

# 煤矿瓦斯预测知识获取模型的应用研究

孙林嘉<sup>1</sup>, 李 茹<sup>1,2</sup>, 屈元子<sup>1</sup>

(1. 山西大学计算机与信息技术学院, 太原 030006; 2. 山西大学计算机智能与中文信息处理省部共建教育部重点实验室, 太原 030006)

**摘要:** 将粗糙集与神经网络结合, 提出由样本更新、粗糙集预处理、神经网络训练、规则提取 4 个模块组成的煤矿瓦斯预测知识获取模型, 将其应用于实时数据进行实验, 结果表明, 该模型实时性好、可靠性及精度高, 可以较好地解决煤矿瓦斯预测知识获取困难的问题, 为煤矿瓦斯预测专家系统知识库的建立奠定基础。

**关键词:** 知识获取; 煤矿瓦斯预测; 粗糙集; 神经网络

## Application Research of Knowledge Acquisition Model for Colliery Gas Forecast

SUN Lin-jia<sup>1</sup>, LI Ru<sup>1,2</sup>, QU Yuan-zi<sup>1</sup>

(1. School of Computer & Information Technology, Shanxi University, Taiyuan 030006; 2. Computer Intelligent and Chinese Information Processing of the Ministry Education Key Laboratory Built Together by Province and Department, Shanxi University, Taiyuan 030006)

**【Abstract】** By combining rough sets and neural networks, this paper develops a knowledge acquisition model for colliery gas forecast, which consists of four modules: sample refreshment, rough sets preprocessing, neural network training and rules extraction. It is applied in the real-time data, whose results show that it solves the problem of the knowledge acquisition difficulty of colliery gas forecast, and has good real-time characteristic, high reliability and perfect precision. The model provides foundation for establishing knowledge database of colliery gas forecast.

**【Key words】** knowledge acquisition; colliery gas forecast; rough sets; neural networks

2004年,山西省实施了煤矿瓦斯监测监控信息网络工程,在一定程度上防范了煤矿瓦斯事故的发生,但是大量的瓦斯数据无论是存储还是应用都存在许多问题,如何浓缩数据规律、提取煤矿瓦斯预测知识是本文研究的主要内容。

### 1 煤矿瓦斯预测知识获取模型

#### 1.1 模型的结构及功能

本文利用粗糙集与神经网络的知识获取能力,结合各自的优势<sup>[1]</sup>,构建了基于粗糙集与神经网络的煤矿瓦斯预测知识获取模型。模型的总体结构如图1所示。

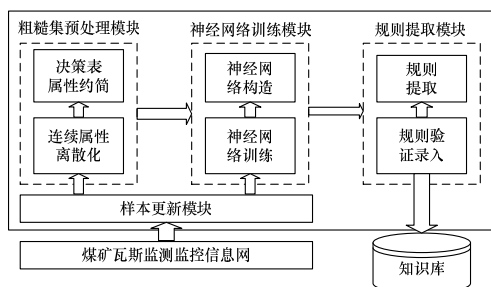


图1 煤矿瓦斯预测知识获取模型

各模块的主要功能如下: (1)样本更新模块。生成训练样本集,对样本进行增加、删除和修改。(2)粗糙集预处理模块。采用粗糙集对样本数据进行预处理,包括连续属性离散化和属性约简。(3)神经网络训练模块。确定神经网络的层数和各层的节点数,得到以权值和阈值形式存在的神经网络知识。(4)规则提取模块。从神经网络中提取易于理解的规则,完成从隐式知识到显式知识的转换,并对规则进行验证,最后录入到知识库中。

#### 1.2 模型的知识获取流程

知识获取流程如图2所示。

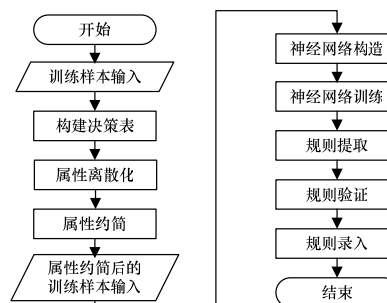


图2 煤矿瓦斯预测知识获取流程

具体步骤如下: (1)从监测点获取原始信息组成训练样本集。(2)构建符合粗糙集处理的信息决策表。(3)对连续型数据进行离散处理。(4)对决策表进行属性约简。(5)根据属性约简结果重新构建训练样本集,即删除已经被约简掉的属性。(6)根据属性约简结果确定神经网络的输入输出层节点数,在训练中动态修正隐含层节点数。(7)用属性约简后的训练样本集对已经确定的神经网络进行训练。(8)从训练好的网络中提取规则。(9)检验规则并录入知识库。

