第26卷 第4期 地质与勘探 Vol.26,No.4
1990年4月 GEOLOGY AND PROSPECTING April, 1990

模糊模式识别在成矿预测中的应用*

王硕儒 刘文滨 姜效典 张维岗
（青岛海洋大学）
曹洛华
（地矿部物化探研究所）
葛宗侠
（湖北地质研究所）

本文在模糊模式识别的理论基础上，提出了矿床预测的一种数学方法。基于F—PFS法，在特征因素和权重的调整下求出最佳分类。聚类中心即为标准模式。按单因素判对率作为该因素权重（逆距离权重），依标准模式模糊向量与已知单元模糊向量之间确定关系产生不等式方程组求解各因素的权重（距离权重），从而实现单元的贴近度识别。从鄂东南地区铜及多金属矿的预测表明，这类方法是有效的。

关键词：成矿预测；模糊模式识别；聚类分析；贴近度；权重

工作方法

模糊性存在于成矿预测之中，因此模糊模式识别可用于成矿预测。进行模糊识别，首先要用模糊数学方法建立标准模式。建模时，要正确选取显示矿床（矿田）存在的地质因素、地球化学特征、地球物理异常及其他信息的特征因素集。

特征因素集的选定是建立在矿床成矿规律的认识和矿床特征的基础上，并与工作目的紧密相关，小比例尺的成矿预测，主要解决区域性的资源评价和发现新的成矿区，因此应把反映区域性控矿因素和区域性物化探特征作为重点；较大比例尺的成矿预测，主要是寻找新的矿床，则应侧重表现矿床地质特征和物化探局部异常等因素，因素集应细微些。对小岩体的识别，则应包括岩体含矿性的表现因素，着重岩性、岩石地球化学、围岩蚀变和物化探异常特征等。不同比例尺、不同对象、目的的成矿预测，特征因素应有所区别，因素集的最终确定还必须通过实践逐步完成。

模糊聚类的识别模型

在聚类分析中，最简单的识别方法是按照待识别单元在聚类时的归属来确定其类型。这种识别极为直观，根据同类中已知单元类型，即可在一定程度上直接下结论。

然而，笔者采用聚类的主要目的却是为了按预测类型和目的建立标准模式，从而在

* 本文系国家“七·五”重点科技攻关55—05项目部分成果。

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

已聚成类的基础上对单元进行识别或再识别。

在聚类方法中，建立在模糊关系基础上的聚类法是从相似矩阵出发的。从相似矩阵到传递矩阵，虽然传递闭包具有传递性可以分类，但由于它是转换得来，且一般离相似矩阵甚远，是否能形成一个合理的分类尚待研究。实践已经表明，在样品数量较多时，其分类结果太粗，难以利用[1]，而且该法所得分类结果还受 λ 的选定所制约。

建立在自动迭代基础上的F—PFS聚类法是在最小二乘法上的最佳分类，很大程度上克服了系统聚类的缺点与不足。这是一种多次调用Isodata聚类法，能同时获得最佳分类数和分类的方法。每个单元隶属度为

$$u_{ij}=\frac{1}{\sum_{k=1}^{c}\left(\frac{\|X_j-V_i\|}{\|X_j-V_k\|}\right)^{2/(m-1)}} \quad (1)$$

且各类聚类中心为

$$V_i=\frac{\sum_{j=1}^{n}u_{ij}^m X_i}{\sum_{j=1}^{n}u_{ij}^m} \quad (2)$$

其中最佳分类数 c 按下式

$$\text{F-PFS}=\frac{t_r(S_B)(n-c)}{t_r(S_W)(c-1)}\ [2] \quad (3)$$

确定。当F—PFS出现极大值时，其对应的 c 即为最佳分类数。式中 $t_r(\cdot)$ 为矩阵 $(\cdot)$ 的迹，而

$$S_B=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^m V_i V_i^T \quad (4)$$

