

DOI:10.3969/j.issn.1673-4785.201304019
网络出版地址: <http://www.cnki.net/kcms/doi/10.3969/j.issn.1673-4785.201304019.html>

基于支持向量机和有序聚类的岩层识别

张多, 韩逢庆
(重庆交通大学 管理学院, 重庆 400074)

摘 要: 由于支持向量机进行分类前需要先使用训练样本训练分类器, 而在岩层识别问题中没有训练样本, 针对此问题, 提出一种基于有序聚类的支持向量机岩层识别分类算法。首先利用有序聚类算法对经滤波和归一化后的测井数据进行初步分层, 然后根据初步分层结果获取训练样本, 最后用训练后的支持向量机分类器对测井数据进行第2次分层。应用该算法对选取的3口井的岩性进行自动识别, 并将该算法的识别结果与其他算法进行比较。仿真实验结果表明, 该算法具有较高的准确率, 每种岩层的平均准确率能达到85%, 解决了岩层识别前必须采用已知类别的数据对支持向量机进行训练的弊端。

关键词: 岩层识别; 支持向量机; 有序聚类; 训练样本; 分类器
中图分类号: TP631 **文献标志码:** A **文章编号:** 1673-4785(2014)01-0098-06

中文引用格式: 张多, 韩逢庆. 基于支持向量机和有序聚类的岩层识别[J]. 智能系统学报, 2014, 9(1): 98-103.
英文引用格式: ZHANG Duo, HAN Fengqing. Stratum identification based on the SVM and ordered cluster[J]. CAAI Transactions on Intelligent Systems, 2014, 9(1): 98-103.

Stratum identification based on the SVM and ordered cluster

ZHANG Duo, HAN Fengqing
(School of Management, Chongqing Jiaotong University, Chongqing 400074, China)

Abstract: The support vector machine (SVM) needs training samples to train itself before identifying stratum, while there are no training samples with stratum identification. Focusing on this problem, this paper puts forward a vector machine classifier based on the ordered clustering algorithm. Firstly, the ordered clustering algorithm is used to get preliminary layered logging data which have been filtered and normalized. Secondly, the training samples are obtained according to preliminary layered outcomes. Finally, the data are layered again by the trained SVM classifier. The algorithm is used to automatically identify the lithology of the selected three wells, and compared with the results of the other algorithms. The results of the simulation experiment show that the algorithm overcomes the drawbacks that the labeled data has to adopt when training SVM, and improves the accuracy of each stratum, reaching 85% on average.

Keywords: stratum identification; support vector machine; ordered clustering; training samples; classifier

岩层识别问题是地球物理勘探的重要课题之一, 准确识别岩层对于地球物理勘探十分重要。较长时间以来, 测井分析工作者主要根据测井曲线的形态变化特征及其与周围岩层之间的差异特征进行人工分层。这种十分原始的手工分层方法, 会由于不同测井分析工作者在使用分层曲线和掌握分层标准的不同造成分层结果的差异^[1]。相对于人工分层, 自动分层可以避免人为分层的随意性, 并在很大程度上提高工作效率^[2]。

自从1982年Wolf等^[3]首次根据测井数据自动判定地层岩性以来, 目前已发展了很多分层方法, 如有序聚类法^[4]、极值方差聚类法^[5]、支持向量机法^[6-8]等。有序聚类法和极值方差聚类法是在没有先验类别条件下, 通过测井数据内在的一些属性和联系, 对岩层进行自动分层。这些方法在数学上比较严格, 其分层结果是完美的数学结果, 但并不能完全反映出测井数据与岩层类别之间的对应关系。

支持向量机(support vector machine, SVM)是Vapnik等^[9]提出的一种基于结构风险最小化原理(structural risk minimization, SRM)和统计学习理论的VC维理论的机器学习方法, 能够充分挖掘事物

