Разность между фактическим значением этой характеристики и вычисленным по формуле дает значение Δ-характеристики. Сравним распределения исходных характеристик и Δ-характеристик.

Распределения исходных характеристик в разных классах в двумерном пространстве $(x_i; x_j)$ отличаются явной тенденцией к упорядоченности. Распределения Δ-характеристик $(\Delta x_i; \Delta x_j)$ не имеют подобной упорядоченности. Этот факт естественно объяснить в первом случае влиянием сильного фактора структуры, позволяющего разделить месторождение на зоны свода и крыльев структуры. Во втором случае это влияние отсутствует. Таким образом, можно считать, что в наборе Δ-характеристик влияние структурного фактора сведено к минимуму. В таком случае задача становится сложной и близкой по характеру второй граничной ситуации.

Набор Δ-характеристик $\{x\}$ обрабатывается с помощью алгоритма кодирования, использующего регрессионный анализ, по принципу, описанному выше, для получения набора параметров $\{t\}$. В сформированном таким образом пространстве параметров строится решающее правило с помощью алгоритма классификации, которое использовалось для разделения контрольных объектов.

Результирующая эффективность решения задачи оказалась более 90%. Увеличение эффективности можно было бы достичь, наращивая количество признаков с большим числом входящих в них параметров $\{t\}$. Но это усложнило бы правило и тем самым, возможно, уменьшило его устойчивость.

Таким образом, пример решения задачи разделения на основе полевых геофизических данных показывает эффективность системы преобразования, предназначенной специально для использования в сложных задачах дихотомии.

Ю.Л. Васильев, А.Н. Дмитриев

ПРОСТОЙ СПОСОБ СРАВНЕНИЯ ОБЪЕКТОВ, ОХАРАКТЕРИЗОВАННЫХ НАБОРОМ ПРИЗНАКОВ

1. Рассматривается вопрос об уточнении диагностики на основе той или иной числовой меры. Ранее авторы ввели на объектах и характеризующих их признаках числовую меру и дали способ вычисления ее, основанный на итерациях. Эта мера в некотором смысле отражает естественное соотношение между сравниваемыми объектами и оправдывается при практических применениях.

Способ вычисления отличается малой трудоемкостью и делает доступными задачи с большим числом признаков. На обрабатываемые таблицы налагается некоторое естественное структурное ограничение. Основная идея состоит в сведении рассматриваемого вопроса к задаче о собственных векторах (о спектре) линейного преобразования, и поэтому подход назван спектральным.

В дополнение к изложенному здесь дано некоторое обоснование быстрой сходимости итераций, наблюдаемой при работе с геологическими таблицами; приведены сведения о некоторых новых приложениях.

2. Пусть имеется h объектов $W_1, \ldots, W_h$ охарактеризованных l двузначными признаками $P_1,\ldots,P_l$. Пусть $T = (t_{ij})$ – таблица размера $h \times l$, которая составлена из единиц и нулей и в которой i-тая строка $(t_{i1},\ldots,t_{il})$ отвечает объекту W_i $(i = 1,\ldots, h)$, j-тый столбец $(t_{ij},\ldots,t_{hj})$ отвечает признаку P_j $(i = 1, l)$,), таблица Т отражает выраженность признаков у объектов, если для i-того объекта и j-того признака она превышает некоторый уровень, то $t_{ij} = 1$, а если меньше, то $t_{ij} = 0$. Условимся, что $h \geqslant 2$, $l \geqslant 2$ и что в таблице Т нет строк и нет столбцов, составленных сплошь из нулей. Если признаки K значны $(K > 2)$, то t_{ij} может принимать соответствующее число значений из отрезка $[0,1]$. Ограничимся рассмотрением случая $K = 2$, так как переход на случай $K > 2$ будет очевиден.

Таблица описывает объекты в связи с тем или иным фактором, и задача состоит в оценке последнего по картине, даваемой таблицей Т. Предполагается, что признаки $P_1,\ldots,P_l$ существенны для X, что набор их достаточно полон и что объекты $W_1,\ldots,W_h$ родственны по отношению к X. Пусть Т – геологическая таблица, X – запасы месторождений, объекты – месторождения одного и того же полезного ископаемого, которые родственны по признакам и сравнимы по запасам, но оценка последних требует уточнения. Нас интересуют случаи, когда число объектов сравнительно невелико (10 – 20), число признаков велико (десятки, сотни), влияния признаков находятся в сложном переплетении.