**基金项目:** 山西省高校拔尖人才基金资助项目(0710003); 太原市科技明星专项基金资助项目(08121005); 山西大学大学生科研训练基金资助项目(0803012)

**作者简介:** 孙林嘉(1983—),男,硕士研究生,主研方向:智能信息处理;李 茹,教授;屈元子,硕士研究生

**收稿日期:** 2008-10-30 **E-mail:** xqysl-1998@163.com

## 2 煤矿瓦斯预测知识获取模型算法

### 2.1 基于信息熵的连续属性离散化算法

运用粗糙集理论处理决策表时，要求决策表中的值用离散数据表达。文献[2]从决策表信息熵的角度提出了一种连续属性离散化算法，算法考虑了由连续条件属性和离散条件属性构成的混合知识系统，实验结果表明此算法不仅有效，而且当样本量和属性个数较大时仍有较高的计算效率。本文采用基于信息熵的连续属性离散化算法对煤矿瓦斯数据中的连续条件属性进行离散化。

### 2.2 基于信息熵的属性约简算法

属性约简是粗糙集理论的核心内容之一。属性约简是由属性的重要性确定的，为考虑正域之外的元素对属性重要性的影响以及“脏数据”对信息表中不可分辨关系的干扰，引入信息熵作为属性重要性的判定准则。本文采用基于信息熵的属性约简算法进行条件属性约简，具体算法参照文献[3]。

### 2.3 神经网络构造及训练算法

3层神经网络可以很好地解决一般问题<sup>[4]</sup>。隐含层节点数的选择迄今为止尚未找到好的方法。因此，本文在训练过程中动态确定隐含层节点数，就是以标准BP算法训练网络，定时判断当前收敛值是否达到要求。若达到要求，则进行下一轮测试，否则加入一个节点，直到网络达到收敛要求。

在训练结束时，可能使网络添加了多余的隐含层节点，所以，有必要在误差达到要求后删除多余的节点。本文采用欧拉函数计算各个向量之间的相似度，若2个向量的相似度达到一个阈值，则将它们归为一类，分类后，每类保留其中一个节点而删除其他同类节点。

经过处理后的网络与处理前的网络不一定完全相同，即处理前收敛达到要求的网络经过处理后未必达到要求，需重新训练，直到网络收敛，这时网络结构达到最简。

算法的具体步骤如下：

(1)根据约简后的决策表初始化BP网络。

(2)采用标准BP算法进行训练，训练中动态地加入隐含层节点，直到网络收敛。

(3)对隐含层节点进行聚类，删除多余节点。

(4)采用标准BP算法对修正的网络进行训练。

(5)重复步骤(3)~步骤(5)，直到网络经过步骤(3)、步骤(4)的处理而没有变化为止。

(6)采用标准BP算法对已经确定的神经网络进行训练，直到网络收敛。

### 2.4 规则的提取

训练结束后，需要从稳定的神经网络中提取易于理解的规则。基于结构分解的规则提取方法<sup>[5]</sup>，以隐含层节点和输出层节点为单位，将网络分解为若干单层网络的集合，对每一子网提取规则，最后对规则进行组合以描述整个网络的特性。最简单的一种算法是文献[6]提出的子集法，算法由规则生成和规则重写2个步骤组成。本文在使用该算法时，要求从规则中消除所有的隐含层节点。

## 3 实验及分析

已知山西省霍州东兴煤矿矿井的原始数据记录7529条，随机抽取7000条生成训练样本，其余529条作为测试样本，数据格式如表1所示，其中，瓦斯压力(WSYL)、放散初速度(FSSD)、巷道类型(XDLX)、煤层倾角(MCQJ)、工作方式(GZFS)、坚固系数(JGXS)、开采深度(KCSD)、地质构造(DZGZ)8个参数为条件属性，突出强度(TCQD)为决策属性。

