию оптимальности [4, 9]) , полагают, что множество возможных решений каким - либо образом упорядочено, например, по надёжности результатов (критерию точности), минимуму суммарных затрат

Таблица 1

Расчёт безразмерных координат по фактору X_j

1	Интервалы фактора X_j	$l(omX_0 до X_1)$	$2(omX_1 до X_2)$	3.....	$r(omX_{r-1} до X_r)$	Σ
2	Число объектов класса А в интервале	n_{A1}	n_{A2}	n_{Ar}	N_A
3	Число объектов класса В в интервале	n_{B1}	n_{B2}	n_{Br}	N_B
4	Значение безразмерной координаты Z_{ij}	n_{A1} / n_1	n_{A2} / n_1	n_{Ar} / n_r	
5	Обозначения	Z_{1j}	Z_{2j}	Z_{rj}	

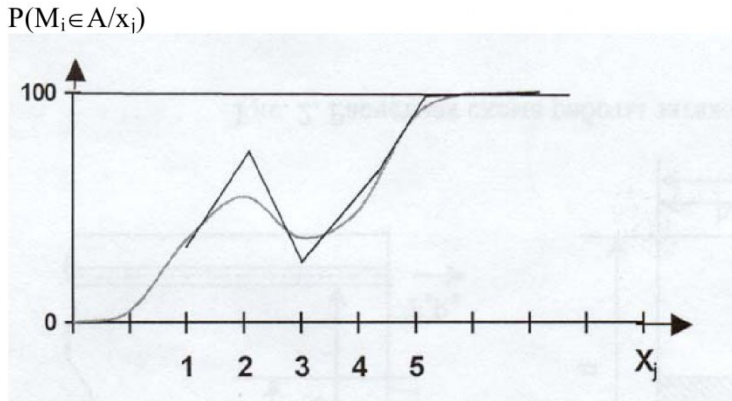


Рис. 1 График эмпирической (ломанная линия) и теоретической (кривая) вероятности принадлежности объектов обучающей выборки к классу A по фактору X_j к классу A по отдельно взятому фактору X_j .

прогноза и принятия решений (критерию экономической эффективности), максимуму безопасности (социальная эффективность) и др.

При решении прикладных задач, в том числе военнотехнических, методами распознавания образов выделяются следующие четыре основных этапа.

1. Обоснование критерия разделения на классы и формирование обучающей и контрольной выборок; введение безразмерных координат.

2. Выбор эффективной системы факторов для прогноза.

3. Разработка методов вероятностного прогноза и правила принятия оптимальных альтернативных решений о классе объекта.

4. Оценка надёжности прогноза и принятия решений, а также экономического и социального эффекта и возможности управления каждым объектом или группой объектов.

Частотный способ введения безразмерных координат. В многофакторной инженерной

задаче факторы могут иметь разную размерность, различный диапазон изменения; наряду с количественными могут быть и качественные факторы. Поэтому вводятся безразмерные величины (координаты), учитывающие опыт прошлого, зафиксированный в обучающей выборке распознаваемых классов. Эти безразмерные координаты являются вероятностно-статистическими количественными показателями принадлежности объектов к одному из выделенных классов и их расчёт основан на анализе вариационных рядов значений факторов в размерных координатах в общей обучающей выборке и в выборке, соответствующей прогнозируемому классу [7, 9].

При традиционном расчёте частотных безразмерных координат определяют экстремальные x_{min} и x_{max} значения каждого фактора и диапазон разбивают на L интервалов с длиной, определяемой по формуле Стёрджесса

$$\Delta x = \frac{x_{max} - x_{min}}{1 + 3,2 \lg N}, \quad (1)$$

где N – общее число объектов классов A и B (для качественных факторов число интервалов определяется числом градаций факторов).

В каждом интервале r определяется число объектов n_{Ar} класса A и n_{Br} класса B, а также $n_r = n_{Ar} + n_{Br}$ – общее число объектов. Новой безразмерной координатой является z_{ij} – условная частота принадлежности i -го объекта (M_i) по j -му фактору (X_j) к классу A:

$$z_{ij} = P(M_i \in A / X_j),$$

рассчитываемая по формуле:

$$z_{ij} = \frac{n_{Ar}}{n_r} \quad (2)$$

Если число объектов в интервале меньше заданной величины $H \approx 5$, то интервал объединяется с соседним.

Затем каждое значение данного фактора (j) на каждом объекте (i) x_{ij} обучающей выборки заменяется соответствующей безразмерной координатой z_{ij} – частотой принадлежности объекта с данным значением фактора к классу A.

По составленным таблицам по каждому фактору строятся графики зависимости принадлежности объекта M_i .

