

基于粗集约简的群智能算法的储层识别

袁可红^{1,2}, 李艳晓¹, 诸克军²

(1. 洛阳理工学院 数理部, 河南 洛阳 471003; 2. 中国地质大学 管理学院, 湖北 武汉 410074)

摘要: 提出了一种基于粗集约简的粒子群储层识别方法, 即应用粗糙集进行属性约简, 应用粒子群 (PSO) 聚类算法对约简和正规化后的数据进行处理。实验表明, 约简后的 PSO 聚类较约简前在识别率上有明显的提高。

关键词: 粒子群算法; 属性约简; 粗糙集; 聚类

中图分类号: TP18

文献标识码: A

文章编号: 1673-9833(2008)05-0046-03

Reservoir Identification of the Swam Intelligence Algorithms Based on Rough Set Reduction

Yuan Kehong, Li Yanxiao, Zhu Kejun

(1. Department of Mathematic and Physics, Luoyang Institute of Science and Technology, Luoyang Henan 471003, China;
2. School of Management, China University of Geosciences, Wuhan 430074, China)

Abstract: The unconventional reservoir identification is presented based on rough set attribute reduction and particle swarm optimization(PSO), which means utilizing the rough set attribute reduction approach to reduce data space and using PSO clustering algorithm to deal with processing normalized data. Experiment shows that identification rate of the unconventional reservoir with reduced attributes is much higher than all feature attributes in PSO clustering algorithm.

Key words: particle swarm optimization; attribution reduction; rough set; clustering

0 引言

石油勘探的全过程, 实际上是通过多种手段(地质、地震和测井)获取数据——从这些数据中提取信息——将信息转化成对石油储层的认识的动态过程。因此, 如何从油田获得的数据提取出信息, 最后根据提取的信息实现对地下储层的认识, 是一个科学决策过程, 储层识别是该决策过程的重要环节。储层识别问题是多个属性下的聚类问题, 不同的学科, 用于识别的方法也不同, 属性的个数和识别算法是影响识别效果的两大因素。本文将粗糙集属性约简(Rough Set Attribute Reduction)用于储层属性的特征提取, 将粒子群优化算法(Particle Swarm Optimization, PSO)作为

识别工具, 二者的有机结合, 提高了储层类型的识别率。这样, 既节省了储层识别过程中用于属性探测的成本, 又得到了置信度高的储层类型识别结果。

1 粒子群优化算法与聚类模型

1.1 粒子群优化算法

粒子群优化算法是美国的 Kennedy 和 Eberhart 在 1995 年提出的。若将粒子搜索到的最好解记作 P_{id} , 粒子群经历的最好位置记作 P_{gd} , 粒子的速度记作 V_{id} 。在文献[1]中, 该算法通过公式(1)和(2)的设置来完成:

$$V'_{id} = \omega V_{id} - \eta_1 \text{rand}() (P_{id} - X_{id}) + \eta_2 \text{rand}() (P_{gd} - X_{id}), \quad (1)$$

式(1)中: V_{id} 表示第 i 个粒子在第 d 维上的速度; ω

收稿日期: 2008-07-18

基金项目: 国家自然科学基金资助项目(70573101)

作者简介: 袁可红(1977-), 男, 河南睢县人, 洛阳理工学院助教, 硕士, 主要研究方向为信息管理与智能算法应用;
李艳晓(1982-), 女, 河南禹州人, 洛阳理工学院助教, 硕士, 主要从事小波分析与分形几何研究。

为惯性权重; η_1, η_2 为调节 P_{id} 和 P_{gd} 相对重要性的参数; $rand$ 为随机数生成函数。粒子移动的下一位置:

$$X'_{id} = X_{id} + V'_{id}, \tag{2}$$

粒子的移动方向由其原有的速度 V_{id} 、经历的最佳距离 ($P_{id} - X_{id}$) 及群体经历的最佳距离 ($P_{gd} - X_{id}$) 决定, 其权重分别为 ω, η_1, η_2 。

1.2 聚类问题的数学模型

聚类问题, 就是将给定的数据集,

$X = \{x_1, x_2, \dots, x_N\}, x_i \in R^d (i=1, 2, \dots, N)$ 分成 k 个类^[2]。这里的 N 是样本的个数, d 是样本的维数。一般的, 通过设置有效性函数和样本的不断更替, 使得模式样本到聚类中心的距离之和最小。其数学模型可概括为:

$$F = \min \sum_{j=1}^k \sum_{x_i \in S_j} \|x_i - m_j\|^2, \tag{3}$$

其中: k 为聚类数目; m_j 为 j 类样本的中心; S_j 为第 j 类样本 x_i 构成的集合; $\|x_i - m_j\|^2$ 为类内样本 x_i 与中心 m_j 的偏差平方和。

