

夏店煤矿断层的支持向量机预测研究

韩成阳¹⁾, 邹冠贵^{1,2)}, 薛静雯¹⁾

1) 中国矿业大学(北京)地球科学与测绘工程学院, 北京, 100083;

2) 中国矿业大学(北京)煤炭资源与安全开采国家重点实验室, 北京, 100083

关键词: 地震属性; 支持向量机; 断层预测

近年来夏店煤矿在开采过程中,发现采区内断层特别发育,严重影响了煤矿的高效开采与经济效益。因此亟须通过物探手段来解决当下遇到的困局,不过传统的地震解释周期长,人工多,不能够满足煤矿的迫切需求。不过随着计算机科技的进步,将机器学习技术应用于地震数据解释开始受到关注,例如使用和深度学习方法在弹性波阻抗数据体上,开展地震相分类。基于支持向量机技术的断层特征自动提取,从地震数据中进行断层检测,已有结果表明,机器学习不仅能够发挥地震多属性的优势,开展分类识别,而且还能够提高工作效率,在地震勘探中体现出可喜结果(Zou et al., 2019),是未来发展趋势。因此在夏店矿区采用 SVM 预测研究,既能满足煤矿生产需要,还能为后续的智能识别工作提供帮助。

1 支持向量机

支持向量机于 20 世纪 90 年代被首次提出(Cortes and Vapnik, 1995),它的基本思想是,先将较低维空间中的原始数据映射到近似线性可分的一个高维空间中,然后在高维空间中基于最大间隔原则的线性学习。支持向量机预测技术是基于已知数据,通过算法来解析训练数据集,进行训练学习,获得认可的模型,并利用该模型对测试数据进行分析。在地震剖面或者沿层属性上识别断层,可以视为支持向量机的两分类问题。基于支持向量机的分类特性,开展自动识别断层,需要做好数据集和算法构建,进而形成支持向量机模型。

支持向量机(SVM)是解决分类问题的有用工具。支持向量机加入了核函数的概念,核函数的引入,使得非线性分类可以“先升维后内积”,极大

地减少运算量,提高了运算速度,降低了对计算机运算能力的要求,特别适用于开展分类问题研究。那么如何设置最佳的 SVM 参数也决定了 SVM 分类器的分类预测能力,好的参数设置能有效提高分类器的性能。

因此在断层识别的优化过程中就面临两个问题:一是如何选择最佳的特征子集;二是如何设置最佳的 SVM 参数。特征子集的选取在很大程度上影响到 SVM 的分类精度和运行时间,SVM 参数设置决定了分类器的分类性能。

2 模型的构建

2.1 样本数据评估与选择

通过对矿区内目标煤层的地震数据提取地震属性,包括瞬时相位、方差、混沌体、最大振幅、均方根振幅、平均振幅、振幅和、瞬时频率、倾角、弧长等形成工区属性数据。将已知区域的属性值和断层信息汇总成 SVM 模型的样本数据,记录每个取样点处,是否存在断层,“存在”用‘1’表示,“不存在”用‘0’表示,共提取 7963 个样本数据。通过分析样本数据,计算各属性间的相关系数,然后进一步利用 R 型聚类分析,评估各属性间的相关性,得出:方差、混沌体、平均能量、瞬时相位、瞬时频率、均方根振幅相关性较低,相对独立,同时 6 种属性都具有断层的地质意义,因此选择这 6 种属性作为样本数据的属性值。

2.2 SVM 模型的参数优化

由于样本数据的 6 种属性值的量纲不同,直接用于模型训练,会影响网络训练结果。所以在训练之前,对样本进行归一化处理,以消除量纲。从处理后的数据中,选出 2000 个样本作为测试数据,剩下的用作训练数据。利用粒子群算法,把预测断

