

基于 RAGA 的 PPC 模型 在化探异常分析中的应用 ——以凤太矿集区为例

廖时理, 张 均, 张利亚, 姚 涛
(中国地质大学(武汉) 资源学院, 武汉 430074)

摘要: 化探数据处理的核心内容是地球化学背景与地球化学异常的合理区分, 找矿信息提取将直接影响到成矿预测的结果。文章采用基于 RAGA 的 PPC 模型对凤太矿集区 1 : 5 万水系沉积物化探数据进行处理, 初步探讨了投影寻踪聚类方法在地球化学异常识别与元素共生组合方面的应用, 以及投影方向值所代表的地质意义。结果表明, PPC 模型的化探异常识别效果优于传统方法, 为多维化探数据的处理提供了一种新的思路和途径。

关键词: PPC 模型; 化探异常提取; 凤太矿集区; 陕西省

中图分类号: P632; P618.4 **文献标识码:** A **文章编号:** 1001-1412(2011)01-0090-06

0 引言

地球化学元素分布规律的研究是揭示元素矿化富集及空间变化规律的重要途径之一, 传统的多元分析方法是建立在研究总体服从正态分布的基础上的, 然而实际情况中的化探数据大多数并不满足正态假定这一前提, 而且当数据的维数较高时, 即使数据的样本点很多, 散布在高维空间仍然会显得非常稀疏。这种情况下, 传统的多元分析方法将会受到局限。投影寻踪聚类方法(PPC)的基本思想是把多维数据投影到低维(1~3 维)的子空间上, 寻找出能反映高维数据结构或特征的投影, 从而能够更加直观地研究和分析高维数据^[1]。最近几年来, 投影寻踪方法在水利资源、农业、军事等许多领域都有了广泛应用, 并取得了较好的效果, 但是在地学领域, 特别是金属矿产勘查领域的研究应用几乎还处于空白状态, 仅少数人进行了一些探索性工作, 如肖凡(2008)^[2]首次将投影寻踪理论引入矿产勘查领域, 并初步探讨了与分形理论结合的可能性。基于此,

本文提出采用基于实码加速遗传算法(Real coding based Accelerating Genetic Algorithm, 简称 RAGA)^[3]的投影寻踪聚类(Projection pursuit cluster, 简称 PPC)^[4]化探数据处理模型的新途径, 并进行了实例研究。

1 投影寻踪聚类(PPC)化探数据处理模型

投影寻踪聚类技术的实质是一种降维处理技术, 即通过投影寻踪技术将多维分析问题通过最优投影方向转化为一维问题, 并利用投影特征值与投影方向向量值进行综合评价。其具体建模过程如下^[1,5-9]:

(1) 样本评价指标集的归一化处理。

(2) 构造投影指标函数 $Q(a)$ 。投影寻踪就是把 p 维数据 $\{x(i, j) | j = 1, 2, \dots, p\}$ 合成以 $a = \{a(1), a(2), a(3), \dots, a(n)\}$ 为投影方向的一维投影值 $z(i)$,

$$z(i) = \sum_{j=1}^p a(j)x(i, j) \quad i = 1, 2, \dots, n$$

收稿日期: 2009-10-26

基金项目: “十一五”国家科技支撑计划课题(编号: 2006BAB01A11)资助。

作者简介: 廖时理(1986), 男, 湖南岳阳人, 硕士研究生, 矿产普查与勘探专业, 从事成矿规律与成矿预测方面研究。通信地址: 中国地质大学(武汉)资源学院数地所, 邮政编码: 430074; E-mail: yyx yzsl@126.com

