

# 一种测井岩性识别的寻优模型

魏志华 张俊儒

(武汉理工大学计算机科学与技术学院 武汉 430070)

**摘要** 岩土作为一种极其复杂的材料,通常会受到外界多种因素的影响而发生变化,这些影响因素既包括断层、裂隙、雨水冲刷和腐蚀等天然环境因素,也包括众多的人为因素,从而导致测井岩性的识别会产生大量的干扰数据。在对大数据量的信息寻优处理的算法中,支持向量机(Support Vector Machine, SVM)是一种受到广泛关注的寻优方法。但是传统的 SVM 寻优方法存在耗时长、收敛慢的缺陷,因此将传统 SVM 寻优当中的留一交叉法改为 K 折交叉法,并利用这种优化的 SVM 对测井岩性数据进行寻优处理,来进行测井岩性的识别。对比试验结果表明,相对于传统支持向量机的寻优算法,该方法具有识别正确率高、收敛速度快等优点。

**关键词** 测井岩性识别, SVM, 最优数据

**中图分类号** TP392 **文献标识码** A

## Optimization Model of Logging Lithological Identification

WEI Zhi-hua ZHANG Jun-ru

(Department of Computer Science and Technology, Wuhan University of Technology, Wuhan 430070, China)

**Abstract** Rock-soil is a kind of extremely complex material, and it will encounter a lot of influence, these factors include the slip, fracture, rain erosion, corrosion and other natural factors and artificial factors. So many impacts result in the identification of lithology will produce a large number of error data. In the large amounts of data information optimization processing algorithm, support vector machine (Support Vector Machine, SVM) is a widely concerned optimization method. But the traditional SVM method is time-consuming, so we optimized the traditional SVM, made the leave one out cross method change to K fold cross method, and used optimized SVM optimize large data searching. Comparative test results show that the method has the advantages of high recognition accuracy, quick convergence speed, compared with the traditional optimization algorithm of support vector machine.

**Keywords** Logging lithology identification, SVM, Optimal data

## 1 引言

测井岩性识别是进行地层评价、油藏描述等工作的一项重要内容<sup>[1,2]</sup>。岩性分析的基本资料来源于测井资料中包含的岩性信息。当今,测井技术飞速发展的同时,合理的应用测井数据成为获得岩性信息的重要途径之一。1982 年 Wollf 等人提出根据测井资料来自动地判定地层岩性。从此以后,利用计算机进行识别岩性的操作已成为测井技术的重要手段,受到国内外相关研究人员的广泛关注。

由于岩土材料的非均质、非线性等特性,测井岩性所探测到的数值参数并不能很好地、直接地引入计算模型,在相关数据中存在大量的干扰数据,因此在实际应用中需要先对原始数据进行预处理。而支持向量机<sup>[3-5]</sup>是一种对规模较大的数据集进行快速寻优预处理的有效方法。

## 2 SVM 算法介绍

支持向量机<sup>[6-8]</sup>(Support Vector Machine, SVM)是一种模式识别方法,它根据 Vapnik<sup>[9]</sup>提出的结构风险最小化原则来提高学习机泛化能力,通过有限样本获取的决策规则,对独

立的测试集计算最小误差,通常比采用经验风险最小化原则的神经网络<sup>[10]</sup>具有更好的泛化能力。该方法在解决小样本、非线性、过拟合以及高维模式识别的问题中都表现出了许多特有的优势,并能够推广应用到函数拟合等其他机器学习问题中。有些研究人员已经讨论了支持向量机算法进行岩性识别的可行性与有效性,并且运用支持向量机算法在测井岩性识别中取得了较好的效果<sup>[11]</sup>。该方法对过渡岩性的识别有较高的精度,但也存在着很多的不足,比如 SVM 算法对大规模训练样本难以实施,并且用 SVM 解决多分类问题也存在困难。

