

# Application of SVM-Based Lithology Identification in Building Rock Physical Property Templates of Junggar Basin

Huizhen Yu, Shuhua Wang, Tao Guo

Western Subdivision of Exploration & Development Research Institute, Shengli Oilfield Company, SINOPEC, Dongying Shandong  
Email: yhzabc@163.com

Received: Feb. 1<sup>st</sup>, 2016; accepted: Feb. 21<sup>st</sup>, 2016; published: Feb. 25<sup>th</sup>, 2016

Copyright © 2016 by authors and Hans Publishers Inc.  
This work is licensed under the Creative Commons Attribution International License (CC BY).  
<http://creativecommons.org/licenses/by/4.0/>



Open Access

---

## Abstract

It is a fundamental work to identify lithology for building rock physical property template and carrying out gravity-magnetic-magnetotelluric-seismic joint inversion & interpretation. Usually we use the cross-plot maps to identify lithology. Its lithology identification standard is given according to the difference of plane distribution of intersection point. However, this method has low accuracy because it is vulnerable to subjective factors, and can't be used in high-dimensional classification; especially if physical property data is complicated. To solve the problem, based on support vector regression machine theory, this paper proposes a more stable lithology identification method, and sets up multiple physical property templates of variant rocks of Junggar basin, and it has been successfully applied to lithology discrimination of Chepaizi area.

## Keywords

SVM, Lithology Identification, Physical Property Template, Junggar Basin

---

# SVM岩性识别方法在准噶尔盆地物性模板构建中的应用

于会臻, 王树华, 郭 涛

中石化胜利油田分公司勘探开发研究院西部分院, 山东 东营

Email: yhzabc@163.com

收稿日期: 2016年2月1日; 录用日期: 2016年2月21日; 发布日期: 2016年2月25日

## 摘要

岩性识别是构建岩石物性模板、开展重磁电震联合反演解释工作的基础。通常采用的交汇图方法进行岩性识别, 该方法根据判断交汇点平面分布差异性来给出岩石岩性识别标准, 易受主观因素影响, 且无法直接应用于多于三类物性数据的岩性识别, 尤其在岩石物性规律复杂时识别精度更低。针对该问题, 基于支持向量机理论提出了一种更稳定的重磁电震物性值岩性识别方法, 以此为基础构建了准噶尔盆地的岩石物性模板, 并在车排子地区的岩性判别中得到了较好的应用。

## 关键词

支持向量机, 岩性识别, 物性模板, 准噶尔盆地

## 1. 引言

重磁电震联合方法是融合多种地球物理技术针对同一地质目标开展同步处理、解释的一体化过程, 相比单一技术来说, 可丰富约束信息, 提高处理、解释的合理性[1]-[3]。Michele De Stefano (2011 SEG) 等指出寻找不同地球物理物性间联系函数是得到合理解空间的关键点。

物性模板的构建是联系这四种地球物理技术的桥梁, 是开展联合解释的基础, 它不仅可为山前带等复杂地质构造建模添加更合理的先验约束信息, 而且还可为预测有利火成岩储层提供解释依据。物性模板构建的前提条件是对岩性进行准确的识别。通常对多类物性数据进行岩性识别常采用的手段是交汇图法, 但该方法不适用于物性数据种类较多及岩石物性值规律较复杂的情形, 针对该问题, 本文提出了一种基于支持向量机(SVM)分类理论的重磁电震岩石岩性识别方法, 并将其成功用于准噶尔盆地的物性模板构建工作中, 为该工区后续复杂地质目标的构造建模、岩性划分工作提供了有力的参考依据。