表1 煤矿瓦斯原始数据

BH	WSYL	FSSD	MCQJ	JGXS	KCSD	GZFS	XDLX	DZGZ	TCQD
1	2.906	3.136	40	0.652	469	打钻	斜巷	断层	91
2	3.321	18.321	15	0.135	493	手摘	石门	断层	337
3	4.810	7.972	24	0.447	602	放炮	斜巷	褶皱	398
4	4.456	5.883	31	0.153	602	手摘	斜巷	褶皱	85
5	3.475	7.845	17	0.354	529	手摘	石门	断层	371
6	3.137	19.029	54	0.300	414	放炮	石门	断层	356
7	0.271	8.577	42	0.840	328	打钻	平巷	褶皱	0
8	2.806	18.718	14	0.273	484	手摘	石门	断层	311
9	0.227	8.637	38	0.780	307	打钻	平巷	褶皱	0
10	3.673	11.712	9	0.469	548	手摘	斜巷	断层	409
11	3.703	10.825	9	0.452	542	手摘	斜巷	断层	401
12	3.918	11.167	12	0.507	583	手摘	斜巷	断层	434
13	1.947	15.286	5	0.112	445	放炮	石门	断层	364
14	4.471	4.129	30	0.617	600	手摘	斜巷	褶皱	83
15	1.312	11.262	48	0.911	375	打钻	平巷	褶皱	0
16	3.719	2.367	26	0.479	569	手摘	斜巷	断层	94
17	1.254	12.275	48	0.838	375	打钻	平巷	褶皱	0
18	1.567	13.527	38	0.252	419	放炮	平巷	断层	218
19	1.169	12.404	37	0.969	391	放炮	平巷	断层	42
20	2.758	18.525	53	0.275	502	手摘	石门	断层	337

### 3.1 粗糙集预处理

采用基于信息熵的连续属性离散化算法<sup>[2]</sup>对连续数据进行离散化，离散区间的设置如表2所示。突出强度按惯例以实际抛出煤重量的大小分为4类：无突出(0 t)，小突出(50 t以下)，中突出(50 t~100 t)，大突出(100 t以上)。

表2 离散区间确定表

条件属性	离散区间			
	1	2	3	4
WSYL	(0,1.35)	(1.350,2.235)	(2.235,3.173)	(3.173,6.25)
FSSD	(0,6.82)	(6.82,10.73)	(10.73,15.12)	(15.12,23.5)
MCQJ	(0,28.5)	(28.5,39.5)	(39.5,50.5)	(50.5,68.5)
JGXS	(0,0.215)	(0.215,0.531)	(0.531,0.764)	(0.764,1.0)
KCSD	(0,380.5)	(380.5,470.5)	(470.5,530.5)	(530.5,650.5)

采用信息熵作为属性重要性的判定准则，计算各个属性的重要性，如表3所示。

表3 各条件属性的重要性

	WSYL	FSSD	JGXS	XDLX	KCSD	MCQJ	GZFS	DZGZ
重要性	0.785	0.724	0.653	0.352	0.06	0.379	0.581	0.040
排序	1	2	3	66	7	5	4	8

按照基于信息熵的属性约简算法<sup>[3]</sup>得到样本数据的5个相对约简：{WSYL, FSSD, JGXS, GZFS}，{WSYL, FSSD, JGXS, GZFS, MCQJ}，{WSYL, FSSD, JGXS, GZFS, XDLX}，{WSYL, FSSD, JGXS, XDLX, MCQJ}，{WSYL, FSSD, JGXS, GZFS, MCQJ, XDLX}，由相对约简集得核属性{WSYL, FSSD, JGXS}。

### 3.2 神经网络训练

采用3层网络结构，输入层节点数由粗糙集对样本数据约简后的指标确定为6个节点。隐含层不同节点数的训练误差如表4所示。

表4 不同隐含层节点数训练误差对比

隐含层神经元数	训练次数	误差逼近
6	4 843	0.425 76
7	4 387	0.257 63
8	3 761	0.123 74
9	2 411	0.039 52
10	4 217	0.388 22
11	4 984	0.452 38
12	4 387	0.543 87
13	6 352	0.708 81

