

基于LS-SVM的回采工作面瓦斯涌出量预测

孙林¹, 杨世元²

(1. 合肥工业大学 应用物理系, 安徽 合肥 230009; 2. 合肥工业大学 仪器科学与光电工程学院, 安徽 合肥 230009)

摘要: 提出了一种基于最小二乘支持向量机 (LS-SVM) 的回采工作面瓦斯涌出量预测新方法。一方面, 该方法基于结构风险最小化, 能较好地解决小样本学习问题, 避免了人工神经网络等智能方法在对回采工作面瓦斯涌出量进行预测时所表现出来的过学习、泛化能力弱等缺点; 另一方面, 该方法用等式约束代替不等式约束, 降低了计算的复杂性, 使得预测容易实现。实验表明, 该方法具有预测精度高、速度快、容易实现等优点, 适合对回采工作面瓦斯涌出量的预测。

关键词: 最小二乘支持向量机; 回采工作面; 瓦斯涌出量; 预测

中图分类号: TD712.5

文献标识码: A

Prediction for gas emission quantity of the working face based on LS-SVM

SUN Lin¹, YANG Shi-yuan²

(1. Department of Applied Physics, Hefei University of Technology, Hefei 230009, China; 2. School of Instrument Science and Opto-electronic Engineering, Hefei University of Technology, Hefei 230009, China)

Abstract: Put forward a new prediction method for gas emission quantity of working face based on least squares support vector machine (LS-SVM). It can solve the small-batch learning better and avoid such disadvantages as over-training, weak normalization capability, ect., which artificial neural networks prediction has, because the latter is based on structure risk minimization. This proposed method is simpler and can be realized easily, because it uses equality restriction instead of inequality restriction. Experiments prove that it offers an effective method for predicting gas emission quantity of working face.

Key words: least squares support vector machine; working face; gas emission quantity; prediction

回采工作面瓦斯涌出量是矿井通风设计、瓦斯防治与管理的重要依据^[1], 所以, 对回采工作面瓦斯来源分析以及瓦斯涌出量的预测变得尤为重要。根据瓦斯分源预测理论, 回采工作面瓦斯涌出有开采煤层、邻近煤层和采空区3种来源^[2]。影响开采层瓦斯涌出量的因素有开采层瓦斯原始含量、煤层厚度、煤层倾角、煤层埋藏深度、工作面长度、推进速度、开采厚度及工作面采出率; 影响邻近层涌出量的因素有邻近层瓦斯原始含量、邻近层厚度、层间距离、层间岩性、煤层倾角、推进速度、开采厚度、工作面长度、顶板管理方式; 影响开采层瓦斯涌出量的因素有煤层瓦斯原始含量、煤层厚度、采出率、开采强度^[3]。煤矿回采工作面瓦斯涌出为一动态变化系统, 受各种复杂因素影响^[4], 要达到准确预测瓦斯涌出量的目的, 必须综合考虑各复杂因素间的不确定关系。支持向量机 (Support vector machine, SVM) 是 Vapnik 等人提出的一种新的通用学习方法, 较好地解决了小样本、非线性、高维数和局部极小点等实际问题^[5-6], 已成为机器学习新的研究热点之一, 并已成功地应用于模式识别 (聚类)、函数逼近和预测等方

面. 瓦斯涌出量预测可以看作是涌出量与其对应影响因素之间复杂的非线性函数关系的逼近问题, 基于此考虑, 笔者尝试将支持向量机理论引入瓦斯涌出量预测问题中.

1 SVM 回归算法

1.1 标准 SVM 回归算法

SVM 的最初研究是针对线性样本点的线性函数拟合问题, 设所有训练数据 (x_i, y_i) , $i = 1, 2, \dots, n$, $\mathbf{x} \in R^l$, $\mathbf{y} \in R$ 能在精度 ε 下无误差地用线性函数 $\mathbf{y} = \boldsymbol{\omega}^T \mathbf{x} + b$ 拟合, 即有

$$\begin{cases} \boldsymbol{\omega}^T \mathbf{x} + b - y_i \leq \varepsilon, \\ y_i - \boldsymbol{\omega}^T \mathbf{x} - b \leq \varepsilon, \end{cases}$$

其中, $\mathbf{x}$ 为输入向量; $\mathbf{y}$ 为输出向量; $\boldsymbol{\omega}$ 为拟合函数法线方向; b 为阈值; ε 为拟合精度. 根据结构风险最小化准则, 这一优化目标在最小化 $\|\boldsymbol{\omega}\|^2/2$ 时可取得较好的推广能力. 考虑到实际应用中允许样本存在逼近误差 ξ 的情况, 则支持向量机优化目标可以表示为

$$\min \frac{1}{2} \boldsymbol{\omega}^T \boldsymbol{\omega} + C \sum_{i=1}^n (\xi_i + \xi_i^*), \quad \text{s. t.} \begin{cases} y_i - \boldsymbol{\omega}^T \mathbf{x}_i - b \leq \varepsilon + \xi_i, \\ \boldsymbol{\omega}^T \mathbf{x}_i + b - y_i \leq \varepsilon + \xi_i^*, \end{cases}$$