式中, x 为单位长度向量, 然后根据 $\{z(i) | i = 1, 2, \dots, n\}$ 的一维散布图进行分类。综合投影指标值时, 要求投影值 $\{z(i) | i = 1, 2, \dots, n\}$ 具有如下散布特征: 局部投影点尽可能密集, 最好凝聚成若干个点团, 而在整体上, 投影点团之间尽可能散开。因此, 投影指标函数可以表达成:

$$Q(a) = S_z D_z$$

式中, S_z 为投影 $z(i)$ 的标准差, D_z 为投影值 $z(i)$ 的局部密度, 即

$$D_z = \sum_{i=1}^n \sum_{j=1}^n (R - r(i, j)) \cdot u(R - r(i, j))$$

$$S_z = \sqrt{\frac{\sum_{i=1}^n (z(i) - E(z))^2}{n-1}}$$

式中, $E(z)$ 为序列 $\{z(i) | i = 1, 2, \dots, n\}$ 的平均值; R 为局部密度的窗口半径, 它的选取既不能使它随着 n 的增大而增加太高, 又要使包含在窗口内投影点的平均个数不太少, 避免滑动偏差太大, R 可以通过实验来决定; $r(i, j)$ 表示样本之间的距离, $r(i, j) = |z(i) - z(j)|$; $u(t)$ 为一单位阶跃函数, 当 $t \geq 0$ 时, 其值为 1, 当 $t < 0$ 时, 其函数值为 0。

(3) 优化投影指标函数。当各指标的样本集给定时, 投影指标函数 $Q(a)$ 只随投影方向 a 的变化而变化。由于高维空间上的点可以通过不同的投影方向映射到低维子空间上, 而不同的投影方向则反映了不同的数据结构特征。其中使投影指标函数达到最大的投影方向即为我们所求的最佳投影方向, 它能够最大可能地暴露高维数据的某类特征结构, 因此, 可以通过求解投影指标函数最大化问题来估计最佳投影方向, 即最大化目标函数 Max:

$$Q(a) = S_z D_z$$

约束条件:

$$s. t. \sum_{j=1}^n a^2(j) = 1$$

这是一个以 $\{a(j) = 1, 2, \dots, p\}$ 为优化变量的复杂非线性优化问题, 用传统的处理方法较难。因此, 本方法应用模拟生物优胜劣汰与群体内部染色体信息交换体制的基于实数编码的加速遗传算法 (RAGA) 来解决高维全局寻优问题。由于文章篇幅在此不做详述, 其具体过程参见文献 [1]。

(4) 分类 (优序排列)。把步骤 (3) 求得的最佳投影方向 a 代入下式:

$$z(i) = \sum_{j=1}^p a(j) x(i, j) \quad (i = 1, 2, \dots, n)$$

由此即可得个样本点的投影值 $z(i)$ 。将 $z(i)$

和 $z(j)$ 进行比较, 二者越接近, 表示样本 i 与 j 越倾向于分为同一类, 若按 $z(i)$ 值从大到小排序, 则可以将样本从优到劣进行排序。

2 区域地质概况

凤太层控金-铅锌多金属成矿区位于秦岭泥盆系多金属成矿带中段。区域构造属于扬子古陆边缘秦岭微板块内北半部的构造断陷沉降带中的一级拉分盆地。

出露的地层主要为泥盆系, 中上泥盆统 (D_{2-3}) 约占总出露面积的 85%; 其次为石炭系、侏罗系及白垩系^[10]。中上泥盆统 (D_{2-3}) 以酒奠梁-王家院断裂和赵家庄-都督门断裂为界, 可划分为 3 部分: 南部泥盆系为一套碳酸盐岩、浅变质细碎屑岩; 中部泥盆系为一套细碎屑岩-碳酸盐岩-细碎屑岩夹碳酸盐岩-细碎屑岩; 北部泥盆系为一套浅变质的细碎屑岩, 是该区的主要含矿层位^[11]。其中, 古道岭组 (D_{2g}) 灰岩与星红铺组 (D_{3x}) 千枚岩的接触面是区内铅锌矿的主要产出层位。