С помощью рассчитанных таблиц перехода от размерных координат объектов x_{ij} к безразмерным z_{ij} или по построенным графикам составляются обучающие матрицы объектов образов A и B в безразмерных координатах для всей обучающей выборки (объём обучающей выборки $N = N_A + N_B$)

Оценка степени информативности факторов и ранжирование факторов по степени информативности (под информативностью фактора понимается его влияние на принадлежность объекта к тому или иному классу). Степень информативности определяется по критериям Пирсона χ^2 , Шен-

$$A = \begin{pmatrix} \text{ОБРАЗ A} \\ z_{11} \dots z_{1n} \\ z_{21} \dots z_{2n} \\ \dots \\ z_{N_{a1}} \dots z_{N_{an}} \end{pmatrix} \quad B = \begin{pmatrix} \text{ОБРАЗ B} \\ z_{N_{a+1,1}} \dots z_{N_{a+1,n}} \\ z_{N_{a+2,1}} \dots z_{N_{a+2,n}} \\ \dots \\ z_{N_{,1}} \dots z_{N_{,n}} \end{pmatrix}$$

нона, Колмогорова–Смирнова, Кульбака, Стьюдента, Фишера, Уилкоксона [3, 7].

После определения информативности факторов по всем критериям производится ранжирование факторов в порядке уменьшения значений оценок их информативности по каждому эксперту (критерию). Вычисление коэффициентов компетентности экспертов первого приближения осуществляется по формулам [9]:

$$K_j = \frac{1}{\lambda} \sum_{i=1}^n x_{ij} \cdot x_i, \quad (3)$$

где

$$\lambda = \sum_{i=1}^n \sum_{j=1}^m x_{ij} \cdot x_i,$$

$$x_i = \sum_{j=1}^m x_{ij} \cdot K_j^0, \quad K_j^0 = 1/m$$

– коэффициент компетентности j -го эксперта, n – количество факторов, m – число экспертов, j – номер эксперта, i – номер фактора, K_j^0 – начальное значение коэффициентов компетентности экспертов, x_{ij} – число, представляющее собой оценку i -го фактора j -м экспертом.

Вычисление сумм централизованных рангов для i -го фактора производится по формуле

$$r_i = \sum_{j=1}^m K_j \cdot r_{ij}, \quad (4)$$

где r_{ij} – ранг, присвоенный j -м экспертом i -му фактору.

Далее факторы упорядочиваются по цепочке неравенств

$$r_1 < r_2 < \dots < r_i < \dots < r_n, \quad i=1, 2, \dots, n.$$

Количественной мерой согласованности мнений экспертов является дисперсионный коэффициент конкордации

$$W = \frac{12S}{m^2(n^3 - n) - m \sum_{j=1}^m T_j} \quad (5)$$

где

$$T_j = \sum_{k=1}^{H_j} (h_k^3 - h_k)$$

– показатель связанных рангов в

j -й ранжировке, H_j – число групп равных рангов в j -й ранжировке, h_k – число равных (одинаковых) рангов при ранжировке j -м экспертом:

$$S = \sum_{i=1}^n \left(\sum_{j=1}^m r_{ij} - \bar{r} \right),$$

где

$$\bar{r} = \frac{1}{n} \sum_{i=1}^n \sum_{j=1}^m r_{ij}.$$

Коэффициент конкордации W изменяется в пределах $0 \leq W \leq 1$: $W=1$, если все ранжировки экспертов одинаковы, и $W=0$, если они различны.

Оценка значимости коэффициента конкордации W при $n \geq 7$ производится по критерию Пирсона. В случае отсутствия одинаковых рангов

$$\chi_{расч}^2 = Wm(n-1), \quad (6)$$

при наличии одинаковых рангов

$$\chi_{расч}^2 = \frac{12S}{mn(n+1) - \frac{1}{n-1} \sum_{j=1}^m T_j} \quad (7)$$

Число степеней свободы равно $\nu = n - 1$.

Практически реализация оценки информативности выполняется в два этапа.

I. Для каждого фактора X_j по соответствующим экспериментальным данным оценивается информативность факторов по критерию Пирсона. Здесь

а) рассчитывается фактическое значение критерия:

$$\chi_{расч}^2 = \frac{\sum_{k=1}^r (n_{kA} - n_{kA}^T)^2}{n_{kA}^T} \quad (8)$$

где $n_{kA}^T = \frac{N_A}{N} \cdot n_k$ – теоретическая частота объектов класса A в каждом интервале, вычисленная в предположении о равномерном распределении объектов класса A по интервалам, т. е. в предположении, что

значения данного фактора не влияют на разделение классов;

б) определяется табличное значение χ^2 - критерия, т. е. его предельное допустимое значение [3], при котором фактор X_j можно считать не влияющим на процесс распознавания A и B ;

в) определяется диагностическая величина:

$$\Delta \chi^2(X_j) = \chi_{расч}^2(X_j) - \chi_{табл}^2(X_j) \quad (9)$$

г) делается вывод об информативности или неинформативности фактора для процесса распознавания: если

$$\chi_{расч}^2(X_j) > \chi_{табл}^2(X_j)$$

(т. е. $\Delta \chi^2(X_j) > 0$), то гипотеза о равномерном распределении объектов класса A по интервалам значений фактора X_j отвергается. Следовательно, это распределение неравномерно и фактор X_j влияет на распределение классов, т. е. является информативным.