之间的对应关系。其根据有限的样本信息在模型的复杂性(即对特定训练样本的学习精度)和学习能力(即无错误识别任意样本的能力)之间寻求最佳折衷,以获得最好的分类能力^[10]。目前,该算法被广泛应用于文本分类、图像分类、基因分析、字符识别、人脸识别等领域^[11]。相对于许多传统的分类器,如参数估计和神经网络,它不仅有完善的理论支持,而且表现出良好的分类性能和推广能力,是一种专门针对有限样本分类的方法。SVM 在很大程度上解决了非线性、高维数以及局部极值等问题^[12]。近年来,更多的研究者开始致力于支持向量机对岩层识别的研究。宋延杰等^[6]选取大庆油田中 7 口井的探井曲线资料作为样本数据,并依据先验经验,人为选取部分样本数据作为训练集,然后利用学习后的支持向量机预测岩层。赵军等^[7]利用 P 油田 18 口井的解释结果,筛选出训练集和测试集,验证支持向量机在岩层识别中的有效性。文政等^[8]利用同地区 5 口井的样本数据对支持向量机进行训练,并对剩余样本进行预测验证。然而,上述研究没有给出提取训练样本的具体方法,训练样本的选取依赖于先验经验或周边已知分层情况井的数据样本。为了克服支持向量机方法在岩层识别问题中的这种弊端,本文提出了一种基于有序聚类的支持向量机算法对岩层进行识别。

1 相关理论

1.1 支持向量机

支持向量机的基本思想是寻找一个超平面,将属于 2 个不同类别的样本无误地分开,且分类间隙要最大。对于非线性问题,可以通过非线性映射 $\varphi: \mathbf{R}^N \rightarrow \mathbf{H}$ 把数据从原空间 (\mathbf{R}^N) 映射到某个高维空间 (\mathbf{H}) 里,在变换空间求最优分类面。

对于非线性样本集:
$$T = \{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)\}$$
式中: $\mathbf{x}_i \in \mathbf{R}^N$ 为 N 维向量, $y_i \in \{1, -1\}$ 。在高维空间中,则存在最优分类面:

$$(\varphi(\mathbf{x}), \mathbf{w}) + b = 0$$

满足条件:

$$\min \varphi(\mathbf{w}) = \frac{1}{2}(\mathbf{w} \cdot \mathbf{w})$$
$$y_i(\varphi(\mathbf{x}_i), \mathbf{w}) + b \geq 1, i = 1, 2, \dots, n \quad (1)$$

式(1)是凸二次优化问题,引入拉格朗日函数:

$$L(\mathbf{w}, b, \mathbf{a}) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n a_i [y_i(\mathbf{w} \cdot \varphi(\mathbf{x}_i) + b) - 1]$$

式中: $a_i \geq 0$ 为拉格朗日乘子。为求 $L(\mathbf{w}, b, \mathbf{a})$ 的最小值,分别对 \mathbf{w} 、 b 、 \mathbf{a} 求偏导,得

$$\begin{cases} \frac{\partial L}{\partial \mathbf{w}} = 0 \Rightarrow \mathbf{w} = \sum_{i=1}^n a_i y_i \varphi(\mathbf{x}_i) \\ \frac{\partial L}{\partial b} = 0 \Rightarrow \sum_{i=1}^n a_i y_i = 0 \\ \frac{\partial L}{\partial a_i} = 0 \Rightarrow a_i [y_i(\mathbf{w} \cdot \varphi(\mathbf{x}_i) + b) - 1] = 0 \end{cases} \quad (2)$$

根据式(1)和式(2)的约束条件,可转化成对偶问题为

$$\begin{cases} \max \sum_{i=1}^n a_i - \frac{1}{2} \sum_{i,j=1}^n a_i a_j y_i y_j K(\mathbf{x}_i \cdot \mathbf{x}_j), \\ a_i \geq 0, i = 1, 2, \dots, n \\ \sum_{i=1}^n a_i y_i = 0 \end{cases}$$

这是一个二次函数极值问题,故存在惟一解。若 a_i^* 为最优解,则有

$$\mathbf{w}^* = \sum_{i=1}^n a_i^* y_i \varphi(\mathbf{x}_i)$$

式中: a_i^* 是不为零的样本,称为支持向量。 b^* 是分类阈值,可由约束条件 $a_i [y_i(\mathbf{w} \cdot \varphi(\mathbf{x}_i) + b) - 1] = 0$ 求解。

解得上述问题后可得到二分类最优分类函数为

$$f(\mathbf{x}) = \text{sgn} \left(\sum_{i=1}^n a_i^* y_i K(\mathbf{x}_i, \mathbf{x}) + b^* \right)$$