区内的基本构造格架为古盆河-殷家坝大型复式向斜, 构造线方向为近 EW 向, 次级褶皱极为发育, 褶皱形态紧闭至倒转, 呈线状排列; 该复式向斜西被区域性大断裂所截, 东被华阳花岗岩侵入^[10]。断裂构造发育, 南北两侧均受深大断裂控制, 北为凤县-山阳断裂, 南部为酒奠梁-狮子坝断裂 (向东延伸并入江口-镇安-板岩镇 EW 向断裂中)。此外, 区内还广泛发育有一系列的次级断裂, 主要有 NWW 向、NNE 向、NW 向和 NE 向等 4 组^[11]。

区内岩浆岩主要有分布在古盆河-殷家坝复向斜东端的狮子岭花岗闪长岩体 (图 1), 出露面积约 150 km², 还有若干中酸性小岩体。区内普遍发育花岗斑岩脉和闪长玢岩脉, 尤其在河口-八方山-王家楞一带发育较多。岩体分布与区域主体构造线方向一致, 明显受区域性深大断裂的控制^[12]。此外, 矿集区西侧深断裂带及其附近有小型基性-超基性岩体出露, 东北部北侧有苏家沟超基性岩体出露^[10]。

3 处理结果与解释

本文所采用的数据为凤太矿集区东部 1:5 万水系沉积物测量数据, 每个样品共分析有 16 个元

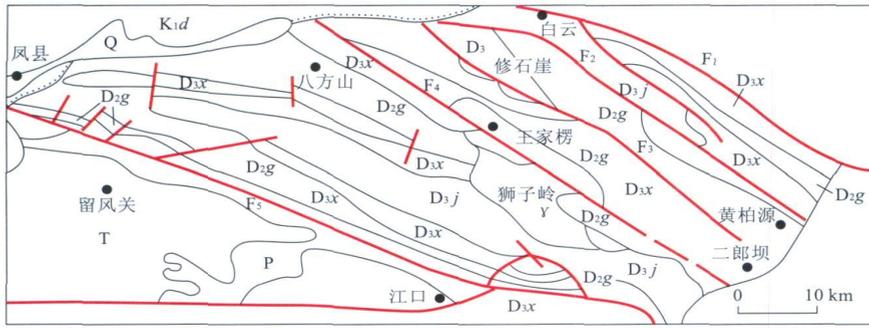


图1 陕西凤县太白地区地质略图(据八方山详查报告,有修改)

Fig 1 Geological sketch map of Taibai and Fengxian in Shanxi Province

F₁. 湘子河—黄柏源断裂 F₂. 修石崖—观音峡断裂 F₃. 王家楞—二郎坝断裂 F₄. 倒回沟—拓梨园断裂
F₅. 洒奠梁—江口断裂 T. 三叠系 P. 二叠系 D_{3j}. 上泥盆统九里坪组 D_{2x}. 中泥盆统星红铺组
D_{2g}. 中泥盆统古道岭组 D_{2w}. 下泥盆统王家楞组 Y. 印支期花岗岩

表1 凤太矿集区东部1:5万水系沉积物测量主要元素特征统计表

Table 1 Statistics of main elements analysis of river sediment samples collected at scale 1:50000 in eastern Fengtai mineral cluster region

元素	样品数	最小值	最大值	平均值	标准差	变异系数	偏度系数	峰度系数	异常下限
Au	3530	0.8	3.7	1.43	0.758	0.530	1.794	2.487	3.016
Ag	3530	50	120	61.2	18.700	0.306	1.866	2.900	98.315
Cu	3530	15	30	22.9	4.430	0.193	0.002	-0.756	31.760
Pb	3530	20	75	35.1	13.000	0.370	1.836	3.095	61.100
Zn	3530	50	256	87.97	48.250	0.549	2.592	5.951	184.470
As	3529	6.5	45	16.2	9.000	0.556	1.891	3.401	34.200
Sb	3529	0.5	3.8	1.45	0.852	0.587	1.329	1.113	3.154
Hg	3529	20	403	68.75	98.610	1.434	2.509	5.186	265.970
Mn	3530	400	850	621.5	122.800	0.198	0.187	-0.649	867.100
Mo	3527	0.5	1	0.67	0.160	0.232	0.730	-0.490	0.990
Bi	3529	0.2	0.4	0.32	0.071	0.219	-0.395	-0.971	0.462
Cr	3529	70	100	89	7.990	0.090	-0.617	-0.004	104.980
Co	3530	15	25	19.8	3.570	0.181	0.062	-1.052	26.940
Ni	3530	20	40	30.4	5.230	0.172	-0.031	-0.433	40.860
V	3529	75	110	95.3	9.443	0.099	-0.436	-0.384	114.186
Ti	3529	3500	6000	4956.4	623.560	0.126	-0.662	0.096	6203.520