II. Результаты расчётов информативности сводятся в таблицу и делается вывод не только об информативности каждого фактора, но и об относительной информативности, основанной на величине превышения $\Delta \chi_j^2$ (табл. 2).

Определение степени взаимосвязи факторов. Устанавливается корреляционная связь между факторами в исходной обучающей выборке и из каждой пары зависимых факторов оставляют какой-либо один (решается на профессиональном уровне), так как в нем заключена информация об обоих факторах. Поскольку факторы могут быть измерены в различных шкалах (отношений, интервалов, наименований, порядка и дихотомии), то в зависимости от сочетания различных шкал измерения двух факторов вычисляются различные меры связи:

Таблица 2

Ранги факторов, соответствующие степени информативности

Фактор	Число интервалов r	Число степеней свободы $f = r - 2$	$\chi^2_{расч.}$	$\chi^2_{табл.} (0,05)$	$\Delta\chi^2$	Ранг R_j
Z_1	3	1	13,2	3,8	+9,4	I
Z_2	4	2	6,6	6,0	+0,6	III
Z_3	5	3	10,8	7,8	+3,0	II
Z_4	6	4	9,6	9,5	+0,1	IV
Z_5	7	5	9,8	11,1	(-1,3)	V

а) – выборочный коэффициент линейной корреляции, б) выборочный коэффициент ранговой корреляции Спирмена; в) точечный бисериальный выборочный коэффициент корреляции Пирсона; г) выборочный коэффициент корреляции Крамера; д) φ – выборочный коэффициент сопряженности Пирсона [3, 4, 9].

При этом в качестве основного показателя степени взаимосвязей факторов берётся выборочный коэффициент линейной корреляции r_{ij} , который для факторов, заданных в шкале отношений или интервалов, вычисляется по формуле [3]

$$r_B = \frac{S_{jk}}{S_{jj} \cdot S_{kk}}, \quad (10)$$

где

$$S_{jk} = \sum_{i=1}^n (x_{ij} - T_j)(x_{ik} - T_k) -$$

$$- \frac{1}{n} \sum_{i=1}^n (x_{ij} - T_j) \cdot \sum_{i=1}^n (x_{ik} - T_k)$$

$$T_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ij},$$

$$T_k = \frac{1}{n_k} \sum_{j=1}^{n_k} x_{ij},$$

S_{jj}^2 и S_{kk}^2 – выборочные

дисперсии.

Построение матрицы и графа взаимосвязей. Ввод безразмерных частотных координат вместо размерных координат различных шкал измерения с методической точки зрения позволяет в качестве количественной характеристики си-

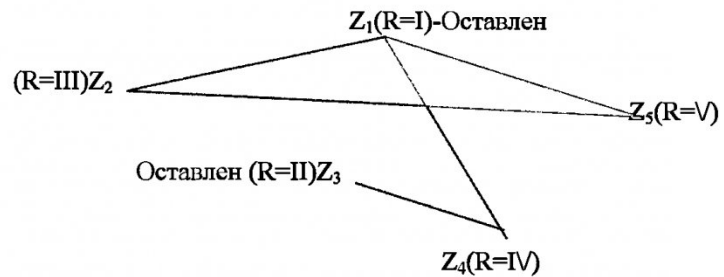


Рис. 2 Граф взаимосвязей факторов с учетом их рангов по информативности

лы (степени) взаимосвязи факторов использовать широко распространённый в прикладных исследованиях выборочный коэффициент линейной корреляции, разработанный для количественных факторов, т. к. факторы, измеренные в безразмерных координатах, непрерывно изменяются на интервале $[0; 1]$ ($z \in [0; 1]$).

Здесь вычисляются парные коэффициенты и строится корреляционная матрица и граф взаимосвязей.

Можно рекомендовать построение упрощённой корреляционной матрицы, где в клетке на пересечении j -ой строки и l -го столбца ставится знак +, если X_j и X_l зависимы, то есть если между ними существ-

вует тесная связь (табл.3). Для наглядности проведённых расчётов целесообразно построить граф взаимосвязи в системе факторов. На графе взаимосвязей исследуемой системы факторов (рис.2) соединяют линией (ребром) вершины, если между ними есть тесная связь.

Минимизация признакового пространства. После обнаружения факта наличия связи между признаками из каждой пары зависимых факторов оставить один (более информативный), так как в нём заключена информация не только о нём, но и о другом факторе. Если, например, между факторами, имеющими ранги, приведенные в последнем столбце табл.2, установлены значимые взаимо-

Таблица 3

Матрица взаимосвязей

Признаки	Z_1	Z_2	Z_3	Z_4	Z_5
Z_1	1	+	-	+	+
Z_2		1	-	-	+
Z_3			1	+	-
Z_4				1	-
Z_5					1

связи, зафиксированные в матрице взаимосвязей (табл. 3), то в систему наиболее информативных и независимых факторов будут входить только факторы Z_j и Z_3 (рис.2).

