

文章编号: 1001-1986(2010)03-0066-03

粗糙集与神经网络技术预测煤厚及小断层的方法

王 新^{1,2}, 崔若飞², 陈同俊²

(1. 中国矿业大学计算机学院, 江苏 徐州 221116;
2. 中国矿业大学资源与地球科学学院, 江苏 徐州 221116)

摘要: 提出了基于分析三维地震数据的粗糙集(RS)—神经网络(NN)技术, 预测采区断层和煤层厚度变化。利用粗糙集对地震数据中所包含的大量干扰数据进行约简, 生成低噪音数据; 将约简后的数据输入神经网络进行训练获得断层识别和煤层厚度预测。实际数据验证表明, 该方法具有较高的精度。

关键词: 粗糙集; 神经网络; 断层预测; 煤层厚度预测

中图分类号: TP311 文献标识码: A DOI: 10.3969/j.issn.1001-1986.2010.03.015

Method to predict coal seam thickness and small fault using RS and NN

WANG Xin¹, CUI Ruofei, CHEN TongJun²

(1. School of Computer Science, China University of Mining and Technology, Xuzhou 221116, China;
2. School of Resource and Earth Science, China University of Mining and Technology, Xuzhou 221116, China)

Abstract: The thesis put forward a new method of Rough Sets (RS) and Neural Network (NN) technique to detect small faults and coal seam thickness by analyzing 3D seismic data. This method uses RS to reduce seismic data noise, and after reduction, low noise seismic data can be hold. After inputting those reduced data to NN, a predicting model which can detect small faults and predict coal seam thickness can be achieved after NN's training. After this step, this model was used to detect small fault by 3D seismic data. We find that this method has high precision.

Key words: rough sets; neural network; detecting small fault; predicting coal seam thickness

随着煤矿开采深度的增加和构造复杂矿井的大量投产, 煤矿开采过程中出现的复杂地质构造和煤层厚度变化对煤矿正常生产及矿工生命安全产生了越来越大的影响。因此, 几乎所有的国有大型煤矿都进行了三维地震勘探工作, 以期能够最终解决复杂地质构造和煤层厚度变化等地质问题。但是, 三维地震勘探所获取的煤层地震属性数据中含有大量的噪声数据, 如何将其中代表地质构造异常和煤层厚度变化异常的信息准确识别已成为当前地震资料解释中的一个难点。

本文将粗糙集与神经网络技术引入到预测煤层厚度与小断层识别领域。利用粗糙集约简决策表能够减少输入的特点, 将三维地震数据中的大量冗余数据进行约简, 提取出与煤层构造或煤层厚度变化相关的地震属性。然后, 将约简后的数据输入神经网络训练, 获得小断层识别结果或煤层厚度预测结果。

1 粗糙集与神经网络理论

1.1 粗糙集理论

粗糙集(RS)方法^[1]最初是由波兰数学家 Pawlak 于 1982 年提出, 目前已成为一种处理不精确、不确定与不完全数据的新的数学理论。粗糙集无需提供除问题所需处理的数据集合之外的任何先验信息, 这是粗糙集方法的最重要的优点之一。此外, 粗糙集的另一个优势是能够有效地去除冗余。因此, 粗糙集方法可以克服传统不确定处理方法的不足, 为处理不精确、不完全等数据的决策问题提供更接近人类认知的数学工具。

属性约简是粗糙集理论中最为重要的部分之一, 目前国际上已存在多种属性约简算法。其中, Skowron 等人提出的基于分辨矩阵的经典属性约简算法最为简单、直观, 已成为众多小型数据库数据进行属性约简的首选方法。后来, 不少学者相继提出了基于属性重要性、属性频度等多种改进算法, 一方

收稿日期: 2009-08-16

基金项目: 国家自然科学基金项目(40804026); 中国矿业大学青年科研基金项目(2009A054)

作者简介: 王 新(1979—), 女, 山东临沂人, 讲师, 博士研究生, 从事智能信息处理。

面为缩小算法的执行时间和使用空间;另一方面增加一些启发式思想,保证算法能得到一个约简,使算法更加完备。笔者根据三维地震数据的特点,采用基于属性重要度和属性依赖度的属性约简算法^[2]。基于属性重要度和属性依赖度的决策表属性约简算法可得到一个用户需求的最小约简,优点是简单、直观,能够处理较大的数据集,算法的复杂度较小。算法可以描述为:

输入 决策表 $DT=(U, CUD, V, f)$,

式中 U 是论域; $C=\bigcup_{k=1}^n a_k$ 是条件属性集; D 是决策属性集; V 为属性的值域; f 是一个信息函数。

输出 该决策表的一个最小约简 B 。

第1步:计算决策表 DT 中的每一个条件属性 a_k 的重要度 $sig(a_k, C, D)$;