数据来源:西北有色金属地质勘查局七一一总队;量的单位: $w(\text{Au, Ag, Hg, Ti})/10^{-9}$,其他元素 $w_B/10^{-6}$ 。

素,对数据进行特异值处理后以 $x+2\sigma$ 为异常下限,其统计特征如表1所示。

采用PPC模型的建模方法和过程对本批数据进行处理,为取得较理想的效果,在经过大量的试验后确定对凤太矿集区1:5万水系沉积物化探数据采用的具体参数为,初始种群数400,优秀个体数 $N=20$,密度窗宽 $R=0.1S_z$,交叉率 $p_m=0.8$,变异率 $p_c=0.85$,加速次数为20次。处理结果如下:

投影方向: $a=(0.08617, 0.35821, 0.02508, 0.38521, 0.35112, 0.36367, 0.21673, 0.40255, 0.16241, 0.18104, 0.35303, 0.23035, 0.04984,$

$0.06231, 0.08420, 0.01975)$

投影值: $z=(0.25524, 0.58012, 0.19979, 0.37939, 0.30957, 0.34071, 0.25007, 0.25514, 0.51125, 0.33387, 0.51142, 0.32357, 0.28028, 0.34686, 0.29118, 0.34246, 0.27859, 0.26065, 0.29148, 0.23728, 0.27261, 0.34194, 0.28005 \dots)$

对投影值进行基本的统计分析,采用 $x+2\sigma$ 作为异常下限,其特征信息如表2所示。

分别用传统的多元统计方法和PPC模型投影值做异常图进行对比:

表 2 PPC 模型投影值 z 特征统计表

Table 2 Statistics of characteristic z value of PPC model projection

投影值	样本数	最小值	最大值	平均值	标准偏差	变异系数	偏度	峰度	异常下限
z	3526	0.118	1.857	0.321	0.113	0.351	5.287	41.024	0.546