Формирование окончательной обучающей и контрольной выборки. Окончательно обучающие и контрольные выборки классов А и В имеют вид:

- а) факторы описаны безразмерными координатами,
- б) по минимальной системе факторов,
- в) порядок следования факторов соответствует их рангам: вначале тот фактор Z_j , который имеет ранг I, затем фактор с рангом II и т.д.

Следует заметить, что для большей точности решения задачи распознавания факторы необходимо исследовать на информативность и независимость по нескольким критериям.

Третий этап оптимизации решения прикладной задачи включает в себя разработку алгоритмов вероятностного прогноза выделенного класса объектов по минимальной и эффективной системе факторов и решающего правила альтернативной классификации. При этом для выбора алгоритмов распознавания, приемлемых для решения задачи, или разработки новых алгоритмов необходимо учитывать характеристики взаимного расположения объектов распознаваемых классов в n - мерном признаковом пространстве информативных и независимых факторов.

Характеристики положения образов. На основе значений факторов в окончательной обучающей выборке, переведенных в безразмерные координаты, вычисляются следующие характеристики образа А и образа В (рис.3):

- а) координаты центров тяжести образов O_A и O_B :

$$\begin{aligned} O_A &= \{ \bar{Y}_{A1}, \bar{Y}_{A2}, \dots, \bar{Y}_{Aj}, \dots, \bar{Y}_{AM} \} \\ O_B &= \{ \bar{Y}_{B1}, \bar{Y}_{B2}, \dots, \bar{Y}_{Bj}, \dots, \bar{Y}_{BM} \} \end{aligned} \quad (11)$$

где

$$\bar{Y}_{Aj} = \frac{\sum_{i=1}^{N_A} Y_{Aij}}{N_A},$$

$$\bar{Y}_{Bj} = \frac{\sum_{i=1}^{N_B} Y_{Bij}}{N_B}$$

- безразмерные координаты i -го объекта по j -му фактору классов А и В соответственно.

- б) расстояние от начала координат до центра тяжести образа А и образа В

$$\rho_{OA} = \sqrt{\sum_{j=1}^m \bar{Y}_{Aj}^2},$$

$$\rho_{OB} = \sqrt{\sum_{j=1}^m \bar{Y}_{Bj}^2} \quad (12)$$

- в) расстояние между центрами тяжести образов

$$\rho_{AB} = \sqrt{\sum_{j=1}^m (\bar{Y}_{Aj} - \bar{Y}_{Bj})^2} \quad (13)$$

- г) угол между радиус-векторами центров тяжести образов

$$\cos \varphi = \frac{\sum_{j=1}^m \bar{Y}_{Aj} \cdot \bar{Y}_{Bj}}{\rho_{OA} \cdot \rho_{OB}} \quad (14)$$

С учетом полученных характеристик положения образов в n - мерном пространстве про-

изводятся выбор одного из алгоритмов распознавания для вычисления обобщенных координат объектов по комплексу факторов. При этом все известные алгоритмы распознавания [1, 2, 5] модифицированы нами для вероятностного прогноза.

Алгоритм 1. Распознавание по разности расстояний от объектов до центров тяжести образов (классов).

Объект принадлежит к тому классу, к центру тяжести которого находится ближе.

Для всех объектов классов А и В вычисляются расстояния ρ_{Ai} и ρ_{Bi} от объектов $M_i(Y_{i1}, Y_{i2}, \dots, Y_{ij}, \dots, Y_{im})$ до центров тяжести классов O_A и O_B , $i=1, 2, \dots, N$ (рис.3)

$$\rho_{Ai} = |M_i O_A| = \sqrt{\sum_{j=1}^m (\bar{Y}_{Aj} - Y_{ij})^2}$$

$$\rho_{Bi} = |M_i O_B| = \sqrt{\sum_{j=1}^m (\bar{Y}_{Bj} - Y_{ij})^2} \quad (15)$$

где m – число факторов.

Определяются обобщенные координаты

$$\rho_i = \rho_{Ai} - \rho_{Bi} \quad (16)$$

Алгоритм 2. Распознавание по разности углов между радиус-векторами объектов и центров тяжести образов.

Объект принадлежит к тому классу, где угол между радиус-векторами объекта и центров тяжести образов меньше.