式中: $K(\mathbf{x}_i, \mathbf{x})$ 为核函数。

1.2 有序聚类算法

有序聚类算法是多元统计分析中针对有序样本的一种统计分类方法。其基本思想是:首先将待分类的 n 个样本看作 1 类,然后根据离差平方和类内最小以及类间最大准则分为 2 类、3 类……一直到所需的 k 类为止。

假设每个样本有 m 个特征指标,则 n 个样本形成的数据矩阵如下:

$$\mathbf{X} = (x_{ij})_{n \times m} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1m} \\ x_{21} & x_{22} & \cdots & x_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nm} \end{bmatrix}$$

式中:元素 x_{ij} 表示第 i 个样本第 j 个特征指标值。首先利用这些特征指标值计算层内变差矩阵 $\mathbf{D} = (d_{ij})_{n \times n}$, 其中:

$$d_{ij} = \sum_{\alpha=i}^j \sum_{\beta}^m [x_{\alpha\beta} - \bar{x}_{\beta}(i, j)]^2, 1 \leq i \leq j \leq n$$
$$\bar{x}_{\beta}(i, j) = [1/(j - i + 1)] \sum_{\alpha=i}^j x_{\alpha\beta}, \beta = 1, 2, \dots, m$$

然后记 $b(n, k)$ 是有序样本分为 k 层的某一种方法,则其层内离差平方和为 $L[b(n, k)]$ 。当 $L[b(n, k)]$ 越小,即各层间的离差平方和越小,分层就越合理。要使 $L[b(n, k)]$ 达到极小值的分法,

就是寻求最佳分点 $(j_k, j_{k-1}, \cdots, j_2)$ 。记 $P(n, k)$ 为将有序样本分为 k 层的最佳分法, 为得到最佳分层方法, 利用递推公式:

$$\begin{aligned} L[P(n, 2)] &= \min_{2 \leq j \leq n} (d(1, j-1) + d(j, n)) \quad (3) \\ L[P(n, k)] &= \min_{k \leq j \leq n} (L[P(j-1, k-1)] + d(j, n)) \quad (4) \end{aligned}$$

由递推公式可以看出, 要求得最佳分法 $P(n, k)$, 需先找第 k 层最佳分点 j_k , 使 $L[P(n, k)]$ 达到最小。

$$\begin{aligned} L[P(n, k)] &= L[P(j_k-1, k-1)] + d(j_k, n) \\ \text{最后, 利用类似的方法依次可得到最佳分点} & j_{k-1}, j_{k-2}, \cdots, j_2。 \end{aligned}$$

2 岩层自动分层

基于上述理论, 本文提出了一种基于有序聚类的支持向量机分类算法, 首先对测井数据进行预处理: 异常数据的处理、数据的滤波, 其次选取测井曲线, 并对其进行归一化处理, 利用有序聚类算法初步分层, 获取训练样本, 最后利用训练后的支持向量机进行预测, 实现岩层的自动分层。

2.1 处理异常数据

在对岩层进行探测时, 由于非岩层因素引起的干扰, 会出现异常数据, 即有的数据比相应测井曲线数据的平均值大很多或者小很多。异常数据不仅干扰有效数据, 影响后面的数据处理, 如归一化等, 而且对后面岩层自动分层的精确度产生影响, 故采用邻近数据平均值代替异常数据^[13]。

2.2 平滑滤波

经过异常数据处理后的测井曲线, 从数字信号处理角度来看, 由于自身携带了大量的噪声干扰信号, 故测井曲线出现了毛刺, 为了消除测井曲线的毛刺干扰, 对测井曲线进行中值滤波。

假设测井数据序列: x_1, x_2, \cdots, x_N , 选取窗口长度为 $2n + 1$ 的滤波器对该序列进行滤波, 则步骤为如下:

- 1) 选出以 x_i 为中心的 $2n + 1$ 个数据 $x_{i-n}, x_{i-n+1}, \cdots, x_i, \cdots, x_{i+n-1}, x_{i+n}$, 对这 $2n + 1$ 个数据排序;
- 2) 选取排序后的中间项作为 x_i 的滤波值;
- 3) 自上而下迭代计算测井数据的滤波值。