和

$$S_W=\sum_{i=1}^{c}\sum_{j=1}^{n}u_{ij}^m(V_i-X_j)(V_i-X_j)^T \quad (5)$$

采用F—PFS法仍可能出现明显不合理的分类情形，这与特征因素的选定和因素权重的分配有关。

为此，还需在多数已知单元的聚类中调整因素及权重，并使聚类在原始数据的情况下进行，这样，由于未作正规化处理，各因素值的大小、量级实际是表示其不同权重；也可以对某些因素值作适当放大或缩小（线性或非线性）进行调整。

笔者还提出一种根据单因素择近原则判别的判对数来调整因素和定权重的方法。具体作法是，先求出已知类各因素的标准值（如用平均值）x_{ik}；然后将已知单元各因素值 y_{ij} 与之比较。若有 $d_j=\min|y_{ij}-x_{ik}|$，则第 j 个单元按第 i 个因素应判属 k 类。由此，可对每个已知单元逐个按单因素识别。其判对率为

$$W_i=m_i/m \quad (6)$$

式中，m 为已知单元总数；m_i 为第 i 个因素判别时的判对数。

通过调整后的模糊聚类，常出现两种结果：1）分类较理想，符合实际；2）仍有少量明显的不同类型单元出现在同一类中，这时可对这种个别类单独重新聚类，或将其中的单元作为待识别单元进行贴近度识别。

采用上述方法得到的各类，都应具有一定的地质意义，如代表不同的矿种或不同的预测等级等。各类的聚类中心即可代表各类模型的标准模式。

采用此法在鄂东40个小岩体建立的类模型已取得了较满意的结果。

该组岩体多为已知，特征因素从12个地球化学指标调整到8个。求得F—PFS值与分类数的关系如图1所示，获得4类标准模式（表1）。

A 类全为以Cu为主的中、大型矿（4个岩体）；B 类为含大、中、小型Cu、W、Au矿4处（8个岩体）；C 类为具W（小型）、Cu（大型）矿各1处、矿点（矿化点）多处（11个岩体）；D 类基本为全无意义的岩体（17个岩体）。

对未知单元的识别，除用上述聚类直接

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

各类聚类中心　　表1

类别	Cu	Au+Ag+Bi	Pb	Ag	K_2O+Na_2O	K_2O/Na_2O	Mo	W+Bi
A	409.9	217.6	66.3	0.51	5.28	7.75	6.53	45.9
B	158.4	94.5	57.5	0.26	4.82	6.17	2.31	12.9
C	76.5	69.8	52.4	0.25	4.73	6.79	3.69	13.1
D	16.7	25.5	45.1	0.05	4.69	3.24	1.00	6.2

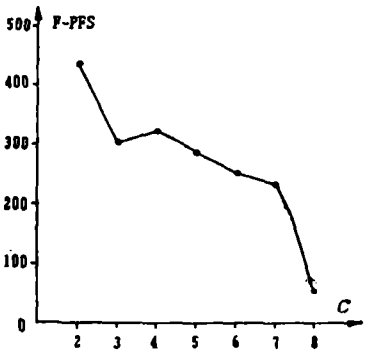

图1　岩体最佳分类数的关系

识别外，可按如下定义的贴近度确定其归属。

设待识别单元集合中第 j 个元素的模糊向量为$(x_{j1}, x_{j2}, \cdots, x_{js})$；标准类单元集合中第 k 个元素的模糊向量为 $(y_{k1}, y_{k2}, \cdots, y_{ks})$。由此，可提出贴近度

$$S(\underset{\sim}{X}, \underset{\sim}{Y})=\sum_{i=1}^{s} W_i'(1-d_{ijk})^P \qquad (7)$$