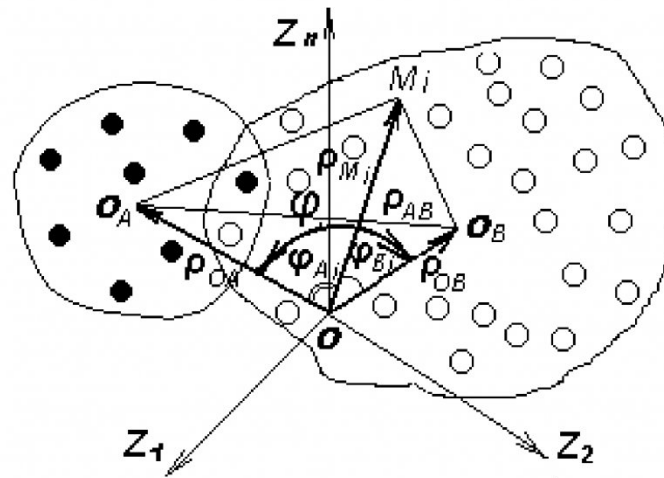


Рис.3 Геометрическая интерпретация образов А и В и их характеристик

Для объектов M_i классов А и В вычисляются косинусы углов между радиус-векторами объектов OM_i и соответственно радиус-векторами центров тяжести образа А – \overrightarrow{OA} , образа

В – \overrightarrow{OB} (рис.3)

$$\cos \varphi_{A_i} = \frac{\overrightarrow{OA} \cdot \overrightarrow{OM}_i}{|\overrightarrow{OA}| \cdot |\overrightarrow{OM}_i|} = \frac{\sum_{j=1}^M \bar{Z}_{Aj} \cdot Z_{ij}}{\sqrt{\sum_{j=1}^M \bar{Z}_{Aj}^2} \cdot \sqrt{\sum_{j=1}^M Z_{ij}^2}} \quad (17)$$

$$\cos \varphi_{B_i} = \frac{\overrightarrow{OB} \cdot \overrightarrow{OM}_i}{|\overrightarrow{OB}| \cdot |\overrightarrow{OM}_i|} = \frac{\sum_{j=1}^M \bar{Z}_{Bj} \cdot Z_{ij}}{\sqrt{\sum_{j=1}^M \bar{Z}_{Bj}^2} \cdot \sqrt{\sum_{j=1}^M Z_{ij}^2}} \quad (18)$$

По значениям $\cos \varphi_{A_i} = C_{A_i}$, $\cos \varphi_{B_i} = C_{B_i}$ определяются величины углов и обобщенные координаты объектов

$$\rho_i = \varphi_{A_i} - \varphi_{B_i} \quad (19)$$

Алгоритм 3. Распознавание по расстоянию от объектов до гиперплоскости, проходящей через середину отрезка, соединяющего центры тяжести образов, нормально к нему.

Объекты, расположенные по разные стороны от такой гиперплоскости, принадлежат разным классам.

Гиперплоскость проходит через середину отрезка $O_A O_B$, т.е. через точку

$$D = \left\{ \begin{array}{l} \frac{1}{2}(\bar{Z}_{A1} + \bar{Z}_{B1}), \\ \frac{1}{2}(\bar{Z}_{A2} + \bar{Z}_{B2}), \dots, \\ \frac{1}{2}(\bar{Z}_{AM} + \bar{Z}_{BM}) \end{array} \right\}$$

и перпендикулярно вектору

$$\overrightarrow{O_A O_B} = \left\{ \begin{array}{l} (\bar{Z}_{B1} - \bar{Z}_{A1}), \\ (\bar{Z}_{B2} - \bar{Z}_{A2}), \dots, \\ (\bar{Z}_{BM} - \bar{Z}_{AM}) \end{array} \right\}$$

Для каждого объекта определяется обобщенная координата

$$\rho_i = \sum_{j=1}^m (\bar{Z}_{Bj} - \bar{Z}_{Aj}) \left(Z_{ij} - \frac{\bar{Z}_{Aj} + \bar{Z}_{Bj}}{2} \right) \quad (21)$$

Алгоритм 4. Распознавание по расстоянию от начала координат до каждого объекта.

Разделяющей образы поверхностью является гиперсфера с центром в начале координат и радиусом F_0 , определяемым порогом, установленным на профессиональном уровне.

Для всех объектов M_i ($Z_{i1}, Z_{i2}, Z_{i3}, \dots, Z_{im}$) вычисляются их расстояния ρ_{Mi} от начала координат, которые и принимаются за обобщенные

$$\rho_i = \rho_{Mi} = \sqrt{\sum_{j=1}^m Z_{ij}^2} \quad (22)$$

Алгоритм 5. Распознавание по условной вероятности совместного влияния всех факторов.

Значения Z_{ij} (2) являются условными вероятностями принадлежности i -го объекта по j -му фактору к классу А.

Для каждого объекта вычисляются вероятности принадлежности объекта к классу А по всем факторам, предполагая, что факторы независимые.

Величина обобщенной координаты равна

$$\rho_i = m \sqrt{\prod_{j=1}^m Z_{ij}} \quad (23)$$

Построение вероятностных прогнозных кривых на основе рассчитанных обобщенных координат объектов по комплексу факторов.