针对上述方法存在的问题,本文在支持向量机算法的基础上加以改进,使得 SVM 能够更好地完成岩性的寻优识别。

## 3 测井岩性模型及寻优

SVM 寻优的关键取决于选取一组合适的参数,包括惩罚因子 C、不敏感参数以及 RBF 核参数等。

传统的 SVM 参数寻优需对(c: 惩罚系数, g: 核函数参数, p: 损失函数参数)组合在给定范围(3 个参数的范围一般为:  $\log_2 c = -1: 1: 7, \log_2 g = 0: -1: -8, \log_2 p = -8: 1: 0$ )内

魏志华(1964—),男,博士,教授,主要研究方向为人工智能, E-mail: wgaodi@sina. cn; 张俊儒(1990—),男,硕士生,主要研究方向为人工智能。

进行穷尽搜索,搜索次数为3个参数向量长度的乘积,搜索时间为搜索次数乘上训练样本个数。这就限制了样本的数目,当样本数目过大时,所需的计算时间就会急剧增加。

图1为传统SVM的流程。在该流程中,对于预测数据的寻优,采取的是基本的留一法交叉验证,它的优势在于每一个模型都是用几乎所有的样本来进行训练的,最接近样本,这样评估所得的结果比较可靠。实验没有随机因素,整个过程是可重复的。但是缺点同样明显即计算成本高,当 $n$ 非常大时,计算耗时。

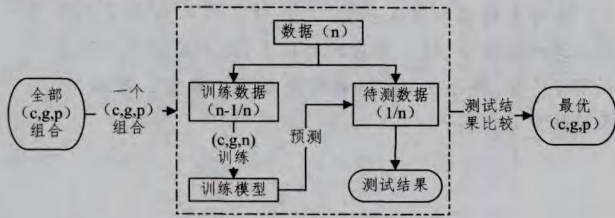


图1 传统SVM流程

针对测井岩性的大数据量,本文作者采取 $K$ 折交叉验证来提高速率。

### 3.1 交叉验证

交叉验证主要用于预测,即可以估计一个预测模型在实际应用中的准确度。它是一种将数据样本切割成较小子集的实用方法,表现为先在一个子集上做分析,而其它子集则用作后续对此分析的确证与验证。

一个交叉验证将大数据样本集划分为两个互补的子集,分别用作训练模型和验证模型。包含大部分数据的子集定义为训练集(training set),作为训练模型;反之则为测试集(testing set),作为验证模型。利用测试集来测试训练得到的模型,以此作为模型的性能指标。然后让测试集数据对训练集数据进行预报,记录下测试集的预报误差和它们的平方和。然后循环整个过程,直到所有样本数据都有且仅有一次被预报。

基于以上交叉验证的方法,数据量越多的训练集会使测试的时间大大增加。所以对于测井岩性寻优模型而言,我们采取如图2所示的 $K$ 折交叉验证( $K \geq 2$ )。

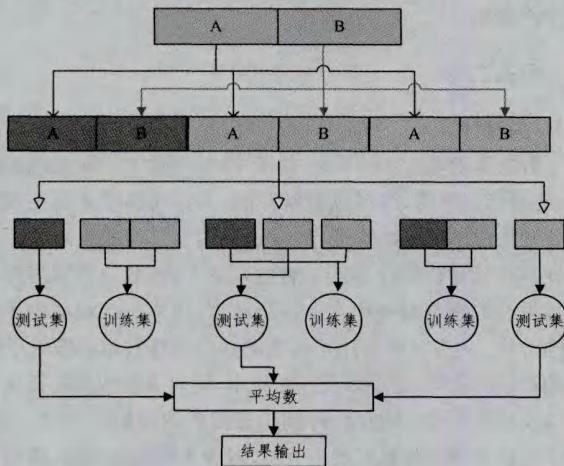


图2  $K$ 折交叉验证基本流程

具体操作过程是:将样本数据集随机划分为 $K$ 个子集(一般是均分),将一个子集数据作为测试集,其余的 $K-1$ 个子集作为训练集;将 $K$ 个子集轮流作为测试集,重复上述过

