

# 粗糙集方法在瓦斯突出预测中的应用

李云明<sup>1</sup> 张晓明<sup>2</sup>

<sup>1</sup>(南京化工职业技术学院成人教育学院 江苏 南京 210048)

<sup>2</sup>(中国科学院合肥智能机械研究所 安徽 合肥 230031)

**摘要** 煤矿瓦斯突出是一种多发的地质灾害,其影响因素众多,但是突出机理尚未研究清楚。将粗糙集理论和技术引入煤矿瓦斯突出区域预测,对影响瓦斯突出的关键因素进行约简,抽取与瓦斯突出相关的决策规则,建立了煤矿瓦斯突出区域预测模型,并通过数据试验和对比,表明粗糙集方法预测瓦斯突出具有较高的预测准确率,同时方法也为瓦斯突出预测提供了新的研究思路。

**关键词** 粗糙集 瓦斯突出 预测 规则

## APPLYING ROUGH SET THEORY IN FORECAST OF GAS OUTBURSTS

Li Yunming<sup>1</sup> Zhang Xiaoming<sup>2</sup>

<sup>1</sup> (Adult Education College, Nanjing College of Chemical Technology, Nanjing 210048, Jiangsu, China)

<sup>2</sup> (Institute of Intelligent Machines, Chinese Academy of Sciences, Hefei 230031, Anhui, China)

**Abstract** Gas outburst in coal mines is a frequently occurred geological disaster and is affected by many factors, however the mechanism of gas outburst is still unclear and is being studied. In this paper we introduce rough set theory and technology to forecasting gas outbursts regions in coal mines. By the reduction of crucial factors impacting gas outbursts and the extraction of decision-making rules pertaining to gas outbursts, a forecasting model of gas outbursts regions in coal mines is built. Through data experiments and contrasts, the rough set approach is demonstrated to have quite high precision in forecasting the gas outbursts. Meanwhile, this approach also provides new research conception to forecasting gas outbursts.

**Keywords** Rough set Gas outburst Forecast Rule

## 0 引言

瓦斯突出是发生在煤矿井下的一种复杂的地质灾害。国内煤矿的瓦斯事故中,许多情况都是瓦斯突出引起的<sup>[1]</sup>。瓦斯突出影响因素众多,但是突出机理尚未彻底研究清楚。随着人工智能技术的发展,灰色预测理论<sup>[2]</sup>、人工神经网络<sup>[3]</sup>、支持向量机<sup>[4]</sup>等理论和技术进入煤矿安全领域,在煤与瓦斯突出预测等方面取得了一定的成绩,但仍然存在进一步改进和完善的地方,比如推理技术单调、学习能力弱等缺点,难以处理矿难事故中普遍存在的大量半结构化或非结构化问题,因此效果并不理想。

粗糙集理论是 50 年代初由波兰数学家 Pawlak 提出的一种处理模糊和不精确性问题的数学工具<sup>[5]</sup>。粗糙集理论不需要预先给定某些特征或属性的数量描述,而是直接从给定问题的描述集合出发,通过不可分辨关系和不可分辨类确定问题的近似域,从而找出问题中的内在规律。近年来,粗糙集理论在模式识别、机器学习、决策分析和决策支持、知识获取等领域逐渐显示了它的重要性,其理论和应用研究都已经取得了较大的发展。因此将粗糙集理论和技术引入煤矿瓦斯突出预测,用来发现隐蔽的事故发​​生机理、诱因和演变规律,为我国煤矿安全生产提供预测预警的科学方法和技术支撑。

## 1 粗糙集理论要点概述<sup>[6]</sup>

### 1.1 知识表达系统

粗糙集理论的要点是将知识与分类联系在一起。知识被认为是一种对对象进行分类的能力,对象用其属性集合表示,分类用来产生概念,概念构成知识的模块。知识是由对象论域的分类模块组成的,它提供关于现实的明显的事实,同时也具有明显事实推导出模糊事实的推理能力。一个知识表达系统定义为:

$$S = \langle U, C, D, V, f \rangle$$

式中,  $U$  是对象的集合,  $C \cup D$  是属性集合(等价关系集合), 子集  $C$  和  $D$  分别称为条件属性和结果属性,  $V = \bigcup_{a \in A} V_a$  是属性值域的集合,  $V_a$  表示了属性  $a \in A$  的范围,  $f: U \times A \rightarrow V$  是一个信息函数,它指定  $U$  中的每一对象  $x$  的属性值。这种定义方式使对象的知识可以方便地以数据表格形式描述,这种数据表称为知识表达系统。