После вычисления обобщенной координаты по каждому алгоритму составляется вариационный ряд этой координаты и строится график зависимости вероятности принадлежности к классу А для каждого интервала значений обобщенной координаты, т.е. вероятностная кривая прогноза. Для этого

1. рассчитаем шаг по обоб-

щённой координате $f(z_i) = \rho_i$

$$\Delta \rho_i = \frac{\rho \max - \rho \min}{1 + 3,32 \ln N}$$

и построим интервальный вариационный ряд для ρ ;

2. рассчитаем координаты эмпирической прогнозной кривой по обобщенной координате ρ

$$\rho_k = \frac{m_k(A)}{m_k}$$

(ρ_k - эмпирические условные вероятности - относительные частоты объектов класса А в k -ом интервале вариационного ряда для ρ);

3. определим параметры теоретической (нормальной) кривой

$$a = \frac{\rho \max + \rho \min}{2},$$

$$\sigma = \frac{1}{6}(\rho \max - \rho \min); \quad (24)$$

4. сгладим эмпирическую кривую теоретической

$$P_{теор.} = \frac{1}{2} + \Phi\left(\frac{\rho_k - a}{\sigma}\right), \quad (25)$$

$\Phi(t)$ - интегральная функция Лапласа ([3], табл. 2),

$$t_k = \frac{\rho_k - a}{\sigma} \quad - \text{условный}$$

параметр k -го интервала, т.е.

$$P_k^{теор.} = \frac{1}{2} + \Phi(t_k);$$

5. по интервалам $\Delta \rho$ и частотам P_k построим гистограмму зависимости условной частоты принадлежности объекта к классу А от обобщенной координаты $F_i = f(z_i)$, которая сглаживается теоретической вероятностной кривой (рис.4).

Построение решающего правила классификации. После оценки для каждого объекта вероятности отнесения его к классу А - количественный прогноз, следует осуществить и итоговый альтернативный прогноз, т.е. принять решение - к какому из двух классов относится объект M_i , если для него рассчитана $P(M_i \in A)$.

Для этого используется

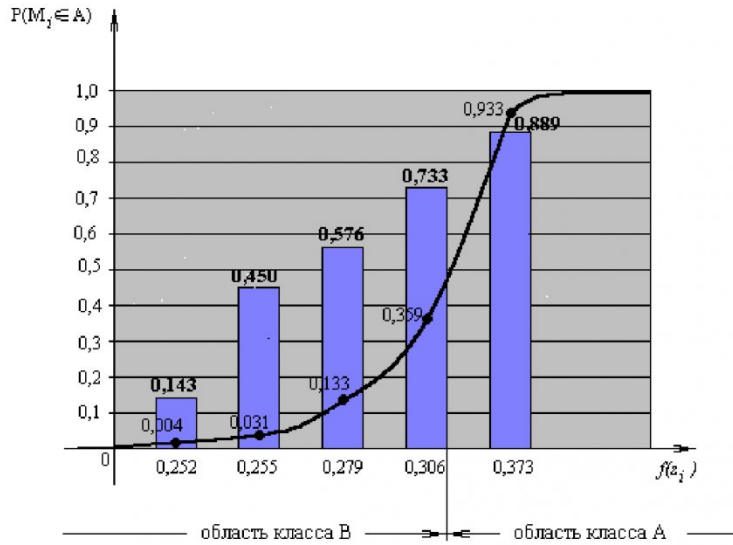


Рис.4 График теоретической вероятностной кривой прогноза класса объектов

простейшее решающее правило классификации:

Решение:

$$\begin{cases} \text{класс A, } P(M_i \in A) > P_0 \\ \text{класс B, } P(M_i \in A) \leq P_0 \end{cases}$$

$$(P_0 = 0,5) \quad (26)$$

В случае известных стоимостей ошибок I и II рода c_1 и c_2 априорных вероятностей классов q_1, q_2 - решающее правило уточняется с учетом этой информации. При этом достаточно знать только соотношение стоимостей ошибок (во сколько раз одна из них больше другой), а не реальные (экономические) показатели, т. е.

$$P_0 = \frac{1}{1 + \frac{q_1 c_1}{q_2 c_2}} \quad (27)$$

Дополнительный учёт взаимосвязей факторов для повышения надёжности прогноза и осуществления управления многофакторным процессом. Он возможен при учёте сочетаний значений факторов и реализуется в логических алгоритмах теории распознавания и многофакторного статистического анализа

Решающей функцией в логических алгоритмах является конъюнкция: сочетание значений факторов или интервалов значений факторов. Так, например, в медицине широко из-

вестны под понятием “синдром”, сочетания двух, трех и более факторов при диагностике какого-либо заболевания или при дифференциации одного заболевания от другого. Аналогичные сочетания можно считать в распознаваемых классах объектов любой природы. При этом ищутся и используются только такие сочетания, которые встречаются максимальное число раз в своем классе и минимальное число раз - в “чужом” (пример диагностического сочетания в медицинской диагностике - “возраст” X_1 до 50 лет при нижней величине артериального давления X_2 менее 80).