由表4可知，隐含层节点数为9个时，网络的误差性能最好，训练次数最少。输出层节点数由输出类别确定为4个。隐含层采用Sigmoid型函数，输出层采用线性函数。7000条

原始数据经离散化后得到 4 281 条数据, 经约简后得到 759 条数据, 学习率为 0.05, 个体误差为 0.01, 经过 1 375 次迭代, 网络达到收敛, 总体误差为 0.014 8, 训练时间为 364 s。

### 3.3 与单独使用BP神经网络的比较

以预处理前的原始数据作为输入, 对单独使用 BP 神经网络进行训练, 网络结构定为 8-13-4, 隐含层仍采用 Sigmoid 型函数, 输出层采用线性函数。2 种模型的比较结果如表 5 所示, 误差曲线如图 3 和图 4 所示。比较图 3、图 4 可以看出, 在精度要求一样的条件下, 采用粗糙集进行数据预处理后的神经网络比未采用粗糙集进行数据预处理后的神经网络能更快速地达到要求, 节省了大量的时间。

表 5 2 种模型实验结果比较

网络模型	总体误差	训练次数	训练精度	训练时间/s
单独使用 BP 网络	0.246 7	6 600	0.01	1 728
本文 BP 网络	0.014 8	1 375	0.01	364

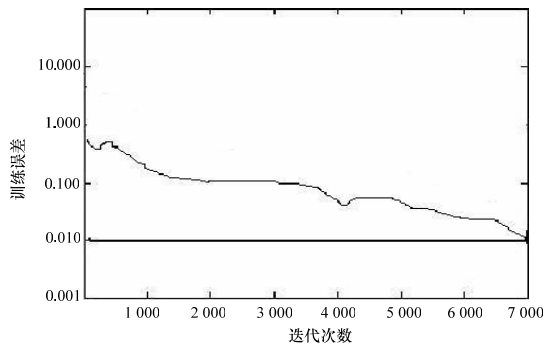


图 3 约简前神经网络训练误差曲线

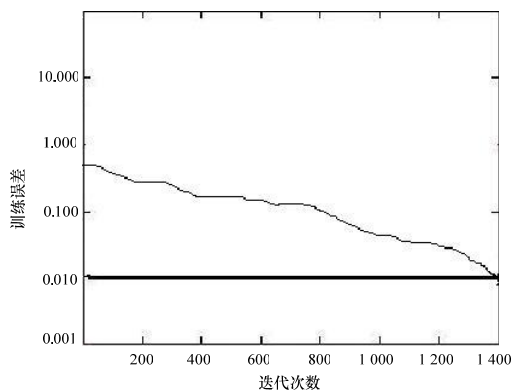


图 4 约简后神经网络训练误差曲线

### 3.4 规则提取

本文采用文献[6]的子集法对最终的神经网络进行规则提取, 得到如下一个规则集, 共计决策规则 20 条:

- (1)if WSYL=3 ∧ FSSD=1 ∧ JGXS=2 ∧ MCQJ=3 ∧ XDLX=1 then TCQD=中
- (2)if WSYL=4 ∧ FSSD=4 ∧ JGXS=1 then TCQD=大
- (3)if WSYL=2 ∧ FSSD=3 ∧ JGXS=1 ∧ MCQJ=3 ∧ GZFS=2 then TCQD=大
- (4)if WSYL=4 ∧ FSSD=1 ∧ JGXS=1 ∧ MCQJ=2 ∧ XDLX=1 then TCQD=中
- (5)if WSYL=4 ∧ FSSD=2 ∧ JGXS=2 ∧ GZFS=3 ∧ XDLX=3 then TCQD=大
- (6)if WSYL=1 ∧ FSSD=2 ∧ JGXS=1 ∧ MCQJ=1 ∧ XDLX=1 ∧ GZFS=2 then TCQD=无
- (7)if WSYL=1 ∧ FSSD=2 ∧ JGXS=4 ∧ GZFS=2 ∧ XDLX=2 then TCQD=无