3. Вектор $\vec{a}\ (a_1,\ldots,a_m)$ называется положительным, если $a_1, a_2,\ldots,a_m > 0$. Норма вектора $\vec{a}\ (a_1,\ldots,a_m)$ – число $|\vec{a}|$, равное $\max\limits_{1 \leqslant i \leqslant m} |a_i|$; вектор $\vec{a}$ называется нормированным, если $|\vec{a}| = 1$.

Числовую меру для h объектов и l признаков естественно задавать в виде положительных нормированных векторов $\vec{\omega} = (\omega_1,\ldots,\omega_h)$ и $\vec{\pi} = (\pi_1,\ldots,\pi_l)$. Ниже мы определим их как нагрузку строк и нагрузку столбцов таблицы Т. По ним предлагается судить о проявленности фактора X в объектах $W_1,\ldots,W_h$, а также о степени влияния на нее признаков $P_1,\ldots,P_l$.

4. Исходя из таблицы Т и пары векторов $\vec{\omega} = (\omega_1,\ldots,\omega_h)$, $\vec{p} = (p_1,\ldots,p_l)$, определим пару векторов $'\vec{\omega} = ('\omega_1,\ldots,'\omega_h)$, $\bar{p} = (p_1,\ldots,p_l)$, а также нормирующие множители α, β и пару векторов $\vec{\omega}$ и $\bar{p}$:

$$'\vec{\omega} = T(\bar{p}), \text{ т.е. } '\omega = t_{i1}p_1 + \ldots + t_{il}p_l, \quad (i = 1,\ldots,h) \qquad \vec{\omega}' = \alpha \cdot {}'\vec{\omega}, \quad |\vec{\omega}'| = 1 \qquad (1)$$

$$'\vec{p} = T^*(\vec{\omega}), \text{т.е. } 'p_j = t_{1j}\cdot\omega_1 + \ldots + t_{hj}\cdot\omega_h \quad (j = 1, \ldots, \ell), \quad \vec{p}' = \beta'\cdot\vec{p}, \quad |\vec{p}'| = 1 \qquad (2)$$

(T^* – матрица размера $\ell \times h$, в которой строками служат столбцы матрицы T).

Определение. Пару нормированных векторов $\vec{\varphi} = (\varphi_1, \ldots, \varphi_h)$, $\vec{\psi} = (\psi_1, \ldots, \psi_\ell)$ назовем фракционирующей парой таблицы T, а сами векторы – фракционирующими векторами, если $\vec{\varphi}' = \vec{\varphi}$ и $\vec{\psi}' = \vec{\psi}$.

Определение. Фракционирующую пару векторов $\vec{\omega}$, $\vec{\pi}$ назовем нагрузкой таблицы T, а сами векторы – нагрузкой строк и столбцов, если эти векторы положительны.

Определение. Таблицу T, нагрузка которой существует и единственна, назовем измеримой.

Таблицу T назовем разрозненной, если среди всех ее h строк найдутся h_1 таких строк, каждая из которых ортогональна[1] каждой из прочих $h - h_1$ строк ($h_1 < h$). Иными словами, совокупность объектов распадается на две такие группы по h и $h - h_1$ объектов, что каждый из признаков обязательно представлен нулями во всех объектах какой-то одной из этих двух групп (иначе говоря, эти две группы "не имеют ничего общего"). Поэтому естественно считать, что "реальные" таблицы разрозненными не являются. Кроме того, можно доказать, что доля разрозненных таблиц среди всех таблиц размера $h \times \ell$ весьма мала (она $\to 0$ при $h \to \infty$, $h \leq \ell$, $\ell \leq 2^{h-1}$). Таблицу, не являющуюся разрозненной, назовем связной.

Нетрудно проверить, что любая разрозненная таблица неизмерима. Однако имеет место теорема 1: любая связная таблица измерима.

