

УДК: 622.142.5:004. 94

ПРИМЕНЕНИЕ МЕТОДА РАСПОЗНОВАНИЯ ОБРАЗОВ ДЛЯ КЛАССИФИКАЦИИ ОБЪЕКТОВ НА ЦИФРОВЫХ ПЛАНАХ

Игнатов Ю.М., к. т. н., доцент, Мякишева Л.Е., к. т. н., доцент
Латагуз М. М., ст. преподаватель,
Кузбасский государственный технический университет имени Т.Ф.
Горбачева, г. Кемерово
Гагарин А.А., ОАО «СУЭК-Кузбасс»

На цифровых маркшейдерских планах нами выполняется размещение полей геометрических и геомеханических характеристик по материалам разведочных скважин и горных работ и строятся цифровые информационно-геомеханические модели горного массива. В разработанном методе за упорядоченную форму хранения и представления информации принят регулярно-ячеистый принцип организации данных. Для описания данных за минимальную единицу площади принимается ячейка правильной геометрической формы, а пространственно-распределенные признаки объектов хранятся в узлах регулярной сетки из ячеек. Созданные матрицы импортируются в геоинформационные системы (ГИС) в виде отдельных слоев с названиями «Матрица-1», «Матрица-2» и др.

Одним из инструментов ГИС является возможность выполнить прогноз по выбранному участку угольного пласта с помощью структурированного языка запросов SQL. Для решения конкретной задачи и создания запроса необходим метод классификации объектов.

Решение задач поиска опасных зон по выбросоопасности и по устойчивости кровли нами выполняется методами распознавания образов. При решении выделяются следующие четыре основных этапа:

1. Обоснование критерия разделения на классы и формирование обучающей и контрольной выборок в безразмерных координатах;
2. Выбор эффективной системы факторов для прогноза;
3. Разработка методов вероятностного прогноза и правила принятия оптимальных альтернативных решений о классе объекта;
4. Оценка надежности прогноза и принятия решений.

В многофакторной инженерной задаче факторы могут иметь разную размерность, различный диапазон изменения; и могут быть и качественные факторы. Поэтому вводятся безразмерные величины (координаты), учитывающие опыт прошлого, зафиксированный в обучающей выборке

распознаваемых классов. Эти безразмерные координаты являются вероятностными количественными показателями принадлежности объектов к одному из выделенных классов.

При расчете частотных безразмерных координат определяют экстремальные x_{\min} и x_{\max} значения каждого фактора и диапазона разбивают на L интервалов с длиной, определяемой по формуле Стэрджесса:

$$\Delta x = \frac{x_{\max} - x_{\min}}{1 + 3,2 \ln N},$$

где N - общее число объектов 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 .

$$Z_{ij} = P_j(M_i \in A)$$

Рассчитываем по формуле:

$$Z_{ij} = \frac{n_{Ar}}{n_r}$$

Затем каждое значение данного фактора (j) на каждом объекте (i) x_{ij} обучающей выборки заменяется соответствующей безразмерной координатой Z_{ij} – частотой принадлежности объекта с данными значениями фактора к классу A . По составленным таблицам по каждому фактору строятся графики зависимости принадлежности объекта M_i

С помощью расчетных таблиц перехода от размерных координат объектов X_{ij} к безразмерным Z_{ij} или по построенным графикам составляются обучающие матрицы объектов классов A и B в безразмерных координатах для всей обучающей выборки (объем обучающей выборки $N = N_A + N_B$).

Далее производится оценка степени информативности факторов и ранжирование факторов по степени информативности. Степень информативности определяется по 7 критериям [3, 7, 9].

После определения информативности факторов по всем критериям производится ранжирование факторов в порядке уменьшения значений оценок их информативности по каждому эксперту (критерию). Количественной мерой согласованности мнений экспертов является дисперсионный коэффициент конкордации. 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.

Затем производится 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 [3,9].

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

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 В:

- а) координаты центров тяжести образов O_A и O_B :
- б) расстояние от начала координат до центра тяжести образов А и В
- в) расстояние между центрами тяжести образов
- г) угол между радиусами векторами центров тяжести образов

С учетом полученных характеристик положения образов в n -мерном пространстве производится 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 вычисляют 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 , $i=1, 2, \dots, N$

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

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

где m – число факторов.

Определяются обобщенные координаты

$$\rho_i = \rho_{Ai} - \rho_{Bi}$$

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.

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-ОА, образа В-ОВ.

$$\cos \varphi_{Ai} = \frac{\overrightarrow{OA} \cdot \overrightarrow{OM_i}}{|\overrightarrow{OA}| \cdot |\overrightarrow{OM_i}|} = \frac{\sum_{j=1}^M Z_{Ai} \cdot Z_{ij}}{\sqrt{\sum_{j=1}^M Z_{Aj}^2} \cdot \sqrt{Z_{ij}^2}}$$

$$\cos \varphi_{Bi} = \frac{\overrightarrow{OB} \cdot \overrightarrow{OM_i}}{|\overrightarrow{OB}| \cdot |\overrightarrow{OM_i}|} = \frac{\sum_{j=1}^M Z_{Bi} \cdot Z_{ij}}{\sqrt{\sum_{j=1}^M Z_{Bj}^2} \cdot \sqrt{Z_{ij}^2}}$$

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

$$\rho_i = \varphi_{Ai} - \varphi_{Bi}$$

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

Объекты, расположенные по разным сторонам от такой гиперплоскости, принадлежат разным классам.