Можно перейти к булевым переменным, задав порог равным, например, 45 лет, т. е. (если $X_1 \leq 45$, то $X_1=0$, а если $X_1 > 45$, то $X_1=1$), а по X_2 -порог 60 (т.е. при $X_2 < 60$ $X_2=0$, а по $X_2 > 60$ $X_2=1$.) Тогда это же сочетание будет описано как “ $X_1=1 \wedge X_2=0$ ”.

При решении практических задач такой перебор возможных сочетаний значений факторов под силу только ЭВМ с высокой степенью быстродействия, поэтому для решения многих задач можно использовать более простые, наглядные и не менее эффективные логические алгоритмы распознавания, реа-

лизуемые в виде графа - “дерева” прогноза и принятия решений.

Алгоритм 6. Граф прогноза и принятия решений (логический алгоритм)

Для построения графа прогноза используются положения системного подхода к построению математических моделей: последовательность, иерархичность, наглядность и оперативность. Принципиальная схема разработки графа прогноза и принятия решений подробно описаны в комплексе программ прогноза и управления многофакторными процессами (КузГТУ) [7] и в разработанном кафедрой математики КФВУС учебном пособии по применению статистических методов теории распознавания образов к решению задач прикладной направленности [11].

На практике использование принципов системного подхода означает использование ранее полученных результатов.

1. Ранжирование факторов по степени информативности каждый из факторов прогноза определяет один из иерархических уровней в соответствии с его информационным рангом.

2. Рассчитанные при вычислении безразмерных координат граничные значения между областями A и B по каждому фактору определяют ветви графа по каждому уровню.

3. Сочетания возможных значений всех факторов определяют обобщённые ветви графа - реально встречающиеся сочетания значений факторов, характерные для решаемой задачи. При этом значения факторов могут как усиливать, так и ослаблять их совокупное влияние на итоговый показатель принадлежности объектов к выделенному классу.

4. Итоговые вероятности принадлежности объекта к классу A в данном методе рассчитываются как отношение числа n_A объектов класса A к общему числу n объектов по выделенным ветвям:

$$P(X_i \in A) = n_A / n.$$

Априори	Оценка по математике	Оценка по физике	Оценка по физ. подг-ке	Конкурсный балл	№ обобщённой ветви	Решение					
P_0	$P(X_4)$	$P(X_4; X_3)$	$P(X_4; X_3; X_2)$	$P(X_4; X_3; X_2; X_6)$		$P(X_i \in A)$	Кл.				
$P_0 = 0,62$ (40/65)	$Y \leq 3$	0,43 (18/2)	≤ 3	0,36 (10/28)	≤ 4	0,29 (7/24)	1	0,29	В		
			$\Rightarrow 4$	0,57 (8/14)	≤ 4	0,58 (7/12)	3	0,58	А		
	$Y \Rightarrow 4$	0,96 (22/23)	≤ 3	0,91 (10/11)	≤ 4	0,87 (7/8)	5	0,87	А		
			$\Rightarrow 4$	1,0 (12/12)		1,0 (12/12)	8	1,0	А		
							$\leq 18,3$	0,83 (5/6)	6	0,83	А
							$> 18,3$	1,0 (2/2)	7	1,0	А
								1,0 (12/12)			А

Рис. 5. Граф прогноза класса успешности обучения высшей математике по комплексу факторов.

5. В заключительном столбце графа приведены результаты альтернативного решения о классе объекта, рассчитанный на основе решающего правила классификации.

В качестве примера приведём граф-метод прогноза и принятия решений о классе успешности обучения курсантов в третьем семестре по данным профессионально педагогического отбора и вступительных экзаменов (рис.5)

Первый уровень прогноза - на основе значений главного по информативности фактора X_1 . Граничным значением $X_1^0 = 4$ область его значений разбивается, например, на две ветви. Для каждой из них определяется отношение числа объектов класса А к общему числу объектов:

$$\frac{n_{kA}}{n_{kA} + n_{kB}} = P_k \quad (k=1,2)$$

- вероятности принадлежности к классу А : P_1, P_2 (только по первому фактору)

Второй уровень - учитыва-

ется совместное влияние уже двух факторов. Получаем: $P_{11}, P_{12}, P_{21}, P_{23}$ - четыре ветви второго уровня графа.

Дальнейшая детализация по всем другим информативным факторам позволяет выделить обобщённые (итоговые) ветви графа с соответствующими им вероятностями принадлежности к классу А, т. е.

$$P(M_i \in A) = P(X_{i1}, X_{i2}, X_{i3}, X_{i4})$$

Прогноз нового объекта и осуществляется следующим образом.

1) По значениям фактора прослеживается его путь на графе- определённая его обобщённая ветвь.

2) Вероятность принадлежности M_0 к классу А, т. е. $P(M_0 \in A)$ равна итоговой вероятности обобщённой ветви.