2.3 选取测井曲线

由于测井数据是按井深的递进变化对应着多种测井曲线的。1) 由于测井曲线过多, 导致分析问题的复杂度变大; 2) 实际上有的测井曲线之间相关性较大, 故可对所有的测井曲线做相关性分析, 找出高度相关的测井曲线, 从这些高度相关的测井曲线中选择一种。相关性系数与相关程度对应关系如表 1。

表 1 相关系数与相关程度对应关系
Table 1 Correlation coefficient and degree of correlation

相关系数	相关程度
0.00~±0.30	微相关
±0.30~±0.50	实相关
±0.50~±0.80	显著相关
±0.80~±1.00	高度相关

2.4 归一化处理

由于不同测井曲线的刻度和量纲各不相同, 在采用多种曲线进行自动分层时, 有必要将测井曲线归一化到 $[0, 1]$ 内, 以消除因刻度和量纲所产生的影响。归一化公式为

$$y_i = \frac{x_i - x_{\min}}{x_{\max} - x_{\min}}$$

式中: x_i 为测井曲线上对应的测井深度 i 的测井值, y_i 为 x_i 的归一化值, x_{\max} 和 x_{\min} 分别为测井曲线中的最大值和最小值。

2.5 利用有序聚类初分层

根据上述有序聚类算法, 对岩层进行初步分层, 首先利用测井数据值计算层内变差矩阵 D , 然后根据递推公式(3)、(4), 可逐步求得岩层的最佳分点, 即可获得岩层的初步分层结果 $\{j_k, j_{k-1}, \cdots, j_2\}$ 。

2.6 利用支持向量机二次分层

由于有序聚类算法所要求的分界点只满足层内离差平方和最小以及层间离差平方和最大, 而没有完全反映出测井数据与岩层类别的对应关系, 故利用支持向量机对岩层进行二次分层, 其步骤如下:

1) 获取训练样本: 根据有序聚类初步分层结果, 选取各层质心离差平方和最小的连续 30 个数据样本作为各层的训练样本。

2) 训练支持向量机: 选取“一对一”多类分类器作为支持向量机分类器, 选取 RBF 核函数作为支持向量机的核函数^[14-15], 利用有序聚类算法获得的训练样本来训练支持向量机, 最终获得多类分类器:

$$f_l(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1}^n a_i^* y_i K(\mathbf{x}_i, \mathbf{x}) + b^*\right), l = 1, 2, \cdots, \frac{k(k-1)}{2}$$

3) 对岩层进行分层: 利用训练后的支持向量机分类器去预测岩层类别。

3 实验结果与分析

在 CPU 为 2.53GHz、内存为 2GB、操作系统为 Windows7 的 PC 机上, 运用 MATLAB7.1 软件进行数字仿真。仿真实验所采用的测井数据来源于中国数学建模网, 选取其中 3 口井进行预测。其中每口井

各有 66 项指标。利用相关性分析方法对所有影响岩性的测井曲线进行筛选,最终选取 10 种测井曲线来指示岩性,选取结果见表 2。

表 2 测井曲线的选取

Table 2 Selection of logging curves

序号	测井曲线名称
1	密度测井曲线(DEN)
2	4M 电阻率测井曲线(R4.0)
3	自然电位测井曲线(SP)
4	井斜测井曲线(DEVi)
5	井斜方位测井曲线(AZIm)
6	自然伽马测井曲线(GR)
7	声波测井曲线(AC)
8	微侧向电阻率测井曲线(RML)
9	井径测井曲线(CAL)
10	电阻率测井曲线(RT)

1~3 号井各有 6 种岩层,从一定深度开始,对各口井进行岩层划分和命名,依次为岩 1 层,岩 2 层,⋯,岩 6 层。其中 1 号井的井深范围为 294.0~530.2 m,2 号井的井深范围为 257.0~297.5 m,3 号井的井深范围为 249.4~496.0 m,各口井中每米测试 8 个点,且已知各口井的实际分层情况。根据前述的岩层自动分层步骤,对 1~3 号井的实际测井数据利用新算法进行岩层的自动分层。同时,本文与文献[8]的算法做对比,以验证新算法的有效性。实验结果如图 1、图 2 和表 3 所示(本文提出的新算法简记为算法 1,文献[8]的算法简记为算法 2)。