- (8)if WSYL=3 ∧ FSSD=4 ∧ JGXS=2 ∧ GZFS=3 ∧ XDLX=3 then TCQD=大
- (9)if WSYL=1 ∧ FSSD=2 ∧ JGXS=4 then TCQD=无
- (10)if WSYL=4 ∧ FSSD=1 ∧ JGXS=3 ∧ GZFS=2 ∧ XDLX=1 then TCQD=小
- (11)if WSYL=4 ∧ FSSD=3 ∧ JGXS=2 then TCQD=大
- (12)if WSYL=2 ∧ FSSD=4 ∧ JGXS=2 ∧ GZFS=1 then TCQD=大
- (13)if WSYL=4 ∧ FSSD=1 ∧ JGXS=3 ∧ MCQJ=2 ∧ XDLX=1 then TCQD=中
- (14)if WSYL=4 ∧ FSSD=1 ∧ JGXS=2 ∧ MCQJ=1 ∧ XDLX=1 then TCQD=中
- (15)if WSYL=3 ∧ FSSD=2 ∧ JGXS=1 ∧ MCQJ=3 ∧ XDLX=3 ∧ GZFS=1 then TCQD=大
- (16)if WSYL=3 ∧ FSSD=4 ∧ JGXS=2 ∧ MCQJ=4 ∧ XDLX=3 then TCQD=大
- (17)if WSYL=1 ∧ FSSD=3 ∧ JGXS=4 ∧ GZFS=2 ∧ XDLX=2 then TCQD=小
- (18)if WSYL=1 ∧ FSSD=1 ∧ JGXS=4 then TCQD=无
- (19)if WSYL=3 ∧ FSSD=3 ∧ JGXS=1 ∧ MCQJ=4 ∧ XDLX=3 then TCQD=大
- (20)if WSYL=1 ∧ FSSD=2 ∧ JGXS=2 ∧ GZFS=3 then TCQD=无

由此得出, 影响该煤矿瓦斯突出的主要因素是瓦斯压力、放散初速度、坚固系数。瓦斯压力越大, 放散初速度越快, 坚固系数越低, 突出可能性和危险程度越大。另外, 工作方式、煤层倾角、巷道类型起到了一定的促进作用, 特别是采取放炮工作方式使煤体内部突然解除约束, 加快瓦斯突出。

### 3.5 测试结果及比较

同样的训练数据, 分别使用从单独使用神经网络提取的规则、单独使用粗糙集提取的规则与本模型提取的规则对 529 条测试样本进行测试并比较, 结果如表 6 所示。从中可以看出, 本文模型在规则数目和预测精度上都较采用单一方法优越, 所得规则能够对煤矿瓦斯预测做出准确判断。

表 6 不同方法测试结果比较

方法	规则数目	测试样本总数	可测试样本数	测试准确率/(%)	测试误差率/(%)
本文模型	20	529	529	100.00	0.00
单独使用神经网络	38	529	412	77.88	22.12
单独使用粗糙集	26	529	447	84.49	15.51

## 4 结束语

本文针对现有煤矿瓦斯监测监控信息网络, 提出了基于粗糙集与神经网络的煤矿瓦斯预测知识获取模型, 解决了煤矿瓦斯数据量大、冗余多、有噪声干扰的问题。将模型嵌入煤矿瓦斯预测专家系统中是下一步的工作。

### 参考文献

- [1] 赵卫东, 陈国华. 粗集与神经网络的集成技术研究[J]. 系统工程与电子技术, 2002, 24(10): 103-107.
- [2] 谢宏, 程浩忠, 牛东晓. 基于信息熵的粗糙集连续属性离散化算法[J]. 计算机学报, 2005, 28(9): 1570-1574.
- [3] 梁吉业, 孟晓伟. 信息熵在粗糙集理论中的应用[J]. 山西大学学报, 2002, 25(3): 281-284.
- [4] Hagan M T, Demuth H B, Beale M. Neural Network Design[M]. 北京: 机械工业出版社, 2003.
- [5] 周志华, 陈世福. 神经网络规则抽取[J]. 计算机研究与发展, 2002, 39(4): 398-405.
- [6] Fu Limin. Rule Generation from Neural Networks[J]. IEEE Trans. on Systems, Man and Cybernetics, 1994, 24(8): 1114-1124.

编辑 张帆