3) Альтернативное решение о классе успешности нового объекта принимается на основе единого вероятностно-альтернативного решающего

правила классификации (26-27), учитывающего динамику изменения стоимости ошибок I и II рода и априорные вероятности распознаваемых классов.

В схеме граф-метода этап принятия решений проводится на основе заключительного столбца (рис. 5)

Оценка надежности методов прогноза по контрольной выборке объектов - это заключительный этап оптимизации решения задачи. На нём проверяется адекватность разработанных на основе обучающих выборок моделей многофакторного прогноза. Для контрольной выборки берутся новые объекты, не вошедшие в исходную обучающую выборку.

Данные контрольной выборки переводятся в безразмерные координаты по таблицам перевода, составленным по окончательной обучающей выборке. Вычисляются условные вероятности $W_i = P(X_i \in A)$ - принадлежности объекта к классу А. Значения W_i сравниваются с порогом принятия ре-

шения p_0 , установленным из соображений специфики решаемой задачи на профессиональном уровне. Если $W_i < p_0$, то объект относится к классу В, если $W_i \geq p_0$, то к классу А (В большинстве случаев $p_0 = 0,5$).

Затем производится сопоставление прогнозируемых значений с фактическими. Определяется количество ошибок прогноза по каждому классу. При этом различают ошибки I и II рода: ошибка I рода – если объект класса А прогнозируется как объект класса В, ошибка II рода – если объект класса В прогнозируется как объект класса А.

По каждому j -му алгоритму вычисляются вероятности ошибки прогноза I рода

$$p_1 = \frac{n_1}{N_1}, \text{ II рода } p_2 = \frac{n_2}{N_2};$$

вероятность суммарной ошибки

$$P_{\text{ош.}} = \frac{n_1 + n_2}{N_1 + N_2} \quad (28)$$

и надёжность прогноза

$$H = (1 - P_{\text{ош.}})100\% \quad (29)$$

где n_1 и n_2 – число ошибок I и II рода, соответственно, при прогнозе по данному алгоритму, а N_1, N_2 – общее число объектов в контрольных выборках классов А и В [2].

Выбор оптимального метода прогноза и принятия решений для решаемой прикладной задачи проводится на основе алгоритма, обеспечивающего не только наибольшую надёжность прогноза, но и минимум затрат на принятие решений. На практике имеет значение оперативность и наглядность проведения прогноза и принятия решений о классе новых объектов, а также возможность управления ими, т. е. такому изменению параметров процес-

са, которое при заданных начальных условиях и имеющихся экономических, технических и социальных условиях позволит перевести прогнозируемый объект в "желаемый" класс.

В наших исследованиях сравнительная оценка надёжности, экономической эффективности, а также социального эффекта при решении многофакторных задач, касающихся учебно-воспитательного процесса и научно-исследовательской работы студентов и учащихся школ, показала, что наиболее надёжным из рассмотренных в данной статье алгоритмов распознавания образов является граф-метод прогноза и принятия альтернативных решений, допускающий к тому же возможность управления прогнозируемым процессом.

СПИСОК ЛИТЕРАТУРЫ

1. *Загоруйко Н. Г.* Методы распознавания образов: состояние и перспективы. – М.: Сов. радио, 1972. – 120 с.
2. *Васильев В. М.* Распознающие системы. – Киев: Наукова думка, - 180 с.
3. *Гмурман В. Е.* Теория вероятностей и математическая статистика. -М.: Высш. шк., 1972.- 386 с.
4. *Зельдович Я. Б., Мышкис Я. С.* Элементы прикладной математики. – М.: Наука, 1967. – 646 с.
5. *Ковалевский В. А.* Методы распознавания изображений.- М.: 1975-380с.
6. *Васильев С. К., Захаров В. Н., Прохоров Ю. Ф.* Кибернетика в системах военного назначения. – М.: Воениздат. 1980
7. Комплекс программ прогноза и управления многофакторными процессами. – Кемерово. КузГТУ. 1980. – 116 с.
8. *Ковалевская В. А., Славолубова Н. Б., Севидова И. И.* Методические указания по выполнению ВНР курсантами КФВУС.- Кемерово. – 2001 – 20 с.
9. *Андерсон Т.* Введение в многофакторный статистический анализ. – М.: Наука, 1968. – 500 с.
10. *Батугин С. А.* Вопросы классификации распознавания / Механика горных пород.- Кемерово, 1972. с.- 40-44.
11. *Ковалевская В. А., Кубрак В. М.* Применение статистических методов теории распознавания образов к решению задач прикладной направленности. Учебное пособие.- Кемерово, 2003.- 57 с.

□ Авторы статьи:

Ковалевская
Валентина Алексеевна
-проф. каф. математики
инженерной графики КВУС

Кубрак
Валентина Михайловна
- зав. каф. математики
инженерной графики КФВУС

Мякишева
Любовь Евтафьевна
- доц. каф. прикладной математики