## 2. 支持向量机物性分类方法

岩性识别过程是物性规律总结的逆向推理。交汇图法是最为常用的手段, 该方法首先选用以确定岩性的岩石物性数据作为输入变量, 利用交汇的方式进行展示, 然后根据物性数值分布范围圈定岩性识别区间, 当有未知岩性的岩石物性数据输入时, 按照圈定的岩石识别边界来确定岩性归属。图 1 示例为利用密度与磁化率的交汇图识别火成岩, 研究人员通过经验给出岩性识别边界, 如图中红、黄、绿、蓝色所示, 分别对应了安山岩、玄武岩、凝灰岩及流纹岩, 在不同岩石的物性值区分程度较大时, 识别较为容易, 但在物性值交叉点分布规律复杂时, 无法合理的给出识别标准, 图中黑色的圆点就无法准确的判断其为安山岩类还是凝灰岩类。此外, 在输入数据维数增多时, 例如当数据种类多于三类时, 交汇图无法进行直观展示, 利用重磁电震四种物性数据(密度、磁化率、电阻率及速度)进行识别变得愈加困难。

改善此类问题通常可利用神经网络[6] [7]、聚类分析、K 值近邻等数学统计工具。但这些算法仍存在着一定的问题。如神经网络虽然在小样本识别中比较有效, 但其网络拓扑结构尚无理论指导, 且易陷入局部最优解等缺点, 使其在岩性识别中应用效果不佳; 聚类分析方法适用于训练样本较多的情形, 而采集的重磁电震的物性数据十分有限, 属于小样本学习过程, 故限制了该方法在岩性识别中的应用。因此,

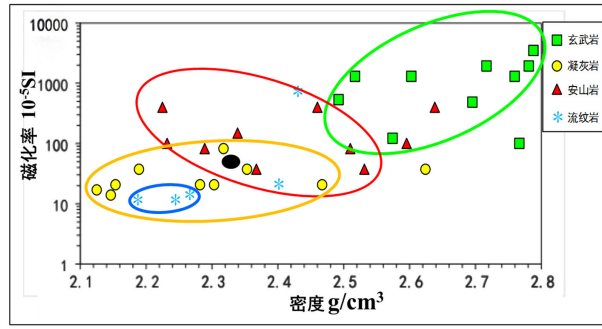


Figure 1. Lithology identification with by cross-plot maps  
图 1. 岩性圈定交汇图

本文尝试利用支持向量机(SVM)进行岩性分类。SVM 是 Vapnik [4] [5]在统计学习理论的 VC 维理论和结构风险最小原理基础上提出的一种先进的机器学习算法,其基本思想是根据实际情况来选择相应的核函数并将原始数据映射至高维空间,之后通过寻找高维空间中的最优分类超平面来保证所得模型既能满足已知样本的分类精度又具有较强的泛化能力。该方法在解决函数逼近[8]、小样本识别分类[9]-[11]及高维模式识别等问题中有着广泛的应用。下面具体介绍一下 SVM 是如何解决重磁电震物性数据的岩石岩性识别问题。

(1) 简单岩性识别问题,即物性数据差异较大的情况。此时,可通过寻找一组其间的超平面  $g(x)$  来将岩石物性数据分为两部分,其满足的数学表达式为:

$$g(x) = \langle w, x \rangle + b = 0 \quad (1)$$

其中,  $w$  为权重,即超平面的斜率;  $b$  为偏移量,即超平面的截距;  $x$  为岩石物性数据值。求  $g(x)$  的过程就是求  $w$  和  $b$  (实数)两个参数的过程,但实际上只要求  $w$ , 求得以后找某些样本点代入就可以求得  $b$ 。因此,  $w$  才是变量。权重系数  $w$  由样本本身与样本所属类别有关,利用拉格朗日乘子对其进行表示:

$$w = \alpha_1 y_1 x_1 + \alpha_2 y_2 x_2 + \dots + \alpha_n y_n x_n \quad (2)$$

其中,  $y_i$  为识别标志(按岩性名称赋予不同的数值标签,如安山岩为 1,流纹岩为 2 等),  $x_i$  为已知物性数据。将该式代入(1),  $g(x)$  可表示为,

$$g(x) = \langle w, x \rangle + b = \left\langle \sum_{i=1}^n (\alpha_i y_i x_i), x \right\rangle + b \quad (3)$$