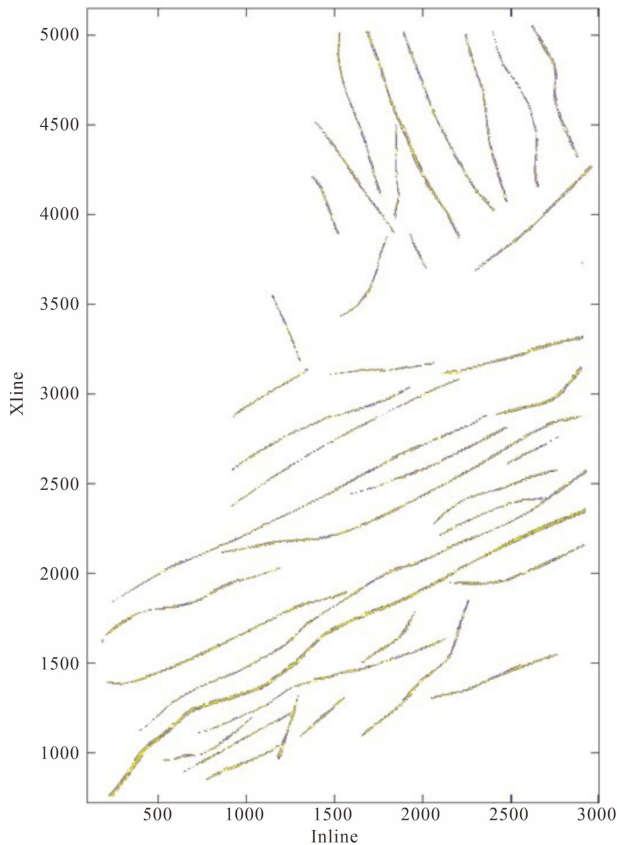


图 1 夏店矿区目标煤层构造展布预测图

层的支持向量机模型的正确率作为适应度函数，将模型的惩罚参数 c 和核函数参数 g 作为待优化参数，优化预测断层的支持向量机模型。通过 100 次的粒子群算法的迭代优化，得到最佳的惩罚参数 $c=4.3687$ ，核函数参数 $g=11.2608$ 。

2.3 SVM 模型的测试分析

确定最佳参数后，我们利用测试集测试模型准确率情况并记录，最后将 10 次的测试结果统计并求取平均准确率可以达到 93.43%。为了更好的评价各 SVM 模型的断层识别能力，本研究选用了 4 个参数评价模型性能。它们分别是准确率、查准率、查全率和 F 值。其中，准确率是指预测正确的样本点在总样本点中占的比例，是评价预测效果的常用指标。查准率是指预测正确的断层样本在所有预测为断层样本中占的比例。查全率是指预测正确的断层样本在确实为断层的样本中占的比例，代表了模型在断层样本的预测能力。 F 值通过查准率和查全率两项指标计算调和平均数得到。模型评价参数的

表 1 模型评价参数

测试次数	准确率	查准率	查全率	F 值
10	0.9343	0.8172	0.9340	0.8717

计算结果如下（表 1）。

2.4 矿区断层的预测结果

将预测结果与坐标点进行关联，形成一个三维矩阵，分别是由 inline、xline 和预测结果组成。利用绘图软件进行绘图，得到了夏店矿区基于支持向量机的断层预测分布图（图 1）。从图中可以看出采取南边的断层主要走向为北东向，而采区北部断层走向近北向，整体分布也与矿区已知的走向分布一致。

3 结论

（1）基于支持向量机模型进行采区断层的自动检测，预测效果较好。基于机器学习的断层自动识别方法是一类能够在地震数据体中自行识别并解释断层的技术，从而提高了断层解释的效率，降低了解释成本和时间。

（2）支持向量机模型预测利用已知的断层数据、非断层数据构成训练数据集和测试数据集，因此融合了已知的地质先验信息，能够进一步提高预测的准确率。

（3）支持向量机模型预测是利用多种地震属性，开展非线性分析，避免了线性假设的前提，提高解释的精度。

参 考 文 献 / References

- Cortes C, Vapnik V. 1995. Support-vector networks. Machine learning, 20: 273~297.
- Zou G, Ren K, Sun Z, et al. 2019. Fault interpretation using a support vector machine: A study based on 3D seismic mapping of the Zhaozhuang coal mine in the Qinshui Basin, China. Journal of Applied Geophysics, 171: 103870.

HAN Chengyang, ZOU Guangui, XUE Jingwen: Study on fault prediction of Xiadian Coal Mine by SVM

Keywords: seismic attribute; SVM; fault prediction