程,这样得到了 $K$ 个模型,并利用测试集得到了 $K$ 个模型的性能指标。

这种验证方法的优点是每一个样本数据在被用作训练数据的同时,也被用作测试数据,避免了过度学习和欠学习状态的发生,得到的结果比较具有说服力。

但是进行交叉验证需要注意:

①训练集中样本数量要足够多,一般至少大于总样本数的50%。

②训练集和测试集必须从完整的数据集中均匀取样。均匀取样的目的是希望减少训练集、测试集与原数据集之间的偏差。当样本数量足够多时,通过随机取样也能实现均匀取样的效果(随机取样,可重复性差)。

### 3.2 仿真实验

首先,需要生成指定数目的随机数,然后把随机数分成两部分,一部分作为训练样本集,一部分作为测试集。其导出数据的过程如图3所示。

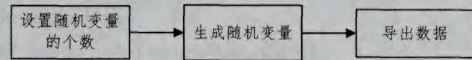


图3 数据生成过程

导出数据图如图4所示。

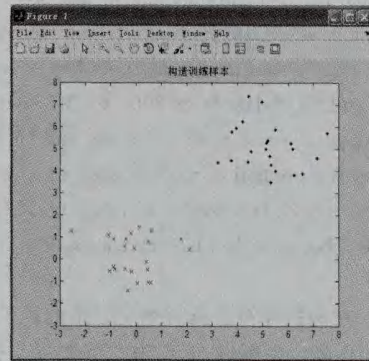


图4 生成训练样本

在此首先在同一区域生成两份同样数量(20个)的随机数,接着把其中一份全部加上5,最后得到了如图4显示的训练样本集和测试集。

图5示出寻优算法的测试流程。

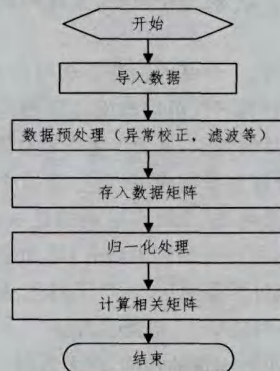


图5 寻优算法的测试流程

经过SVM优化寻优算法的相关处理之后,得到了图6

(下转第450页)

能够验证忆阻器混沌电路对初始值的高度依赖性。

## 参考文献

- [1] Chua L. Memristor-The missing circuit element [J]. IEEE Transactions on Circuit Theory, 1971, 18(5): 507-509
- [2] Strukov D B, Snider G S, Stewart D R, et al. The missing memristor found[J]. NATURE, 2008, 459(7250): 1154
- [3] Itoh M, Chua UA L O. Memristor Oscillators [J]. International Journal of Bifurcation and Chaos, 2008, 18(11): 3183-3206
- [4] Kyprianidis I M, Volos C K, Stouboulos I N. Chaotic Dynamics from a Nonlinear Circuit based on Memristor with Cubic Nonlinearity[J]. AIP Conference Proceedings, 2010, 1203(1): 626-631
- [5] Muthuswamy B, Chua L O. Simplest Chaotic Circuit[J]. Inter-

national Journal of Bifurcation and Chaos, 2010, 20(5): 1567-1580

- [6] Chua L O, Sung M K. Memristive devices and systems[J]. Proceedings of the IEEE, 1976, 64(2): 209-223
- [7] Xu Bi-rong. A simplest parallel chaotic system of memristor[J]. Acta Phys. Sin, 2013, 62(19): 99-106
- [8] Adhikari S P, Kim H. Why Are Memristor and Memistor Different Devices? [J]. IEEE Transactions on Circuits and Systems I: Regular Papers, 2012, 59(11): 2611-2618
- [9] Biolkova V, Kolka Z, Biolk Z, et al. Memristor modeling based on its constitutive relation; Advances in Communications, Computers, Systems, Circuits and Devices, 2010[C]// Puerto De La Cruz, SPAIN, 2010