又因只有  $x_i$  和  $x$  为向量,将其余部分从内积中提出,得

$$g(x) = \sum_{i=1}^n \alpha_i y_i \langle x_i, x \rangle + b \quad (4)$$

已经证明,在这组超平面中,分类最优的超平面就是保证其与各类样本数据中离平面最近的向量与超平面之间的几何距离达到最大。这样,最优超平面的寻优过程可表达为,

$$\min \Phi(\omega) = \frac{1}{2} \|w\|^2, \text{ 限制条件为: } y_i (w \cdot x + b) \geq 1, i = 1, 2, \dots, n \quad (5)$$

限制条件是为了保证样本数据点不在超平面边界之内。可见, SVM 方法的核心思想是最大化分类边界, SVM 分类的目标是寻找划分特征空间的最优超平面,这个分类面指的是使不同属性样本得到间隔  $\gamma$  最大化的分界面,它应是一个既保证样本能得到正确分类,又保证分类具有较好泛化能力的分界面。

(2) 复杂岩性识别问题,即物性数据规律复杂,无法用简单的分类线(或平面)直接区别不同岩性。为解决这种问题,可将非线性分类样本数据进行归一化后,再将其从原来所在低维空间转换到一个更高维的特征空间中,即将原来线性不可分的问题变为了高维空间中的线性可分(见图 2)。将公式(1)表示为:

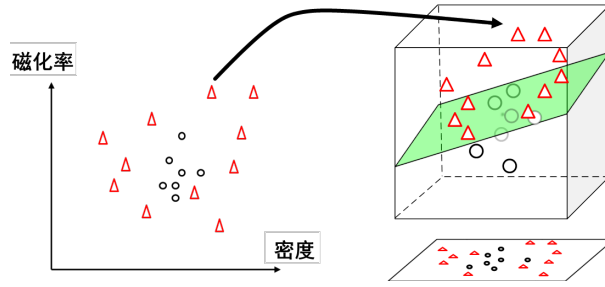


Figure 2. Map non-linear data to the high-dimensional feature space  
图 2. 将非线性数据映射至高维特征空间

$$g(x) = \sum_{i=1}^l \langle w_i, \phi_i(x) \rangle + b \quad (6)$$

其中， $\phi: X \rightarrow F$  为低维空间  $X$  到高维空间  $F$  的非线性映射描述。

同时，由线性学习机的决策规则的对偶形式，可将决策表示为由  $X$  与训练样本的内积表示：

$$g(x) = \sum_{i=1}^l \alpha_i \langle \phi(x_i), \phi(x) \rangle + b \quad (7)$$

在岩性识别可选取径向基核函数  $K = \exp\left(-\frac{\|x-x_i\|^2}{2p^2}\right)$  来代替公式(7)中的内积  $\langle \phi(x_i), \phi(x) \rangle$ ，实验证明核函数较好的解决具有复杂规律的岩性识别问题。

(3) 物性数据出现异常时。物性数据的采集不可避免的因人为或仪器因素出现误差，这将影响分类准则的判断。为了不改变最优分类超平面，SVM 引入了一个非负的松弛因子  $\zeta$  来允许有一定的训练误差。对于样本数据中可以进行线性分类的数据，为了降低松弛因子  $\zeta$  带来的增大分类间隔引入的容错代价，SVM 的优化目标函数引入了一个可变的惩罚因子  $C$ ，(5)式可表达为：

$$\min \Phi(\omega) = \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l \zeta_i, \text{ 限制条件: } y_i(w \cdot x + b) \geq 1 - \zeta_i, i = 1, 2, \dots, n \quad (8)$$

其中，松弛因子  $\zeta_i \geq 0$ ， $C$  为一个正则化参数，由于各类平滑训练误差与最大化分类间隔之间的矛盾。 $C$  越小，最大化分类间隔所占权重越大，从而带来的训练集合上的分类误差也就越大。反之， $C$  越大，由此带来的训练集上的误差就越小，但 SVM 超平面的推广能力就会有所降低，因此应根据实际训练岩石物性样本确定  $C$ ，通常采用交叉验证的方法进行计算。