### 1.2 知识约简

在一个决策系统中,各个条件属性之间往往存在着某些程度上的依赖或关联,约简可以理解为在不丢失信息的前提下,可

收稿日期: 2009-08-24, 李云明, 副教授, 主研领域: 智能信息处理, 软件工程。

以最简单地表示决策系统的决策属性对条件属性集合的依赖性  
或关联度。

对于一个给定的决策系统  $S$  条件属性集合  $C$  的约简是  $C$   
的一个非空子集  $C'$ , 它满足: (1)  $ind(C', \{d\}) = ind(C, \{d\})$ ;  
(2) 不存在  $C'' \subset C'$ , 使  $ind(C'', \{d\}) = ind(C, \{d\})$ 。

$C$  的所有约简的集合记作  $RED_D(C)$ 。  $C$  的所有约简的交  
集叫作核 (CORE), 记作  $CORE_D(C)$ ,  $CORE_D(C) = \bigcap RED_D(C)$ 。

根据约简和核的概念, 粗糙集理论提供了分析多余属性的  
方法, 对知识的处理是通过对决策表中的属性值的处理实现的。  
具体步骤如下: (1) 删除重复的实例; (2) 删除多余的属性; (3)  
删除每个实例多余的属性值; (4) 求出最小约简; (5) 根据最小  
约简, 求出逻辑规则。

## 2 粗糙集预测模型

煤与瓦斯突出事故是不希望发生的一种事件, 但在煤矿采  
掘生产过程中却隐藏着许多可能导致其发生的危险<sup>[7]</sup>。 要保  
证生产的安全 (瓦斯突出事故不发生), 系统了解所有可能导致  
突出发生的主要危险特征和主导因素是必要且关键的; 它对促  
进安全生产、避免今后不再重复类似事故的发生也是极其重要  
的。 同时影响和控制瓦斯突出的因素或条件多而复杂, 其中不  
仅包含定量变量, 而且具有大量的定性变量, 不同的因素、条  
件或组合与瓦斯突出的关系基本都呈非线性变化。 所以采用通  
常的数值分析方法进行计算一般难以得到比较理想的结果, 而且  
构建相关的数学模型也是极其困难的。 粗糙集方法不仅可以很  
好地提取主要的瓦斯突出的主要属性, 而且还可以通过控制训  
练样本数量、主要属性的多少、数据离散化程度等手段提高拟合  
精度, 从而得到比较理想的预测模型。

### 2.1 模型的建立<sup>[8]</sup>

模型的建立包括以下三步:

#### (1) 属性约简

属性约简的目的就是在保持预测能力不变的情况下, 求得  
最佳约简, 本文中条件属性为影响和控制瓦斯突出的 8 个属性,  
其中: 瓦斯压力为  $P$ , 属性编号为 0 煤力学强度为  $f$ , 属性编号为 1;  
煤体破碎性综合特征系数为  $K_{\text{破碎}}$ , 属性编号为 2 煤的透气性  
系数为  $\lambda$ , 属性编号为 3 煤层分叉合并综合特征系数为  $K_f$ , 属  
性编号为 4 煤厚及煤厚变化综合特征系数为  $K_{\text{煤厚}}$ , 属性编号为 5  
断层复杂程度系数为  $K_{\text{断层}}$ , 属性编号为 6 层间滑动综合特  
征系数为  $K_{\text{层滑}}$ , 属性编号为 7。

利用差别矩阵进行属性约简, 算法如下:

输入: 瓦斯突出知识表达系统, 包括条件属性集  $A$ , 决策属  
性集  $D$ ;

输出: 瓦斯突出知识表达系统的约简集  $Reduct$ ;

步骤:

#### Step 1 对连续属性值离散化;

以训练样本瓦斯压力  $P$  为例, 根据训练数据情况, 对数值在  
[0.1, 2.167) 区间的, 离散化为 1, 对数值在 [1.2167, 2.4334) 区  
间的, 离散化为 2 对数值大于 2.4334 的, 离散化为 3。 前五条  
数据的离散化结果由表 1 所示。

表 1 瓦斯压力  $P$  前五条数据离散化结果

序号	1	2	3	4	5
原属性值	1.4000	2.1600	1.4000	2.9000	3.6500
离散化后的属性值	1	2	1	3	3

Step 2 建立差别矩阵  $M$ , 寻找差别矩阵的核  $Core$ , 计算系  
统的核  $Core$  各属性的出现频率  $frequency(ai)$ ;

本文实验中, 核为煤体破碎性综合特征系数和层间滑动综  
合特征系数。

Step 3  $Reduct = Core, F = A - Core$

Step 4  $M = M - m \downarrow$  其中  $m$  是指差别矩阵中所有包含最  
佳约简中属性的元素, 重新计算各属性的出现频率  $frequency$   
( $ai$ );