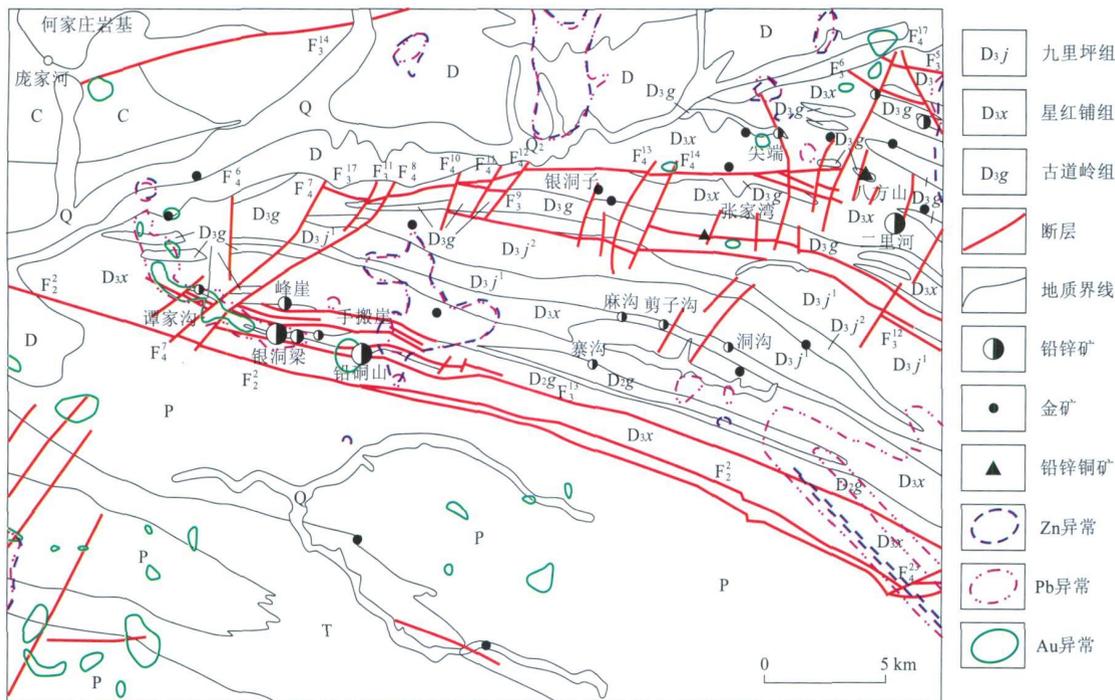


图 2 凤太矿集区 Pb, Zn, Au 综合异常图

Fig. 2 Pb, Zn, Au composite anomaly map of Fengtai area

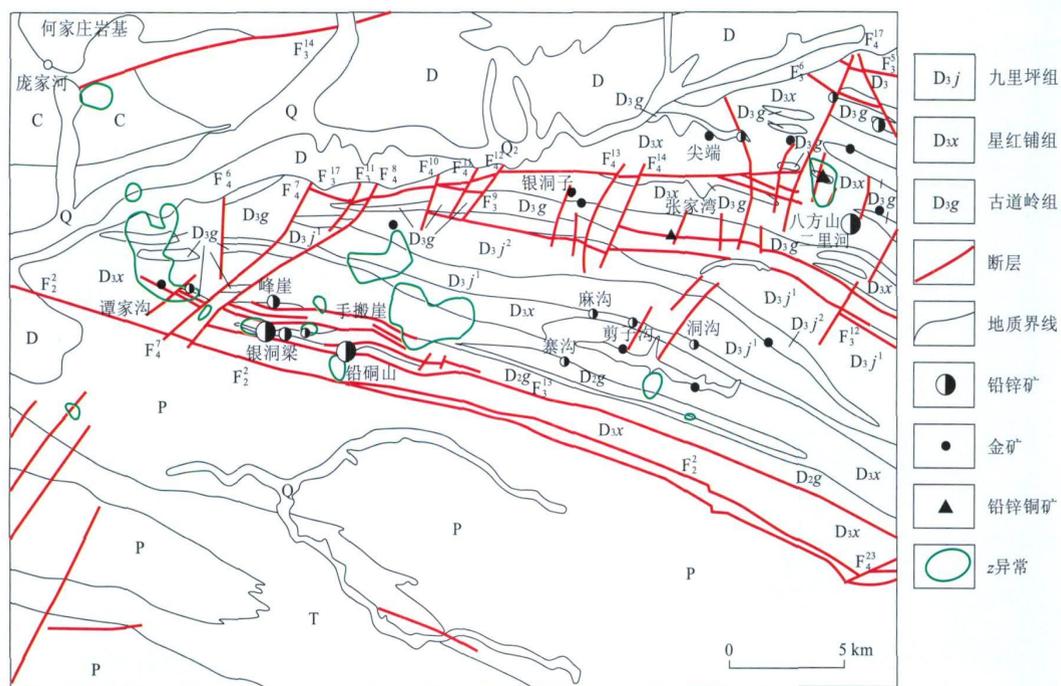


图 3 投影值异常图

Fig. 3 Projection value anomaly map

如图 2 和图 3 所示, 结合研究区的成矿规律与控矿因素^[12, 14], 常规方法所得到的异常中, Pb, Zn 的异常范围较大, 而且在针对本区成矿规律而言应该是无远景的地区亦出现了大片的异常; Au 的异常区域则显得杂乱无章, 与已知矿点的吻合性很差。无论是异常的形态规模还是分布, PPC 模型的预测结果与实际的矿床(点)、分布的吻合度都要比综合异常图的结果好, 比较而言, PPC 模型的分析结果更加突出了一些与成矿有利部位的异常, 减少了异常查证面积而没有丢失异常信息。