从图 1 和图 2 可以直观地看出,新算法分层结果更接近于实际分层情况,即新算法优于文献[8]的算法。同时,从表 3 可以看出,利用新算法进行岩层识别,每种岩层的平均准确率能达到 85%,而利用文献[8]的算法进行岩层识别,其平均准确率只能达到 70%。主要原因在于文献[8]采用同地区其他井中已知岩层类别的测井数据作为训练集,虽然同地区相邻井在某些方面具有相同的属性和联系,但并不能完全反映出待预测井中每种岩层的特性。相对于文献[8]提出的算法,新算法主要优势在于利用有序聚类算法进行初步分层,然后根据初步分层结果来选取训练样本,所选取的训练样本能够很好地反映出各口井中岩层的特性,这样经测井数据训练后的支持向量机才具有更好的分类能力,分层准确率才会更高。

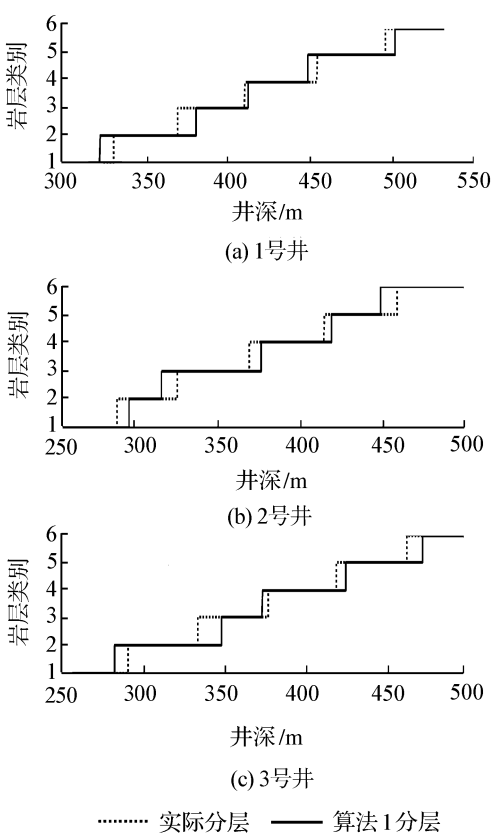


图 1 算法 1 分层结果与实际分层结果对比
Fig.1 Comparison of layered results of algorithm 1 and actual situation

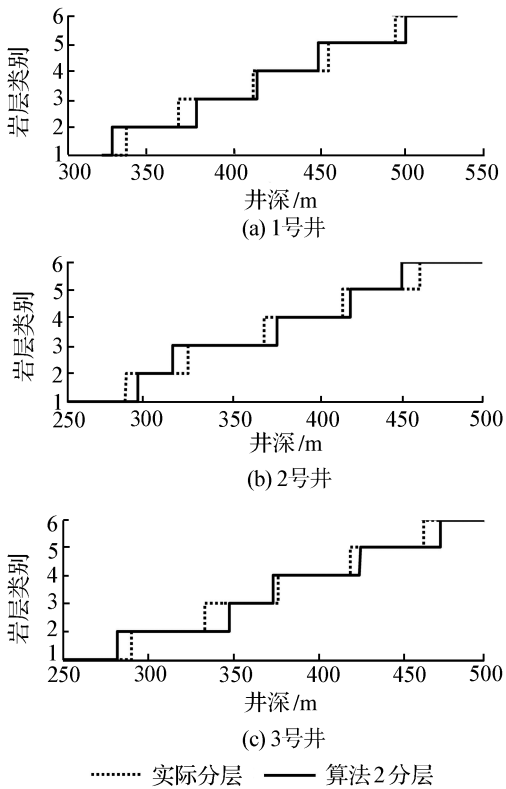


图 2 算法 2 分层结果与实际分层结果对比
Fig.2 Comparison of layered results of algorithm 2 and actual situation

表 3 算法 1 与算法 2 分层准确率比较

Table 3 Comparison of the accuracy of algorithm 1 and algorithm 2

%

预测 井号	岩 1 层		岩 2 层		岩 3 层		岩 4 层		岩 5 层		岩 6 层	
	算法 1	算法 2	算法 1	算法 2	算法 1	算法 2	算法 1	算法 2	算法 1	算法 2	算法 1	算法 2
1 号井	77.8	63.8	100.0	74.1	75.0	60.7	82.6	70.2	100.0	75.6	83.9	68.7
2 号井	100.0	77.2	64.5	59.3	100.0	73.4	83.9	76.9	66.2	61.2	100.0	78.5
3 号井	79.4	66.8	100.0	75.6	68.7	63.2	100.0	79.8	86.9	72.5	71.8	59.6
平 均	85.7	69.3	84.8	69.7	77.9	65.8	88.8	75.6	84.4	69.8	85.2	69.0