式中，$d_{ijk}=|x_{ji}-y_{ki}|$，则$1-d_{ijk}$为距离的逆运算（余）；W_i'为第 i 个特征因素的逆距离权重，$W_i'W=_i \Big/ \sum_{i=1}^{s} W_i$ 可按（6）式确定；P为调整未知单元从属于各类对比度的适当参数。

距离权重求解的贴近度识别法

众知，贴近度识别在模式识别中占有很重要的位置。为使模糊识别模型更趋完善，兹提出在给定已知类型单元情形下，一种新的建模和识别方法。

该法基本包括两个内容：

1．数据模糊化（模糊向量的建立）

原始数据常为确定性的数据，为使其模糊化，可采用类似于数据正规化的方法，即

$$x_{ij}'=\begin{cases} 1 & x_{ij}\geqslant x_{j\max} \\ \dfrac{x_{ij}-x_{j\min}}{x_{j\max}-x_{j\min}} & x_{j\min}<x_{ij}<x_{j\max} \\ 0 & x_{ij}\leqslant x_{j\min} \end{cases} \qquad (8)$$

式中，$x_{j\max}$、$x_{j\min}$分别为各单元中第 j 个特征因素的极限阀值（不必为最大值和最小值）。

这种正规化数值仅表示对“大小”这一模糊概念的隶属程度。要说明的是，这种模糊化应在所有单元（已知和未知）中统一进行。

在数据模糊化的基础上，将各类型各单元（已知）的模糊化数据分别按各特征因素确定各类模糊化数据的平均值，即

$$a_{ij}=\frac{1}{n_i}\sum_{k=1}^{n_i} x_{ij}' \qquad (9)$$

式中，$i=1, 2, \cdots, c$；$j=1, 2, \cdots, s$；n_i 为第i类型的已知单元个数。由此，第 i 个标准模型模式为$\underset{\sim}{A_i}=(a_{i1}, a_{i2}, \cdots, a_{is})$。

2．各因素权重求解的数学原理及方法

设有已知模式$\underset{\sim}{A_1}, \underset{\sim}{A_2}, \cdots, \underset{\sim}{A_c}$，且有单元$\underset{\sim}{X_{A1}}\in \underset{\sim}{A_1}$，$\underset{\sim}{X_{A2}}\in \underset{\sim}{A_2}$，…，$\underset{\sim}{X_{Ac}}\in \underset{\sim}{A_c}$。每一模式的模糊向量为$\underset{\sim}{A_i}=(a_{i1}, a_{i2}, \cdots, a_{is})$，$i=1, 2, \cdots, c$。而任意一单元的模

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net

糊向量为$\underset{\sim}{X}_j=(x_{j1}, x_{j2}, \cdots, x_{js})$。

对$\forall \underset{\sim}{X}_{A1}$，都存在

$$d(\underset{\sim}{X}_{A1},\underset{\sim}{A}_1)<d(\underset{\sim}{X}_{A1},\underset{\sim}{A}_2),\ d(\underset{\sim}{X}_{A1},\underset{\sim}{A}_1)<d(\underset{\sim}{X}_{A1},\underset{\sim}{A}_3),\ \cdots,d(\underset{\sim}{X}_{A1},\underset{\sim}{A}_1)<d(\underset{\sim}{X}_{A1},\underset{\sim}{A}_c)$$

而

$$d(\underset{\sim}{X}_{A1},\underset{\sim}{A}_i)=\sum_{j=1}^{s}W_j\,|x_{A1j}-a_{ij}|^{P_j}$$

$$i=1,\ 2,\ \cdots,\ c$$

同理，对$\forall \underset{\sim}{X}_{A2}$，$\forall \underset{\sim}{X}_{A3}$，…，$\forall \underset{\sim}{X}_{Ac}$，也有如下关系

$$d(\underset{\sim}{X}_{A2},\underset{\sim}{A}_2)<d(\underset{\sim}{X}_{A2},\underset{\sim}{A}_1),\ d(\underset{\sim}{X}_{A2},\underset{\sim}{A}_2)<d(\underset{\sim}{X}_{B},\underset{\sim}{A}_3),\cdots,d(\underset{\sim}{X}_{A2},\underset{\sim}{A}_2)<d(\underset{\sim}{X}_{A2},\underset{\sim}{A}_c)$$