Step 5 剩余属性集  $F$  中挑选出现频率最高的属性  $f$  加入  
到最佳约简中去,  $Reduct = Reduct + \{f\}, F = F - \{f\}$ ;

本文实验中, 剩余属性集中出现频率最高的两个属性为煤  
的透气性系数和煤厚及煤厚变化综合特征系数。

Step 6 回到 Step4 循环, 直到属性个数满足要求;

Step 7 输出约简集  $Reduct$ ;

#### (2) 决策规则约简

决策规则的约简是利用决策逻辑分别消去决策算法中每一  
个决策规则的不必要条件, 它不是整体上的简化属性, 而是针对  
每一个决策规则, 去掉该规则的冗余属性值, 进一步简化决策  
算法。

本文在决策规则的约简过程中, 首先去除重复的决策规则,  
然后对每一条决策规则, 从第一个属性开始, 依次去掉每一个属  
性, 并且检查该属性去掉后决策表是否协调, 如果决策表协调则  
该属性可以去掉, 否则该属性予以保留, 重复此过程, 可以达到  
对每条决策规则进行约简的目的。 经过属性约简后, 得到一个  
约简的决策表, 其中决策表的每一行都是一条决策规则。 3.2  
节中列出了本文实验生成的决策规则。

#### (3) 模型生成

由约简生成的所有瓦斯突出决策规则构建预测模型。 用于  
类别预测时, 将先计算被测对象与各决策规则的匹配度, 即计算  
两者中取值相同的属性的个数与规则内条件的属性总数的比  
值。 然后决策类别相同的规则的匹配度相加, 然后除以类别  
规则的个数, 由值最大的那类瓦斯突出规则判定样本类别。

## 2.2 模型结构图

根据 2.1 节所述, 抽取关键步骤, 建立如图 1 所示的基于粗  
糙集的预测模型结构图<sup>[9]</sup>:

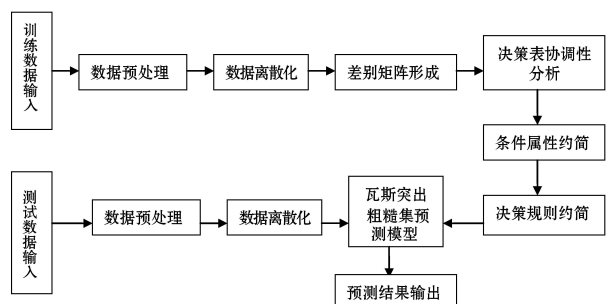


图 1 粗糙集预测模型结构图

## 3 实验与分析

### 3.1 实验数据

以文献 [10] 中所公开的淮北矿业集团公司芦岭矿的相关

资料为基础, 确定实验数据为该矿影响和控制瓦斯突出的 8 个主要因素。

### 3 2 实验结果

实验系统在 Windows 环境下, 利用 JBuilder2007 实现。

采用局部离散化方法, 将属性值离散化为高、中、低三个域, 分别用 3 2 1 表示, 具体参见 2 1 节 step1 突出危险性结果定义为: 不突出用 0 表示, 突出威胁用 1 表示, 一般突出用 2 表示, 严重突出用 3 表示。

经过关键属性提取, 在影响和控制瓦斯的 8 个属性中选择 4 个关键属性如下:

$K_{破碎}$  (编号 2): 为煤体破碎性综合特征系数;

$K_{层滑}$  (编号 7): 为层间滑动综合系数;

$\lambda$  (编号 3): 为煤的透气性系数;

$K_{煤厚}$  (编号 5): 为煤厚及煤厚变化综合特征系数。

经过规则约简, 最终规则由表 2 所示。

表 2 规则表

规则	$K_{破碎}$	$K_{层滑}$	$\lambda$	$K_{煤厚}$	突出危险性
1	0~1.07	0~15.24	0.22~0.43	0~1.31	不突出
2	0~0.53	0~15.24	0.43~0.65	0~1.31	不突出
3	0~0.53	0~15.24	0.22~0.43	1.31~2.63	不突出
4	0.53~1.07	0~15.24	0~0.65	1.31~2.63	突出威胁
5	1.07~1.6	0~15.24	0.22~0.65	0~1.31	突出威胁
6	1.07~1.6	0~15.24	0.43~0.65	1.31~2.63	突出威胁
7	1.07~1.6	0~15.24	0.22~0.43	1.31~2.63	一般突出
8	1.07~1.6	15.24~30.53	0.22~0.43	0~1.31	一般突出
9	1.07~1.6	15.24~30.53	0~0.22	1.31~2.63	一般突出
10	1.07~1.6	0~15.24	0.22~0.43	2.63~3.94	严重突出
11	1.07~1.6	15.24~30.53	0~0.22	2.63~3.94	严重突出
12	1.07~1.6	0~15.24	0~0.22	1.31~3.94	严重突出
13	1.07~1.6	30.53~45.77	0.43~0.65	1.31~2.63	严重突出