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

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

И перпендикулярно вектору

$$\overrightarrow{O_A O_B} = \left\{ \begin{matrix} Z_{B1} - Z_{A1} \\ Z_{B2} - Z_{A2} \dots \\ Z_{BM} - Z_{AM} \end{matrix} \right\}$$

Для каждого объекта определяется обобщенная координата

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

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

Разделяющей образцы поверхностью является гипersфера с центром в начале координат и радиусом F_o , определяемые порогом, установленным на профессиональном уровне. Для всех объектов $M_i \{ Z_{i1}, Z_{i2}, \dots, Z_{im} \}$ вычисляют расстояния ρ_{Mi} от начала координат, которые и принимаются за обобщенные

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

Алгоритм 5. Распознавание по условной вероятности совместного влияния всех факторов. Значения Z_{ij} являются условными вероятностями принадлежности i -го объекта 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

$$\rho_i = \sqrt{\prod_{j=1}^m Z_{ij}}$$

Производится 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.

Построение решающего правила классификации. После оценки для каждого объекта вероятности отнесения его к классу А – *количественный* прогноз, т.е. принять решение – к какому из двух классов отнести объект M_i

Для этого используется простейшее решающее правило классификации:

$$\begin{cases} \text{класс А, } P(M_i \in A) \geq P_0 \\ \text{класс В, } P(M_i \in A) \leq P_0 \end{cases} \\ (P_0=0,5)$$

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.

$$P_0 = \frac{1}{1 + \frac{q_1 C_1}{q_2 C_2}}$$

Прогноз нового объекта осуществляется следующим образом:

- 1) По значениям фактора прослеживается его путь на графе определенная его обобщенная ветвь;
- 2) Вероятность принадлежности M_0 к классу A , равна итоговой вероятности обобщенной ветви;
- 3) Альтернативное решение о классе успешности нового объекта принимается на основе единого вероятностно-альтернативного решающего правила классификации.

Оценка надежности методов прогноза по контрольной выборке объектов – это заключительный этап оптимизации решения задачи. На нем 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 принадлежности объекта к классу A . Значения W_i сравниваются с порогом принятия решения p_0 , установленным из соображений специфики решаемой задачи на профессиональном уровне. Если $W_i < p_0$, то объект относится к классу B , если $W_i \geq p_0$, то к классу A (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: ошибки I рода- если объект класса A прогнозируется как объект класса B , ошибки II рода- если объект класса B прогнозируется как объект класса A .

По каждому j -му вычисляются вероятности ошибки прогноза p_{I-} - I рода, p_{II-} - II рода и вероятность суммарной ошибки:

$$p_I = \frac{n_1}{N_1}, p_{II} = \frac{n_2}{N_2}, p_{\text{ош}} = \frac{n_1 + n_2}{N_1 + N_2}.$$

И надежность прогноза

$$H = (1 - P_{\text{ош}}) 100$$

где n_1 и n_2 - число ошибок I и II рода соответственно,

а N_1 и N_2 – общее число объектов в контрольных выборках классов A и B .

Выбор оптимального метода прогноза и принятия решений для решаемой прикладной задачи проводится на основе алгоритма, обеспечивающего не только наибольшую надежность прогноза, но и минимум затрат на принятие решений. На практике имеет значение оперативность и наглядность проведения прогноза [4] и принятия решений о классе новых объектов, а также возможность управлять ими, т.е. такому изменению параметров процесса, которое при заданных начальных условиях позволит перевести прогнозируемый объект в «желаемый» класс.

Список литературы:

1. Загоруйко К. Г. Методы распознавания образов: состояние и перспективы. - М.: Сов. радио, 1972. - 120 с.
2. Васильев В. М. Распознающие системы. – Киев: Наукова думка, - 180с.
3. Гмурман В. Е. Теория вероятностей и математическая статистика. – М.: Высш. шк. 1972.-386 с.
4. Батугин А.С. Тектонофизическая модель горно-тектонических ударов с подвижками крыльев крупных тектонических нарушений. // ГИАБ, «Труды научного симпозиума «Неделя Горняка – 2010». 2010. с.252-264.
5. Ковалевский В. А. Методы распознавания изображений. - М.: 1975.- 380 с.
6. Васильев С. К., Захаров В. К., 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. Андерсон Т. Введение в многофакторный статистический анализ. - М.:Наука, 1968.-500 с.
9. Батугин С. А. Вопросы классификации распознавания / Механика горных пород. - Кемерово, 1972. 40-44 с.