4 结 束 语

岩层识别是地球物理勘探的基础性工作,岩层识别的准确率将直接影响后续工作的有效开展,本文提出了一种基于有序聚类的支持向量机岩层识别算法。首先,该算法首先利用有序聚类进行初步分层;其次,根据初步分层结果提取出支持向量机所需的训练样本;最后,利用训练后的支持向量机进行第 2 次分层,得出相应的分类结果。该算法不同于常规的支持向量机算法,它不需要已知岩层类别的测井数据,也就是说能够间接通过有序聚类对未知岩层类别的测井数据进行初步识别,提高识别精度。仿真实验结果表明,该算法不仅很好地解决了已有算法必须依赖已知岩层类别的测井数据的弊端,而且具有较高的准确率,每种岩层的平均准确率为 85%,明显高于以往研究的 70%。然而,该算法仍具有局限性,如岩层分层数需要人为设定,以及如何选取核函数等,这都将是下一步的研究方向。

参 考 文 献:

[1] 林海燕,戴云,肖慈珣.一种基于沃希变换的测井自动分层方法[J].成都理工学院学报,1999,26(1):52-57.
LIN Haiyan, DAI Yun, XIAO Cixun. A well-log automated identification of rock boundaries method based on Walsh transform[J]. Journal of Chengdu University of Technology, 1999, 26(1): 52-57.

[2] 肖波,韩学辉,周开金.测井曲线自动分层方法回归与展望[J].地球物理学进展,2010,25(5):1802-1810.
XIAO Bo, HAN Xuehui, ZHOU Kaijin. A review and outlook of automatic zonation methods of well log[J]. Progress in Geophysics, 2010, 25(5): 1802-1810.

[3] WOLF M, PELISSIER-COMBESCURE J. Faciolog automatic electrofacies determination [C]//SPWLA 23rd Annual Logging Symposium. Corpus Christi, USA, 1982: 6-9.

[4] HAWKINS D M, MERRIAM D F. Optimal zonation of digitized sequential data[J]. Mathematical Geology, 1973, 5

(4): 389-396.

[5] 张明玉.极值方差聚类法在测井分层取值中的应用[J].新疆石油地质,2002,23(5):429-431.
ZHANG Mingyu. Application of minimax variance cluster analysis method in well log layering[J]. Xinjiang Petroleum Geology, 2002, 23(5): 429-431.

[6] 宋延杰,张剑凤,闫伟林,等.基于支持向量机的复杂岩性测井识别方法[J].大庆石油学院学报,2007,31(5):18-20.
SONG Yanjie, ZHANG Jianfeng, YAN Weilin, et al. A new identification method for complex lithology with support vector machine[J]. Journal of Daqing Petroleum Institute, 2007, 31(5): 18-20.

[7] 赵军,程鹏飞,刘地渊,等.支持向量机在水淹层测井识别的应用[J].物探与化探,2008,32(6):652-655.
ZHAO Jun, CHENG Pengfei, LIU Diyuan, et al. The application of the support vector machine to the recognition of flooding formation[J]. Geophysical and Geochemical Exploration, 2008, 32(6): 652-655.

[8] 文政,高松洋,毕广武.支持向量机在复杂岩性测井识别中的应用[J].大庆石油地质与开发,2009,28(1):134-137.
WEN Zheng, GAO Songyang, BI Guangwu. Application of support vector machine (SVM) in complex lithology identification by well logging[J]. Petroleum Geology and Oilfield Development in Daqing, 2009, 28(1): 134-137.

[9] VAPNIK V. The nature of statistical learning theory[M]. New York, USA: Springer-Verlag, 1995.

[10] 刘跃辉,郑建东.基于支持向量机原理的复杂地层岩性识别方法[J].国外测井技术,2011(4):22-26.
LIU Yuehui, ZHENG Jiandong. The lithology identification method based on support vector machine principle [J]. World Well Logging Technology, 2011(4): 22-26.