预测结果 (方框中的数据为预测结果无法验证的数据):

测试数据 1 预测结果为: 突出威胁

测试数据 2 预测结果为: 突出威胁

测试数据 3 预测结果为: 一般突出

测试数据 4 预测结果为: 严重突出

测试数据 5 预测结果为: 不突出

测试数据 6 预测结果为: 突出威胁

测试数据 7 预测结果为: 一般突出

测试数据 8 预测结果为: 突出威胁

测试数据 9 预测结果为: 严重突出

测试数据 10 预测结果为: 不突出

测试数据 11 预测结果为: 不突出

测试数据 12 预测结果为: 不突出

测试数据 13 预测结果为: 不突出

测试数据 14 预测结果为: 不突出

测试数据 15 预测结果为: 不突出

从上面的预测结果可以看出, 实验瓦斯突出的影响因素进行了关键属性提取, 并得出用于瓦斯突出预测的规则, 建立了相

应的决策模型, 预测结果也比较满意。基于粗糙集的瓦斯突出预测模型的总体预测准确率不低于 86%, 即不低于文献 [10] 中的预测准确率。因为文献 [10] 中预测对的数据项, 本方法都预测对了; 数据项 5 和 10 是文献 [10] 中预测错误的项, 由于无法得到其正确的原始结果, 所以无法说明本方法预测的正确与否。另外, 通过与中国矿业大学的专业教授交流得知, 本方法中将数据项 5 和 10 的结果判断为不突出是合理的, 而且采用基于粗糙集的方法, 系统的计算复杂度大大降低了。本方法的实验也存在一些问题, 简单分析如下:

(1) 数据项离散化方法采用局部离散化, 不能很好地对数据进行离散化, 容易出现信息损失;

(2) 采用别人整理的训练数据数量过少, 虽然方便了与已有结果做对比, 但是却不能保证建立完备的预测模型;

今后, 我们将继续依托国家自然科学基金重点项目, 对基于粗糙集的瓦斯突出预测深入研究, 拟开展如下工作:

(1) 数据调研, 与淮南矿务局协商数据采集事宜, 采集大量的第一手数据, 为建立完备的数据模型做好准备;

(2) 对连续数据离散化方法进行研究, 尽可能减少数据处理后的信息丢失;

(3) 整合其它的智能计算方法, 提高粗糙集方法的约简效率和预测准确率。

### 4 结束语

本文研究了粗糙集理论在瓦斯突出预测中的应用, 粗糙集理论不需要任何先验知识, 即可从大量的样本数据中挖掘出预测规则, 并且通过实验表明该方法具有较高的预测准确率, 为瓦斯突出预测提供了新的研究思路。今后, 该研究工作将继续开展, 早日解决实验中遇到的问题。

### 参 考 文 献

[1] 臧大进, 王耀才. 煤与瓦斯突出预测研究 [J]. 计算机工程与设计, 2007, 28(12): 2983-2987.

[2] 杨凌霄, 沈鹰, 侯国栋, 等. 基于支持向量机的煤与瓦斯突出预测研究 [J]. 河南理工大学学报, 2006, 25(5): 348-350.

[3] 李成武. 煤与瓦斯突出危险性灰色分类方法实验研究 [J]. 煤矿安全, 2007(06): 1-4.

[4] 高雷卓. 煤与瓦斯突出灾害的预测—基于遗传和 BP 算法 [J]. 辽宁工程技术大学学报: 自然科学版, 2002(4): 408-410.

[5] Pawlak Z. Rough sets [J]. International Journal of Information and Computer Science, 1982, 11: 341-356.

[6] 刘清. 粗糙集及粗糙推理 [M]. 北京: 科学出版社, 2001.

[7] Aguado Maria B, Diaz Gonzalez C, Nicieza Control and prevention of gas outbursts in coalmines Riosa-Olloniego coalfield Spain [J]. International Journal of Coal Geology, 2007, 69(4): 253-266.

[8] 杨善林, 倪志伟. 机器学习与智能决策支持系统 [M]. 北京: 科学出版社, 2004.

[9] 曾黄麟. 粗糙理论及其应用 [M]. 重庆: 重庆大学出版社, 1996.

[10] 张瑞林. 现代信息技术在煤与瓦斯突出区域预测中的应用 [D]. 重庆: 重庆大学, 2004.