$$\vdots \qquad\qquad \vdots \qquad\qquad \vdots$$

$$d(\underset{\sim}{X}_{Ac},\underset{\sim}{A}_c)<d(\underset{\sim}{X}_{Ac},\underset{\sim}{A}_1),\ d(\underset{\sim}{X}_{Ac},\underset{\sim}{A}_c)<d(\underset{\sim}{X}_{Ac},\underset{\sim}{A}_2),\cdots,d(\underset{\sim}{X}_{Ac},\underset{\sim}{A}_c)<d(\underset{\sim}{X}_{Ac},\underset{\sim}{A}_{c-1})$$

由此，选取不同类型的已知单元，可得出在约束条件$\forall_j$，$W_j>0$及$\sum_{j=1}^{s}W_j=1$下，统一的线代数不等式方程组

$$\begin{cases} W_1x_{11}+W_2x_{12}+\cdots+W_sx_{1s}<0 \\ W_1x_{21}+W_2x_{22}+\cdots+W_sx_{2s}<0 \\ \vdots \qquad\qquad \vdots \qquad\qquad \vdots \\ W_1x_{m1}+W_2x_{m2}+\cdots+W_s\,x_{ms}<0 \end{cases} \tag{10}$$

选择适当m值可解（10）式，一种方法是按线性规划求解。这样，该数学问题可变为在上述约束条件及（10）式约束下，求解目标函数$_{max}W_j$和$_{min}W_j$课题。

确定权重后，即可按（11）式对未知单元$\underset{\sim}{X}$进行识别。

$$S(\underset{\sim}{X}_j,\underset{\sim}{A}_i)=1-\frac{1}{b}\sum_{k=1}^{s}W_k\,|x_{jk}-a_{ik}|^{P_k}$$

$$i=1,\ 2,\ \cdots,\ c \tag{11}$$

式中，P_k可随因素的不同进行调整其数值；b为调整对比度的适当参数。

用上述方法，对鄂东39个已知铁帽两两对比建立模型和进行识别。

采用Cu，Pb，Zn，Ag，Mo，W，Bi，Au 8种地球化学元素组成特征因素集。铜矿与铅锌矿构成的对比模型使用4个已知铜，3个已知铅锌矿建立不等式组；铜矿与铁矿的识别模型各用3个已知样建立；铅锌矿与铁矿各用4个已知样建立对比识别模型。分别按线性规划求解各模型的权重集。计算中，$P_k=2$（$\forall k$），b取2。例如求解的Pb/Zn—Fe模型的权重集为：$W_{Cu}=0.0102$；$W_{Pb}=0.2143$；$W_{Zn}=0.2435$；$W_{Ag}=0.2556$；$W_{Mo}=0.1408$；$W_W=0.0405$；$W_{Bi}=0.0203$；$W_{Au}=0.0748$。

应用这样的两两对比模型分别对已知样进行检验，其判对率Cu—Pb/Zn模型为89.7%，Cu—Fe模型86%，Pb/Zn—Fe模型100%。

鄂东南区域单元预测

鄂东南地区位于长江中下游铜、铁成矿带的西端。该区燕山期中酸性岩浆活动强烈，岩体侵入于石炭—三叠纪的碳酸盐岩及志留纪砂页岩和早期形成的岩体中，这为形成接触交代型和斑岩型铜及多金属矿床提供了热液来源。因此，区域地层与构造岩浆岩均为该区控矿的直接因素。

将该区划分为78个单元（图2），所有单元都具有类似的地质背景、相似的控矿条件和成矿作用。单元的边界多采用不同岩性的界面或断裂带，局部覆盖地段适当外推，并参考物、化探异常分布特点予以确定。