用基于 RAGA 的 PPC 模型对陕西凤太矿集区水系沉积物中 16 种元素进行分析, 根据最佳投影向量 a 值将它们由大到小排序并作折线图(图 4)。

根据折线图的变化趋势并结合实际情况, 以 0.3 和 0.1 为分界点可将投影元素分为 3 类: I 类为 As, Hg, Zn 和 Pb; II 类为 Ti, Sb, Co, Mn, Bi, Cu, Au 和 Ni; III 类为 Mo, Ag, V 和 Cr。I 类元素的值比较大, 可能与本区为喷流沉积型的铅锌矿床有关, Pb, Zn 为本区的主要成矿元素, 在研究区内明显富集, 显示投影方向值对主要成矿元素具有一定的突出作用。另外, 研究区内 As-Hg-Zn-Pb 元素组合具有层控型铅锌矿床的特征^[13], 由折线图得到的 I 类元素组合则恰好反映了这一特征。II 类元素包括本区的主要成矿元素 Au, Cu 及其他一些相伴生的元素, 投影向量值在突出主要成矿元素的同时, 显示这些伴生元素可能对找矿具有一定的指示作用。III 类元素的投影向量值较小, 显示与矿化的关系不是太明显。

另外, 将投影分类的结果与聚类分析的结果(图 5)进行对比, 两种方法所得到的元素分类结果差异不大, 具有相似的元素组合, 不同的是投影寻踪模型的结果对研究区的主要成矿元素具有一定的突出作

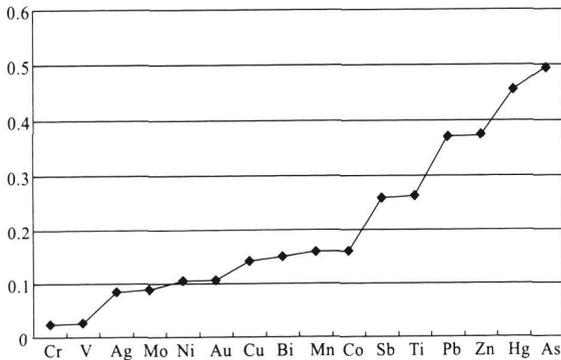


图 4 最佳投影方向向量值分布图

Fig. 4 Distribution of the best projection direction vector

用。

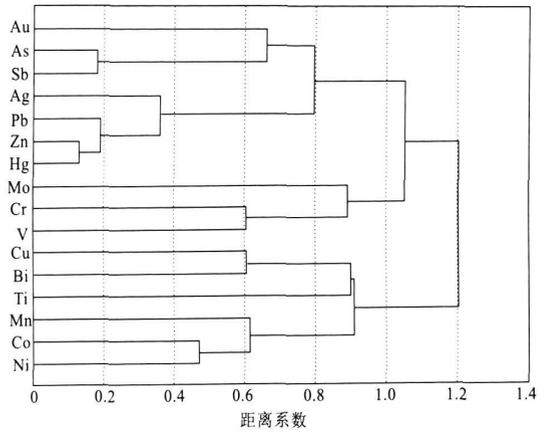


图 5 凤太地区元素聚类分析图

Fig. 5 Diagram showing the element cluster analysis of Fengtai region

4 结论

将 PPC 模型的投影结果与常规的多元处理方法在异常的识别和元素组合方面进行初步对比, 结果表明, 该方法具有如下优势:

(1) 该模型不仅能识别出异常强度较大的区域, 而且能识别异常强度较小的区域, 从而便于圈定异常区; 并且在同样的条件下, 相对于常规的多元统计方法, 所得到的结果更稳定。