Способ вычисления нагрузок дает теорема 2: для любой измеримой таблицы T нагрузка строк $\vec{\omega}$ и нагрузка столбцов $\vec{\pi}$ являются пределами соответственно последовательностей векторов

$$\vec{\omega}^0, \vec{\omega}^1, \ldots, \vec{\omega}^n, \ldots, \text{ где } \vec{\omega}^0\overbrace{(1, \ldots, 1)}^{h \text{ единиц}} \text{ и } \vec{\omega}^n (\vec{\omega}^{n-1})' = T(\vec{\pi}^{n-1}),$$

$$\vec{\pi}^0, \vec{\pi}^1, \ldots, \vec{\pi}^n, \ldots, \text{ где } \vec{\pi}^0\overbrace{(1, \ldots, \ldots, 1)}^{\ell \text{ единиц}} \text{ и } \vec{\pi}^n (\vec{\pi}^{n-1})' = T^*(\vec{\omega}^{n-1}).$$

5. Следует отметить быструю сходимость итераций для всех геологических таблиц (около 50), которые были изучены данным методом (в получаемых числах третий знак устанавливался между 5-й и 10-й итерациями, а к 30-й итерации устанавливались 8-й – 9-й знаки). Это обеспечивает весьма малую трудоемкость счета, т.к. отдельная итерация весьма проста, а число итераций мало. Однако желательно не только эмпирическое обоснование быстрой сходимости. В линейной алгебре известны приме-

---

[1] Два вектора $(x_1, \ldots, x_\ell)$ и $(y_1, \ldots, y_r)$ ортогональны, если $x_1 y_1 + \ldots + x_n y_n = 0$.

ры, требующие даже при малой неточности огромного числа итераций. Нужны доводы, что такого рода примеры среди геологических таблиц маловероятны или невозможны.

Скорость сходимости итераций для таблицы T тем больше, чем меньше величина $|\lambda_2/\lambda_1|$, где λ_1 и λ_2 – собственные значения матрицы $\Phi = T \cdot T^*$, причем λ_1 – собственное значение, наибольшее по модулю; λ_2 – собственное значение, следующее по модулю за λ_1.

Известная теорема А. М. Островского для матрицы $\Phi = (\varphi_{ij})$, составленной из положительных чисел, дает оценку

$$\left|\frac{\lambda_2}{\lambda_1}\right| \leq \frac{M^2-m^2}{M^2+m^2}, \text{ где } M = \max_{ij} \varphi_{ij}; \quad m = \min_{i,j} \varphi_{ij}.$$

В применении к таблице T это означает, что скорость сходимости итераций будет большой, если всевозможные скалярные произведения строк будут нулевыми и будут по возможности мало отличаться друг от друга. Но эти свойства можно характеризовать как математическое выражение имеющего место в геологических таблицах переплетения влияний признаков. Это означает, что быстрота сходимости итераций в известной мере обеспечивается спецификой геологических таблиц. В этой связи были бы интересны оценки величины $|\lambda_2/\lambda_1|$ для неотрицательных матриц, а также соотношения, связывающие соответствующие параметры матриц Φ с параметрами исходных таблиц T.

6. Практические возможности спектрального подхода проверялись на геологических задачах различного содержания. В основном выбирались таблицы, для которых числовые оценки объектов или признаков были получены другими методами. Проводилось сравнение результатов спектрального и тестового подходов.

Наиболее полное применение спектрального подхода осуществлено для задачи сравнительного изучения гигантских месторождений нефти. В проблеме рудной геологии подход использован для задач прогноза и поиска месторождений. Так, по списку петрохимических и петрологических признаков была проведена диагностика трапповых интрузий в связи с их рудоносностью. В задаче сравнительного изучения россыпей Забайкалья спектральным методом были уточнены границы классов россыпей. По специализированному набору признаков была произведена предварительная разбраковка алмазоносных трубок Якутии. Уточнение ранжирования месторождений олова по запасам было проведено для Приморья. В задаче изучения редкометальных оруденений получены списки рудоконтролирующих признаков последовательной их проверкой на способность упорядочивать объекты по важности. При изучении химических составов тектитов различных групп было обнаружено подразделение австралитов и меймечитов на две подгруппы, что оказалось возможным после решения числовых таблиц большого размера.