(上接第 431 页)  
所示结果。

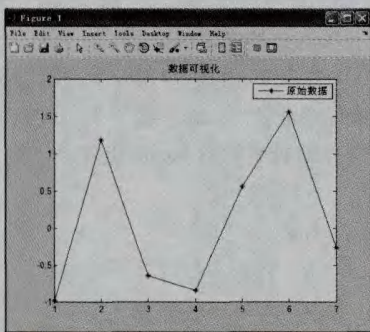


图 6 寻优结果(1)

这就是 20 个数据经过寻优之后的结果:只剩下 7 个最优的数据。如果把数据提升到了 100 个,会生成图 7 所示结果。

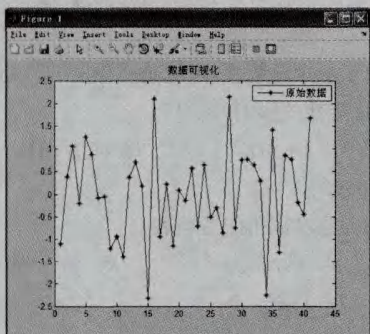


图 7 寻优结果(2)

可以看出,当数据量增大时,寻优下来的数据也会增多,而且该算法处理的结果符合寻优结果。

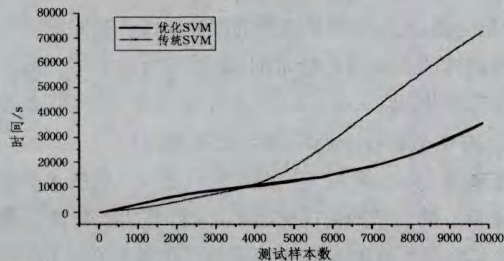


图 8 寻优速率对比图

经过仿真实验,得到图 8 所示的寻优速率对比图。实验

结果表明,测试数量增加的同时,传统 SVM 寻优的时间也会慢慢呈现阶数级递增,而优化之后的 SVM 寻优所需时间的增长幅度则要慢得多。

**结束语** 在测井岩性识别的应用中,相对于传统支持向量机 SVM 算法来说,经过优化之后的 SVM 方法在保证了处理精度的同时,显著提高了系统处理速率,有效节省了处理时间。

## 参考文献

- [1] Jatinder N D, Gupta, Sexton R S. Comparing backpropagations with a genetic algorithm for neural network training[J]. Omega, 1999, 27(6): 678-684
- [2] Lian Zeng-zeng, Tan Zhi-xiang, Deng Ka-zhong. Closeness Degree Methods to Identify Logging Lithology in Coalfield Base on Fuzzy Pattern Recognition[J]. Well Logging Technology, 2013 (3)
- [3] Keerthi S S, Shevade S K, Bhattacharyya C, et al. A fast iterative nearest point algorithm for support vector machine classifier design[J]. IEEE Transactions on Neural Networks, 2000, 11(1): 124-136
- [4] Shevade S K, Keerthi S S, Bhattacharyya C, et al. Improvements to the SMO algorithm for SVM regression[J]. IEEE Transactions on Neural Networks, 2000, 11(5): 1188-1193
- [5] Akyildiz I F, Su W. Wireless sensor networks a survey [J]. Computer Networks Journal, 2002, 38
- [6] Joachims T. Support Vector Machine[M]. 2008
- [7] Christopher, Burges J C. A Tutorial on Support Vector Machines for Pattern Recognition[M]. New York: Springer-verlag, 1998
- [8] Lipkowitz K B, Cundari T R, Wiley. Applications of Support Vector Machines in Chemistry [J]. Reviews in Computational Chemistry, 2007, 23: 291-400
- [9] Vapnik V N. The nature of statistical learning theory[M]. New York: Springer Verlag Press, 1995
- [10] Wu Xiao-juan, Zhu Xin-jian, Cao Guang-yi. Modeling a SOFC stack based on GA-RBF neural networks identification[J]. Institute of Fuel Cell, 2007, 167(1): 145-150
- [11] 于代国,孙建孟,王焕增. 测井识别岩性新方法——支持向量机方法[C]//大庆石油地质与开发, 2005(5)