第2步:记 C 相对于 D 的核为 $C_0=\{a_k|sig(a_k, C, D)=1\}$, 令 $B=C_0$, $M=C-B$;

第3步:使用 $sig(a_k, C, D)$ 作为启发式信息

a. 选择使 sig 最大的属性 a_k , 且 $B=B \cup \{a_k\}$;

b. 计算决策属性 D 对条件属性 C 的属性依赖度 $r_c(D)$;

c. 若决策属性 D 对最小约简 B 的属性依赖度 $r_B(D)=r_c(D)$ 终止, 否则转 b;

第4步:输出最终得到的 B 就是 C 相对于 D 的一个最小约简, 算法结束。

1.2 神经网络理论

神经网络(NN)是由大量的、同时也是很简单的处理单元广泛地互相连接而形成的复杂网络系统,它反映了人脑功能的许多本质特征,是一个高度复杂的非线性动力学系统。

神经网络具有大规模并行、分布式存储和处理、自组织、自适应和自学习能力,特别适用于处理需要同时考虑许多因素和条件的、不精确和模糊的信息处理问题。神经网络的发展与神经科学、数理科学、认知科学、计算机科学、人工智能、信息科学、控制论、机器人学、微电子学、心理学、光计算、分子生物学等有关,是一门新兴的边缘交叉学科。

BP网络是一种多层前馈神经网络,包含了神经网络理论中精华的部分,每一个训练样本在网络中都经过两次传递计算:首先向前传播计算,从输入层开始,通过各层并经过处理后,产生一个输出,并得到实际输出与期望输出之间差距的误差矢量;而后,则通过反向传播计算,从输出层到输入层,利用误差矢量对权值进行逐层的修改。BP网络可以实现从输入到输出的任意非线性映射,采用的是误差

逆向传播算法。由于BP网络结构简单、可塑性强,BP网络得到了广泛的应用。特别是它的数学意义明确、步骤分明的学习算法更使其具有广泛的应用背景。

1.3 粗糙集与神经网络的结合

粗糙集方法与人工神经网络方法在处理信息时,存在如下两方面的差异^[3]: a. 人工神经网络处理信息时,一般不能将输入信息空间的维数减少,所以当输入的信息空间维数较大时,网络不仅结构复杂,而且训练时间很长,而粗糙集方法通过挖掘数据间关系,不仅可以去掉冗余输入信息,而且可以简化输入信息的表达空间;b. 在实际问题处理中,人工神经网络方法有较好的抑制噪音干扰能力,而粗糙集方法对噪音比较敏感。通常用无噪音的训练样本学习推理结果去处理有噪音环境中的信息,一般应用效果不佳。

因此,将粗糙集与神经网络结合起来处理三维地震数据,能够实现优势互补。图1是粗糙集与BP神经网络方法预测煤层厚度及小断层识别的算法流程图。

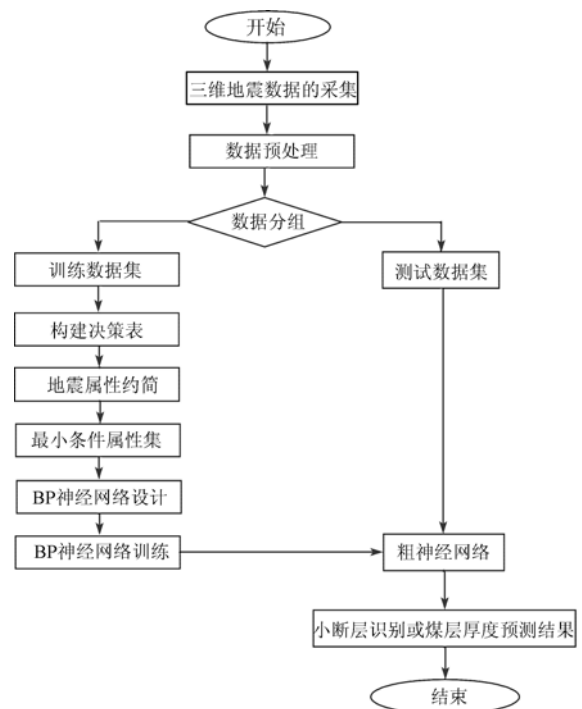


图1 粗糙集与BP神经网络的算法流程图

Fig. 1 Flow chart of algorithm using RS and BP NN

2 粗糙集与神经网络在煤层小断层识别中的应用

梁家煤矿位于山东龙口,区内小断层极为发育,故选取梁家煤矿东六采区的地震属性数据来检验粗糙集与神经网络方法识别小断层的可靠性。

2.1 样本预处理

选取梁家煤矿东六采区地震数据的地震波运动

学、地震波动力学、地震波时间域和频率域等 18 种地震属性,共 419 组数据作为样本。根据样本数据的实际情况采取了最小-最大规范化方法,将各属性值映射到[0, 1]区间。由于实际地震资料难免存在干扰,所以,对样本数据通过中值滤波方法试图减少干扰对预测精度的影响。