2 基于粗糙集的属性约简

属性约简就是从条件属性集合中发现部分必要的属性, 使得根据这部分条件属性形成的相对于决策属性类和所有条件属性形成的相对于决策属性的分类一致, 所有条件属性相对于决策属性 D 有相同的分类能力。在文献[3]中, 属性约简主要基于以下两个定义:

定义 1 设 U 为一个论域, P 和 Q 为定义在 U 上的两个等价关系簇。 Q 的 P 正域记为 $POS_P(Q)$, 即 $POS_P(Q) = \bigcup_{X \in U/Q} P(X)$ 。进一步, $POS_P(Q) = POS_{\{P, \dots, P\}}(Q)$, 则称 r 为 P 中相对于 Q 可省略的。若 P 中相对于 Q 不可省略的, 则称 P 相对于 Q 独立。

定义 2 设 U 为一个论域, P 和 Q 为定义在 U 上的两个等价关系簇, 若 P 的 Q 独立子集 $S \subset P$, 满足 $POS_S(Q) = POS_P(Q)$, 则称 S 为 P 的 Q 约简, 记为 $RED_Q(P)$ 。 P 的所有 Q 约简的交称为 P 的 Q 核, 记为 $CORE_Q(P)$ 。

3 基于粗集约简的粒子群聚类系统

3.1 基于粗集属性约简的粒子群聚类步骤

1) 选取训练样本, 将连续属性离散化, 对决策表进行属性约简, 得到最小条件属性集;

2) 根据最小属性集, 从原始数据中取出相应的数据, 作为测试样本;

3) 将测试样本随机分为 K 类, 求各类的均值并将其作为类中心^[4], 按顺序将各类中心排列成矩阵并将之作为一个粒子, 反复进行 m 次, 完成初始粒子群位置的设置, 由最小条件属性的数值范围设定粒子速度

和位置的界限 $V_{max}, V_{min}, X_{max}, X_{min}$;

4) 由式 (3), 更新粒子经历的最好位置 P_{id} , 并更新群体经历的最好位置 P_{gd} ;

5) 由式 (1), 更新粒子速度, 并限制在 $[V_{min}, V_{max}]$ 之间;

6) 由式 (2), 更新粒子位置, 并限制在 $[X_{min}, X_{max}]$ 之间;

7) 将上述步骤进行 n 次迭代, 不断调整 P_{id} 和 P_{gd} , 直至目标最优。

3.2 流程图

基于粗集约简的粒子群聚类算法流程见图 1。

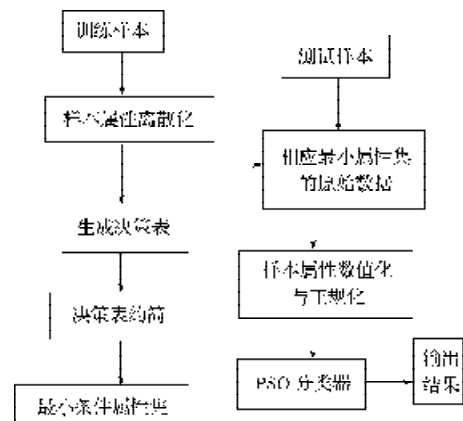


图 1 基于粗集约简的粒子群聚类算法流程图

Fig. 1 The program of PSO clustering based on rough set attribution reduction

4 实验过程及结果

数据来源于江汉油田的 oilsk81, oilsk85, 部分数据见表 1 和表 2。

表 1 oilsk81 井测井解释成果表

Tab. 1 Result of well log interpretation from oilsk81

层号	声波 /($\mu s \cdot m^{-1}$)	中子 /%	深测向 电阻率 / $\Omega \cdot m$	孔隙度 /%	含油 饱和度 /%	渗透率 /($m^2 \cdot \mu m^2$)	储层 类型
1	195	7.5	13.0	6.0	0	0	干层
2	225	10.0	7.3	11.0	0	0	水层
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
31	213	9.5	12.0	9.0	61	2	油层

表 2 Oilsk85 井测井解释成果表

Tab. 2 Result of well log interpretation from oilsk85

层号	声波 /($\mu s \cdot m^{-1}$)	中子 /%	深测向 电阻率 / $\Omega \cdot m$	孔隙度 /%	含油 饱和度 /%	渗透率 /($m^2 \cdot \mu m^2$)	储层 类型
1	225	15.1	10.5	10.7	0	3.2	水层
2	224	13.4	16	10.5	0	2.9	水层
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
34	185	10.3	100	1.6	0	0	干层