[11] 李茂宽,关键.基于模糊 C 均值的支持向量机数据分类识别[J].系统仿真学报,2005,17(7):1785-1787.
LI Maokuan, GUAN Jian. Data classification and recognition of support vector machines based on fuzzy C-means clustering[J]. Journal of System Simulation, 2005, 17

(7): 1785-1787.

[12]柯永振,张加万,孙济洲,等.结合支持向量机与 C 均值聚类的图像分割[J].计算机应用,2006,26(9):2081-2083.

KE Yongzhen, ZHANG Jiawan, SUN Jizhou, et al. Image segmentation combining support vector machines with C-means[J]. Journal of Computer Applications, 2006, 26(9): 2081-2083.

[13]万应明,高峻,董建平,等.多测井曲线合成应用方法初探[J].石油物探,2005,44(1):71-75.

WAN Yingming, GAO Jun, DONG Jianping, et al. The primary discussion to the application of synthetic multiple-logging[J]. Geophysical Prospecting for Petroleum, 2005, 44(1): 71-75.

[14]赵磊,李国和,马现峰.基于支持向量机的地层识别研究[J].计算机工程与应用,2006,35(1):230-232.

ZHAO Lei, LI Guohe, MA Xianfeng. Research of stratum recognition based on support vector machine[J]. Computer Engineering and Applications, 2006, 35(1): 230-232.

[15]郑延斌,李国和.支持向量机在地层识别中的应用[J].河南师范大学学报:自然科学版,2009,37(2):37-39.
ZHENG Yanbin, LI Guohe. Application of support vector machine to stratum recognition[J]. Journal of Henan Normal University: Natural Science, 2009, 37(2): 37-39.

作者简介:



张多,女,1988年生,硕士研究生,主要研究方向为机器学习、人工智能。



韩逢庆,男,1968年生,重庆市工业与应用数学学会理事,重庆市运筹学会理事。主要研究方向为机器学习、人工智能、小波理论及应用等,发表学术论文30余篇,其中被SCI、EI、ISTP检索20余篇。

第 10 届全国机器翻译研讨会

The 10th China Workshop on Machine Translation (CWMT 2014)

第 10 届全国机器翻译研会(CWMT 2014)于 2014 年 11 月在澳门大学举行。

全国机器翻译研讨会自 2005 年召开第一届以来,已连续成功组织召开了 9 届,共组织过 5 次机器翻译评测(2007、2008、2009、2011、2013),1 次开源系统模块开发(2006)和 2 次战略研讨(2010、2012),这些活动对于推动我国机器翻译技术的研究和开发产生了积极而深远的影响。因此,CWMT 已经成为我国自然语言处理领域颇具影响的学术活动。

2005 年由中科院自动化所、计算所和厦门大学联合发起并组织了第 1 届统计机器翻译技术评测及学术研讨会,会议在厦门大学成功举办。随后,会议由中科院计算所、自动化所、软件所、哈尔滨工业大学和厦门大学联合组织。2006—2013 年分别在中科院计算所、哈尔滨工业大学、中科院自动化所、南京大学、中科院软件所、厦门大学、西安理工大学、昆明理工大学成功召开。

CWMT 2014 旨在为国内外机器翻译界同行提供一个交互平台,加强国内外同行的学术交流,召集各路专家学者针对机器翻译的理论方法、应用技术和评测活动等若干基本关键问题进行深入的研讨,为促进中国机器翻译事业的发展,起到积极的推动作用。有别于以往,本次会议特设 2 个专题讲座,将和与会者互动探讨机器翻译最炽热的研究论点,而大会专题论坛汇集顶尖领域专家向大家揭示机器翻译最前沿的蓝图,本会还邀请了国内外知名专家学者做相关的特邀报告。会议期间将选出最佳学生论文,并颁发证书和奖金。另外,专为企业、用户、开发者而设的系统演示环节能让与会者展示及分享宝贵经验。

我们热诚地期待所有对机器翻译感兴趣的学者、企业界人士和学生积极参加本届全国机器翻译研讨会,并欢迎任何单位和个人到会展示您的机器翻译系统。感谢大家对本届会议的关注和支持!

会议网站:<http://www.cis.umac.mo/cwmt2014/ch/index.html>