### 3. 模型验证

现利用实测的流纹岩、安山岩、玄武岩三类岩性的 42 组岩石物性数据进行方法验证。每组数据包括密度、磁化率、电阻率、速度四种数据，同时定义流纹岩为第 1 类，安山岩为第 2 类，玄武岩为第 3 类。其中，各 32 组样本作为已知数据，另外各 10 组数据作为未知数据，即测试数据。支持向量机用于岩性识别的具体流程如下：

- (1) 在工区内选择已知岩石岩性的部分物性数据作为训练样本  $(x_i, y_i)$ ，其中  $y_i$  为不同岩性的类别标签， $x_i$  为物性数据(密度、磁化率、电阻率及速度值)；
- (2) 选择相应的核函数  $K$ ；
- (3) 根据分类识别的泛化目标需求选取惩罚因子  $C$ ；
- (4) 将低维不可分问题转化为二次寻优问题，求解二次寻优问题得到最优分类平面  $g(x)$ ，获取岩性

识别模型；

(5) 将待检测的物性数据作为输入变量，利用所构建的分类识别模型得到岩性识别结果。

交汇图的识别结果如图 3(a)，识别精度为 73.2%；SVM 分类结果如图 3(b)所示，识别精度为 91.3%，提高了 16.7%。可见，SVM 方法可有效的对岩石物性进行区分。

#### 4. 准噶尔西北缘岩石物性模板构建及应用

利用上述方法分区域对准噶尔盆地的岩石物性数据进行了岩性识别，并依次为依据构建了岩石物性模板。其中，图 4 展示出了准噶尔西北缘的岩石密度、磁化率、电阻率及速度模板。

现将将图 4 所示的岩石物性模板应用与西北缘车排子工区的凝灰岩岩性分类中。如图 5 中的测井与岩心数据所示，凝灰岩速度、密度分布范围广，与其它火成岩不易区分，而在磁化率数据来看，八个样本点中有五个是弱异常，有 3 个为中强磁异常。这种差别性使得凝灰岩岩性分类出现了疑问，此时，通过分析车排子地区凝灰岩物性模板，可看出凝灰岩密度、速度、电阻率取值范围广，相比其它岩性，其磁化率值较低，因此，磁化率可以作为区分凝灰岩与其它物性相近岩石的重要标志。再结合岩石物性模板进行分析，计算得到安山岩与凝灰岩磁化率统计值相差  $400 \pm 50(\times 10^{-5}SI)$ ，因此，基本断定图 5 中的三个样本点为安山岩类。

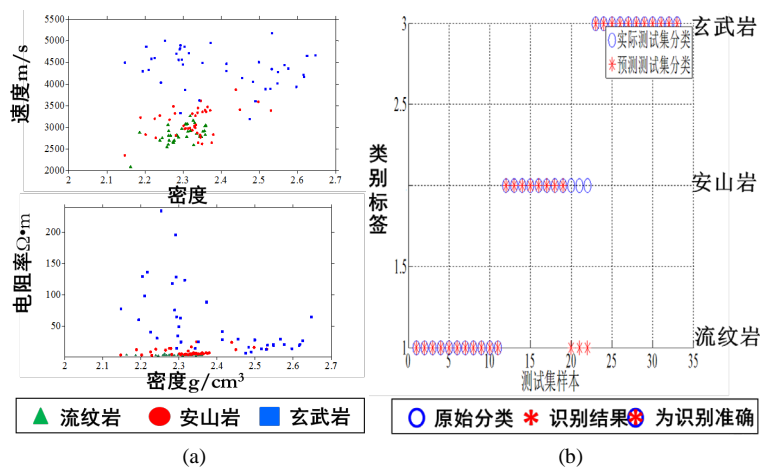
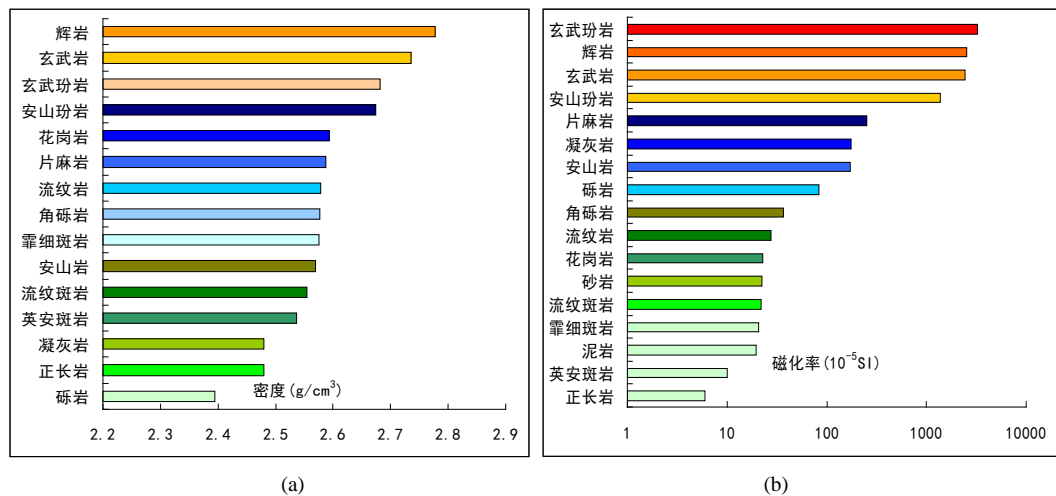