注: 表中的层号表示各口井所处实际位置的序号, 选用声波、中子、渗透率等 6 个测井属性, 储层类型为实际测井过程中做出的结论。

实验数据处理 对声波、中子、电阻率和空隙度

等属性值由 $x_{ij} = \frac{x_{ij} - x_{ij}^{\min}}{x_{ij}^{\max} - x_{ij}^{\min}}$ 正规化, 对数值差别较大的含油饱和度和渗透率属性值, 采用统计平均方法^[5],

即利用公式 $x_{ij} = \frac{x_{ij} - m_j}{x_{ij}^{\max} - m_j} (1 - M) + M$ 正规化, 其中 $M = 0.5$ 。

实验粒子群算法中的参数设置

粒子个数 $P_{\text{sonum}} = 50$, 最大迭代次数 $T = 100$, 权重 ω 采用由大到小的线性递减设置^[6], 即 $\omega = \omega^{\max} - (\omega^{\max} - \omega^{\min}) * t / T$, $\omega^{\max} = 0.9$, $\omega^{\min} = 0.4$, 调节系数 $\eta_1 = \eta_2 = 1.49$, 粒子位置 and 速度在第 d 个属性的最大和最小值设为 $x_{d \max}$, $x_{d \min}$, $v_{d \max}$, $v_{d \min}$, 其关系由文献^[6]给出的

$$v_{d \max} = k \frac{(x_{d \max} - x_{d \min})}{2}, v_{d \min} = -v_{d \max}, 0.1 \leq k \leq 1 \text{ 确定, 这}$$

里 $k = 0.1$ 。适应度函数: $J_p = \sum_{j=1}^k \sum_{i \in S_j} \|x_i - m_j\|^2$, 其中 $\|\cdot\|$ 为欧式距离。

类中心的设置和存放

每个粒子是一个 4×6 的矩阵, 自上而下表示干层、水层、差油层、油层的类中心, 其中 4 为类数, 6 为属性的最大数目。

从表 1 和表 2 中, 随机选取 20% 作为训练样本, 借助粗糙集在软件 Rossetta 下完成对训练样本的约简, 即用 Semi-naïve algorithms 算法离散化原始数据, 用 Genetic algorithm 算法对离散化的数据进行属性约简, 得出各油井的最小属性集为声波、含油饱和度。

对最小属性集和无约简的属性组合分别经 PSO 分类器进行识别, 进行 10 次实验, 计算各自的最低、平均、最高识别率, 实验结果统计见表 3。

表 3 测试结果

Tab. 3 Result of experiment %

识别率	oilk81			oilk85		
	最低	平均	最高	最低	平均	最高
约简后	96.77	96.77	96.77	97.06	97.06	97.06
约简前	93.55	95.16	96.77	59	80.1	88

5 实验结果分析

从选用属性个数上分析, 约简后选用的最小条件属性个数为 2, 约简前选用属性的属性为 6, 可得出如下结论: 在不改变识别率的前提下, 通过属性约简可以大大节省为获取测井属性状态而需要投入的成本。

从识别准确率上分析, 无论是表 1 还是表 2 反映的测井数据, 约简后较约简前在识别率上都有提高, 这是储层识别工作的一个主要目的。

从识别结果的稳定上分析, 约简后较约简前更加稳定, 一是因为粒子群搜索算法能够通过大量的粒子完成对最相似样本的匹配, 使适应度值达到最小; 二是由于通过粗糙集的属性约简, 消除了冗余属性对识别结果的影响, 保留了对识别结果影响大的属性, 克服了约简前将属性对识别影响等同看待的不足之处。算法选取的是小样本, 而样本个数过多和属性个数过多都将降低识别准确率, 寻求大样本和过多属性下的识别算法尚值得研究。

综上所述, 本文采用的基于粗糙集的粒子群储层识别是一种高效且稳定的识别技术。

参考文献:

- [1] 刘靖明, 韩丽川, 侯立文. 基于粒子群的 K 均值聚类算法[J]. 系统工程理论与实践, 2005, 6(6): 54-58.
- [2] 范明, 戴冠中. 基于有效性函数的聚类算法[J]. 计算机科学, 2007, 34(5): 197-207.
- [3] 张文修, 吴伟志. 粗糙集理论与方法[M]. 北京: 科学出版社, 2003.
- [4] 高尚, 杨静宇. 一种新的基于粒子群算法的聚类方法[J]. 南京航空航天大学学报, 2006, 38(7): 62-65.
- [5] 岳超源. 决策理论与方法[M]. 北京: 科学出版社, 2004.
- [6] 杨维, 李歧强. 粒子群优化算法综述[J]. 中国工程科学, 2004, 6(5): 87-94.

(责任编辑: 廖友媛)