不同类型单元在出露地层、断裂构造、矿化露头分布密度、矿化露头地化指标、物性、航磁异常、重力异常以及Cu、Pb、Au、W、Mo等元素原生晕、水系沉积物异常、规模都有一定的差异。

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

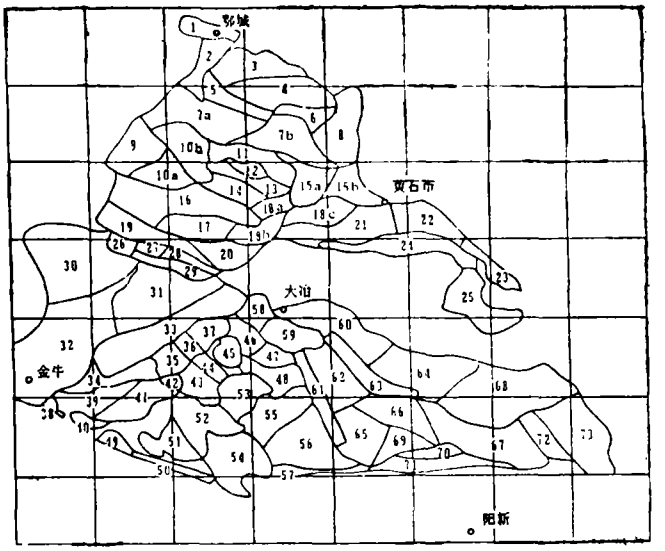

图 2 鄂东南成矿预测单元划分

选用26个因素作F～PFS聚类，确认分6类较为适当。其中A_1类主要反映以Cu为主的矿化集中区，有CuFe、CuMo、CuW等大中型矿床6处。B_1、C_1、D_1类均反映以PbZn矿化为主的单元，但在矿化特征上有差异。B_1可属一级，表现为矿化组份较复杂（Pb、Zn、Ag、Au、Cu），其中71单元为大型铅锌银矿床，这类条件较好。C_1类矿化组份较简单（Pb、Zn、Ag、Au）具有一定矿化意义，D_1类组份简单，矿化较弱。6类中另两类绝大部分为未知单元，难于识别，将其与前4类作贴近度（7）式计算。按择近原则确切识别，由此得出总分类，如表2所示。

表2中已知矿床58单元虽未确切划归A_1，但却极为贴近：$S(D_1,58)=0.5629$；$S(A_1,58)=0.5627$。

根据勘探程度，该区可划分出48个已知单元，又可分为Cu矿单元（A_2：9处），含Cu单元（B_2：12处），已见弱Cu矿化单元（C_2：17处），非Cu单元（已知纯Fe矿单元D_2：10处）。对其分别按第二种数学模型建立模式和识别。

为突出主因素的作用，可在（6）式的基础上筛选组成特征因素集（判对率低的因素舍去）。该特征因素集包括岩石中Cu含量、标准估值，水系沉积物中Cu含量及异常规模；水系沉积物Au和W(Mo)含量；岩石中Pb/Cu值；地层出露面积；断裂构造长度；矿化露头分布密度等10个变量。解（10）式得出各因素权重，依上述顺序分别为0.010、0.054、0.008、0.211、0.189、0.005、0.160、

单元分类最终结果 表 2

类型	聚类结果	贴近度（7）式识别的单元
A_1	14、18a、42、48、59、60、61、70	11、13、35、36、38、44、53、63
B_1	24、32、47、71、72	7a、39、40、49、50、51、54、64、65、66、67、69
C_1	18c、21、22、33	6、19、55、56、73
D_1	8、15a、23、46	1、2、3、4、5、7b、9、10a、10b、12、15b、16、17、18b、20、25、26、27、28、29、30、31、34、37、41、43、45、52、57、58、62、68