Figure 3. Comparison of igneous rocks lithology identification methods  
 图 3. 火成岩岩性识别方法对比



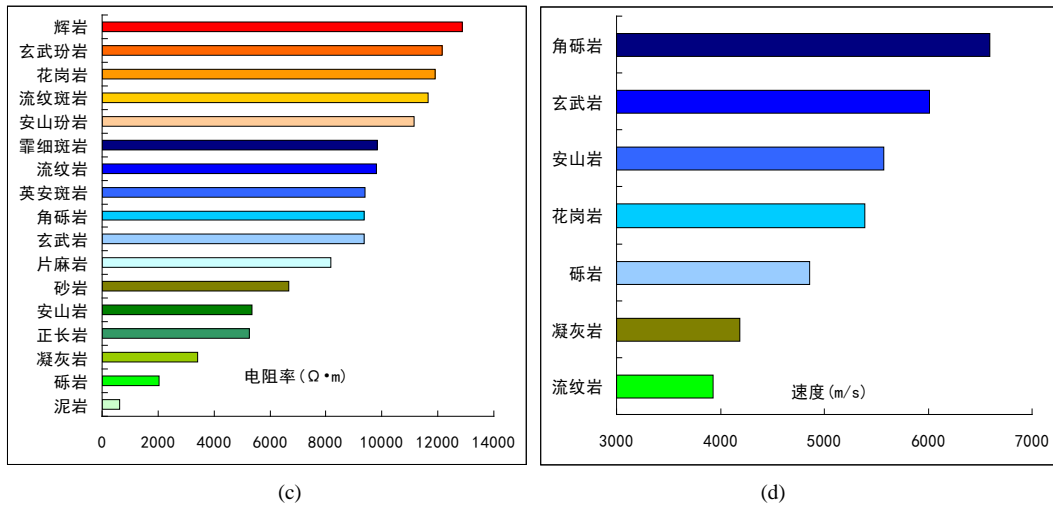


Figure 4. Rock physical property template of Northwestern Junggar basin  
图 4. 准噶尔盆地西北缘岩石物性模板