2.2 利用粗糙集对样本约简

选择基于属性重要度和属性依赖度的约简算法对梁家煤矿东六采区的三维地震数据进行属性约简,将约简结果(平均频率相位(Pha)、主频相位(Phm)、主频($Fmain$)、相关系数(Rxy)、二维分形参数($Ttd2$)、宽频带总能量(Qfw)、容量维(Df))等 7 个属性作为神经网络的输入向量。

2.3 神经网络预测及结果对比

实验所采用的神经网络结构为 7-20-20-1,采用基于数值最优化理论的 trainbfg 训练算法^[4],均方误差 MSE 设为 0.000 01,最大循环次数为 10 000 次。利用其对梁家煤矿 Inline100 线地震剖面进行预测,其预测结果如图 2 所示(横坐标为 CDP 序号,间隔为 5 m,纵坐标为预测值)。在图 2 中,认为预测值大于 0.8 的点为断层。

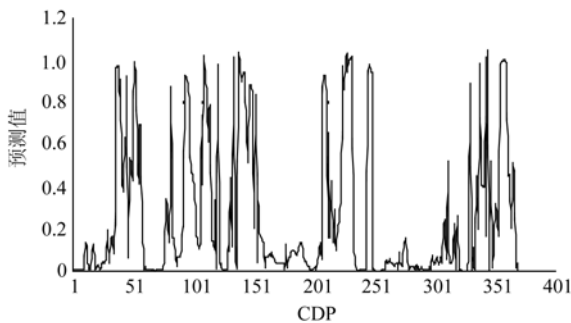


图 2 Inline 100 线神经网络预测结果
Fig. 2 Network's predicting of Inline 100

为了检验此预测结果是否正确,将预测结果与常规小断层解释结果(计算机手工交互解释和相干体算法计算)进行对比分析(表 1),可见,粗糙集与神经网络方法对于识别小断层,其识别能力要明显强于手工解释和相干体算法。但有时由于粗糙集与神经网络方法过于敏感,在少数情况下也可能做出错误预测。

3 粗糙集与神经网络在煤层厚度识别中的应用

徐庄煤矿某采区已知钻井有 14 口(图 3),主采煤层厚度介于 1.2~8.7 m,煤层厚度变化剧烈。为了验证粗糙集与神经网络方法识别煤层厚度的有效性,采用粗糙集与神经网络方法对煤层厚度的预测结果进行交叉验证。

表 1 不同方法识别小断层异常对比表
Table 1 Small faults' comparison among different detecting methods

位置	粗糙集与神经网络	手工解释	相干体算法	验证结果
41 道	是	否	否	是
52 道	是	否	否	是
82 道	是	否	否	是
96 道	是	否	否	否
110 道	是	是	是	是
138 道	是	是	是	是
209 道	是	否	否	否
225 道	是	是	是	是
247 道	是	是	是	是
335 道	是	是	是	是
343 道	是	是	是	是
355 道	是	是	是	是

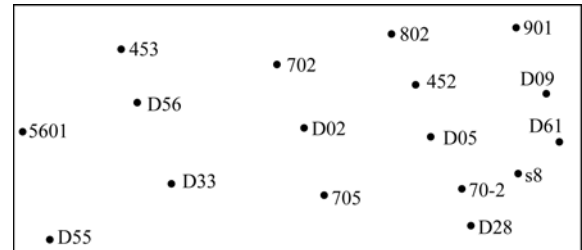


图 3 徐庄煤矿采区钻井分布

Fig. 3 Borehole distribution of Xuzhuang Mine

3.1 样本预处理

选取徐庄煤矿的地震数据的地震波运动学、地震波动力学、地震波时间域和频率域等 18 种地震属性,共 82 619 组数据。根据样本数据的实际情况采取了最小-最大规范化方法,将各属性值映射到[0, 1]区间。

3.2 利用粗糙集对样本约简

选择基于属性重要度和属性依赖度的约简算法^[2]对徐庄煤矿的三维地震数据进行属性约简,将约简结果(波峰振幅(AI)、低频带能量(Qf)、平均频率相位(Pha)、二维分形参数($Ttd2$))作为神经网络的输入向量。

3.3 神经网络预测结果

任意选取井号为 D55、D56、705、D33、702、D28、D05、452、802、901、D09 等 11 口井的实际煤层厚度进行训练,预测井号为 5601、453、D02 的煤层厚度。

实验所采用的神经网络结构为 4-20-1,采用基于数值最优化理论的 trainbfg 训练算法^[4],均方误差 MSE 设为 0.000 01,最大循环次数为 10 000 次。交叉验证时,选取井旁 9 个地震道作为训练样本。

(下转第 72 页)