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved. http://www.cnki.net

0.097、0.101、0.074。

表2中，14，42，48，59，61，70，11，13，36，63，58为CuAu为主的矿床；71为PbZnAg为主的矿床；43,57为WMo为主的矿床。

对所有单元分别用（11）式贴近度按择近原则进行识别。

已知性质单元的判对率：Cu矿类为100%，含Cu类为75%，弱Cu类为88%，非Cu类为70%。总计识别为A_2类和B_2类的单元如表3所示。

（11）式识别的有意义单元　　表 3

Cu矿	11、14、35、36、42、48、58、59、61、63、66、70、71、72
含Cu	13、18a、38、40、44、53

表3中，11、14、36、42、48、58、59、61、63、70、13为CuAu为主的矿床；71为PbZnAg为主的矿床。

在上述两类模型识别的基础上，可以对单元作出综合评价，其方法是对属于成矿极有利的类型求两种评价结果的交集。由此，该区域最有利成矿的单元集为

$U=\{11, 13, 14, 18a, 35, 36, 38, 40, 42, 44, 48, 53, 58, 59, 61, 63, 66, 70, 71, 72\}$

判对的已知矿床单元占总矿床数的85.7%。以Cu为主的矿床无一漏判，错判的皆为以W、Mo为主的矿床。这可认为钨钼矿的特征因素与铜矿是有差异的，这已被该区矿床与变量之间的相关性研究结果证实。

根据鄂东南区域具体情况，各单元的已知含矿性可用其含矿指数来表示。它是该单元内含矿系数（与矿体规模成正比）与其对应含矿数量乘积之和，被该单元面积去除的1个数。该数值直接表明了其单元的含矿性。上述预测单元（未知的）除35单元极不显著外，其余单元的含矿指数分别为18a单元：Cu2.3，PbZn0.7；38单元：Cu3.8；53单元：Cu1.3；72单元：Cu1.3，Au0.4，Ag0.6；66单元：Cu0.3；44单元：Cu1.0，Au1.4，WMo1.0。

结　　语

模糊模式识别是对成矿靶区从定性走向定量预测数学方法又一新的途径。模糊模式识别模型可对岩体和区域单元进行较准确的判别其含矿性。

识别的准确度很大程度上取决于地质，地球化学，地球物理特征因素的选定和各因素权重的正确确定以及建模的方法。因素和权重可通过对区域成矿规律的认识和各类异常来初选，又可按单因素的判对率高低及解线代数不等式方程组予以肯定和赋值。而各类的标准模式则可按其聚类中心或模糊化数据的平均值来表示。

两种数学模型在对鄂东南地区铜及多金属矿识别应用中，判对率高，预测效果也较符合实际，据此，可参照此数学模型对其他矿种予以建模和识别。

参 考 文 献

[1] 王硕儒等，海洋通报，1987，第6卷，第1期，第31～37页

[2] 张伟，模糊数学，1987，第3～4期，第51～56页

Metallogenic Prognosis Using Fuzzy Pattern Recognition

Wang Shuoru　Cao Luohua　Ge Zongxia

Liu Wenbin　Jiang Xiaodian　Zhang Weigang

On the basis of fuzzy pattern recognition, a kind of mathematic methods for metallogenic prognosis are proposed in this present paper. Under the adjustment of characteristic factors and weights and by using F-PFS method an optimum classification may be obtained. The cluster centre is considered as a standard pattern. According to the rate of right judgement of the single factor, which taken as the complement weight of the distance of the factor, and on the basis of the relationship determined between the fuzzy vector of the standard pattern and that of the known cells, a system of linear algebraic inequality could be established and solved for distance weights of different factors. Thus the recognition of unknown cells is realized by proximity degree. These methods have been used in the prognostication of copper and multinetal ores in Southeast Hubei areas and were proved to be effective for this end.

© 1994-2009 China Academic Journal Electronic Publishing House. All rights reserved.　http://www.cnki.net