(2) 应用该模型不仅能够将元素之间的相关性进行区分, 还能够突出研究区的优势矿种, 从而为确定进一步找矿的方向提供依据。

(3) 该方法有着较为确定的取值标准, 且不需要对数据进行预处理, 减少了化探数据处理过程中的主观因素影响, 使处理结果更加具有确定性和合理性。

当然, 由于该模型才开始应用于金属矿产勘查领域, 在实际运用过程中尚存在一些需要进一步探讨和完善的问题, 如在交叉率、遗传率等参数的选择上怎样减小人为主观性的权重, 如何更有效地构建应用于金属矿产领域的投影指标函数, 如何对投影方向值所代表元素的地质意义进行更深入的探讨, 如何将该模型与分形技术更好地结合等等。

参考文献:

- [1] 付强, 赵小勇. 投影寻踪模型原理及其应用[M]. 北京: 科学出版社, 2006: 21-49.
- [2] 肖凡. 基于 RAGA 的 PPC 模型在个旧化探资料处理中的应用研究[D]. 武汉: 中国地质大学, 2008.
- [3] 金菊良, 杨晓华, 丁晶. 基于实数编码的加速遗传算法[J]. 四川大学学报(工程科学版), 2000, 32(4): 1-5.
- [4] Friedman J H, Turkey J W. A projection pursuit algorithm for exploratory data analysis[J]. IEEE Trans. On Computer, 1974, 23(9): 881-890.
- [5] 王顺久. 水资源开发利用综合研究[D]. 成都: 四川大学, 2003.
- [6] 罗玮, 肖健华. 基于核的投影寻踪方法及其在模式分类中的应用[J]. 五邑大学学报, 2003, 17(3): 6-8.
- [7] 李竞生, 姚磊华. 含水层参数识别方法[M]. 北京: 地震出版社, 2003: 66-67.
- [8] 张欣莉, 丁晶, 王顺久. 投影寻踪在岩溶水水质分类中的应用[J]. 工程勘察, 2001(5): 26-27.
- [9] 金菊良, 张欣莉, 丁晶. 评估洪水灾情等级的投影寻踪模型[J]. 系统工程理论与实践, 2002(2): 140-144.
- [10] 王满祥. 陕西省凤县八方山多金属矿床中段地质勘探总结报告[R]. 西安: 西北有色金属地质勘查局, 1989: 1-170.
- [11] 向婷婷, 李春兰, 刘雁. 凤太矿集区金铅锌矿地球化学特征及找矿方向[J]. 物探与化探, 2004, 28(3): 245.
- [12] 于晓琴. 陕西凤太矿集区金、多金属矿产分布特征及成矿作用浅析[J]. 陕西地质, 1997, 25(1): 13-16.
- [13] 王瑞廷, 王涛, 高章鉴, 等. 凤- 太多金属矿集区主要金属矿床成矿系列与找矿方向[J]. 西北地质, 2007, 40(2): 80.
- [14] 隗合明. 秦岭凤太矿田层控铅锌矿床的控矿条件及勘查标志[J]. 陕西地质, 1990, 8(1): 69-75.

Application of RAGA-Based PPC model to analysis of geochemical anomaly

—A case of Fengtai mineral cluster region

LIAO Shi-li, ZHANG Jun, ZHANG Li-ya, YAO Tao

(Institute of Resources, China University of Geosciences, Wuhan 430074, China)

Abstract: To distinguish the geochemical background and geochemical anomaly is the key of geochemical data processing and information extraction has a direct impact on the results of ore prediction. This paper introduces the application of RAGA-based PPC model to procession of 1 : 5 0000 stream sediment data from Fengtai mineral cluster region and probes into use of projection pursuit cluster in recognition of geochemical anomaly and the symbolical elements combination, as well as the geological significance represented by the value of the projection direction. Results show that, PPC model for geochemical anomaly recognition is better than the traditional method and provides a new thinking and approaches for multi-dimensional processing of geochemical data.

Key Words: PPC Models; extraction of geochemical anomalies; Fengtai mineral cluster region; Shaanxi province