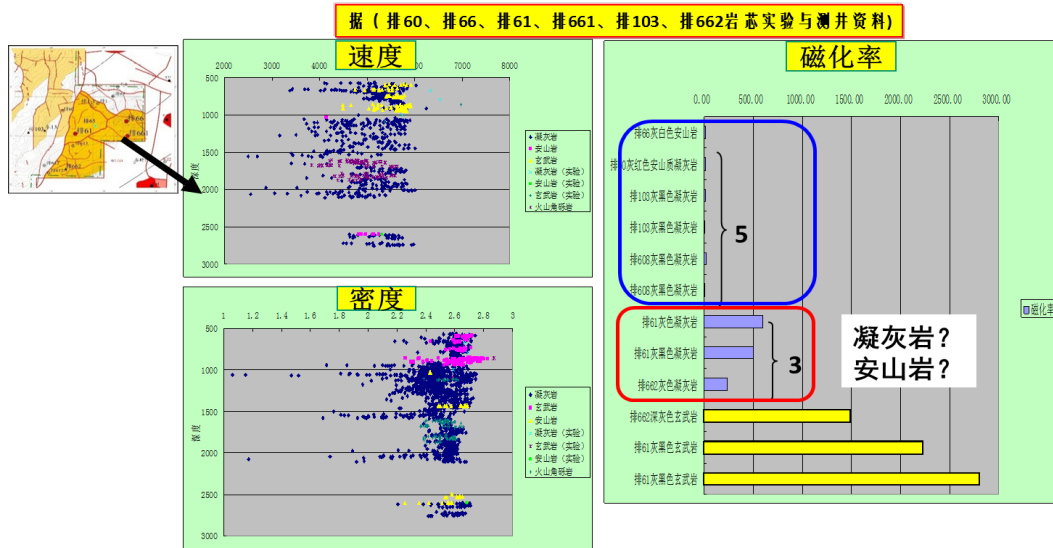


Figure 5. Physical characteristics of tuff of Chepaizi area  
图 5. 车排子地区凝灰岩物性特征

## 5. 结论

为了进一步提高重磁电震物性数据岩石岩性的识别能力，基于支持向量机理论提出了一种重磁电震物性岩性识别方法。选取了准噶尔盆地地区的重磁电震物性资料对该方法进行了验证，结果表明，相比交汇图等方法，本文提出的识别方法不仅可直接对高维数据进行分类，而且可降低人为主观影响，识别精度得到了明显提高。利用该方法对准噶尔盆地的重磁电震数据构建了准噶尔盆地的岩石物性模板。同时，利用该西北缘岩石物性模板分析了车排子地区的部分凝灰岩岩性进行了校正，为下一步开展类似落实岩性工作提供了有利的判别依据。

## 参考文献 (References)

- [1] 于鹏, 张罗磊, 王家林, 等. 重磁电震资料联合反演黔中隆起物性结构[J]. 同济大学学报(自然科学版), 2008,

36(3): 406-412

- [2] 谭绍泉. 乌伦古坳陷石炭系地层特征与勘探方向[J]. 石油天然气学报, 2012, 34(4): 65-68.
- [3] 杨辉, 文百红, 张研, 等. 准噶尔盆地火山岩油气藏分布规律及区带目标优选[J]. 石油勘探与开发, 2009, 36(4): 419-427.
- [4] Vapnik, V. (1995) *The Nature of Statistical Learning Theory*. Springer-Verlag, New York.  
<http://dx.doi.org/10.1007/978-1-4757-2440-0>
- [5] Vapnik, V., Golowich, S. and Smola, A. (1996) Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing. *Advances in Neural Information Processing Systems*, 281-287.
- [6] 蔡煜东. 应用人工神经网络方法预测油气[J]. 石油地球物理勘探, 1993, 28(5): 634-638
- [7] 许建华, 蔡瑞. 有监督 SOM 神经网络在油气预测中的应用[J]. 石油物探, 1998, 37(1): 71-76
- [8] 朱国强, 刘士荣, 俞金寿. 支持向量机及其在函数逼近中的应用[J]. 华东理工大学学报, 2002, 28(5): 555-559.
- [9] 张学工. 关于统计学习理论与支持向量机[J]. 自动化学报, 2000, 26(1): 32-42.
- [10] 胡国胜, 钱玲. 支持向量机的多分类算法[J]. 系统工程与电子技术, 2006, 28(1): 127-132.
- [11] 梅建新, 段汕. 支持向量机在小样本识别中的应用[J]. 武汉大学学报:自然科学版, 2002, 48(6): 732-734.