其中, $\xi_i, \xi_i^* \geq 0$ 为实际应用中允许样本存在的逼近误差; $C > 0$ 被称为平衡因子, 用来控制对超出误差样本的惩罚程度. 对于上式通过建立 Lagrange 方程来求解, 有

$$\begin{aligned} L(\boldsymbol{\omega}, b, \xi_i, \xi_i^*, a_i, a_i^*, \eta_i, \eta_i^*) = & \frac{1}{2} \boldsymbol{\omega}^T \boldsymbol{\omega} + C \sum_{i=1}^n (\xi_i + \xi_i^*) - \sum_{i=1}^n \alpha_i (\varepsilon + \xi_i - y_i + \boldsymbol{\omega}^T \mathbf{x}_i + b) - \\ & \sum_{i=1}^n \alpha_i^* (\varepsilon + \xi_i^* + y_i - \boldsymbol{\omega}^T \mathbf{x}_i - b) - \sum_{i=1}^n (\eta_i \xi_i + \eta_i^* \xi_i^*), \end{aligned}$$

其中, 参数 $\alpha_i, \alpha_i^* \geq 0$ 被称为 Lagrange 乘子, $\eta_i, \eta_i^* \geq 0$ 为引进的临时变量. 求解方程在 $\boldsymbol{\omega}, b, \xi_i, \xi_i^*$ 处的极小值并根据 KKT 条件可得 SVM 回归算法, 即

$$f(x) = \boldsymbol{\omega}^T \mathbf{x} + b = \sum_{i=1}^n (\alpha_i - \alpha_i^*) (\mathbf{x}_i^T \cdot \mathbf{x}) + b = \sum_{\mathbf{x}_i \in SVs} (\alpha_i^* - \alpha_i) (\mathbf{x}_i^T \cdot \mathbf{x}) + b.$$

对于非线性回归, 通过非线性变换 $\mathbf{x} \rightarrow \boldsymbol{\varphi}(\mathbf{x})$ 将原输入空间的回归问题映射到高维特征空间 (Hilbert 空间) 中, 然后在 Hilbert 空间进行线性回归, 即 $f(x) = \boldsymbol{\omega}^T \boldsymbol{\varphi}(x) + b$, 从而取得在原空间非线性回归的效果, 引入核函数 $k(x_i, x_j)$, 可得回归算法

$$f(x) = \boldsymbol{\omega}^T \boldsymbol{\varphi}(x) + b = \sum_{i=1}^n (\alpha_i - \alpha_i^*) k(x_i, x) + b = \sum_{\mathbf{x}_i \in SVs} (\alpha_i^* - \alpha_i) k(x_i, x) + b.$$

1.2 LS-SVM 回归算法

LS-SVM 是标准 SVM 的一种扩展, 它选择误差 ξ_i 的二次范数作为损失函数, 如此优化问题^[7]则变为

$$\min \frac{1}{2} \boldsymbol{\omega}^T \boldsymbol{\omega} + \frac{1}{2} \gamma \sum_{i=1}^n \xi_i^2, \quad \text{s. t. } y_i = \boldsymbol{\omega}^T \boldsymbol{\varphi}(x_i) + b + \xi_i.$$

其中, γ 为调节常数, 它能够在训练误差和模型复杂度之间取一个折衷以便使所求的函数具有较好的泛化能力. LS-SVM 定义了与标准 SVM 不同的损失函数, 并将其不等式约束改为等式约束. 最优的 α, b 可以通过 KKT 条件获得, 求取参数 α 和 b 后, LS-SVM 回归模型即为

$$y(x) = \sum_{i=1}^n \alpha_i k(x_i, x) + b.$$

2 瓦斯涌出量 SVM 预测法在 Matlab 中的实现

用于回采工作面瓦斯涌出量预测的实验数据来自参考文献[3], 根据回采工作面分源瓦斯涌出量及其各个影响因素的指标形成 3 个样本集, 分别针对开采煤层、邻近煤层和采空区, 每个样本容量为 18, 前

13 组样本作为训练样本, 后 5 组为测试样本。建模过程主要分为以下 4 步: ① 根据训练样本利用 Matlab 软件平台编写 SVM 学习程序; ② 输入训练样本, 得到模型参数; ③ 利用模型参数, 建立预测模型; ④ 键入测试样本, 得到预测结果。

模型运行硬件环境为 Pentium M-1.2G CPU, 128 Mb 内存; 软件环境为 Matlab7.0。实验求取了模型对各样本集学习样本、测试样本均方差及 CPU 耗时, 见表 1; 表 2 给出了对预测样本预测值及相对误差; 图 1 为各分源瓦斯涌出量预测效果。

表 1 LS-SVM 模型预测均方差及 CPU 耗时

Table 1 Average squares mistake and CPU time-consumption of LS-SVM model

瓦斯区	均方差		CPU 耗时/s
	训练样本	测试样本	
开采煤层	1.0312×10^{-4}	0.023 8	0.112
邻近煤层	7.9769×10^{-7}	0.003 3	0.090
采空区	2.9574×10^{-4}	0.001 4	0.107
回采工作面	0.007 6	0.019 4	-

表 2 训练样本期望输出与预测结果的对比

Table 2 Comparison of training sample expectation outputs and prediction results

样本号	开采煤层			邻近煤层			采空区			回采工作面		
	实际值	预测值	误差/%	实际值	预测值	误差/%	实际值	预测值	误差/%	实际值	预测值	误差/%
14	5.64	5.756 2	2.06	1.48	1.504 8	1.68	1.39	1.405 4	1.11	8.51	8.666 4	1.05
15	5.34	5.539 6	3.74	1.23	1.263 7	2.74	1.38	1.310 1	5.07	7.95	8.113 4	2.82
16	2.28	2.181 7	4.31	1.56	1.497 6	4.00	0.22	0.201 7	8.32	4.06	3.881 0	2.95
17	3.13	3.321 8	6.13	1.49	1.405 4	5.68	0.30	0.303 1	1.03	4.92	5.030 3	1.87
18	5.30	5.438 2	2.61	1.26	1.320 8	4.83	1.36	1.323 8	2.66	8.04	8.082 8	5.31

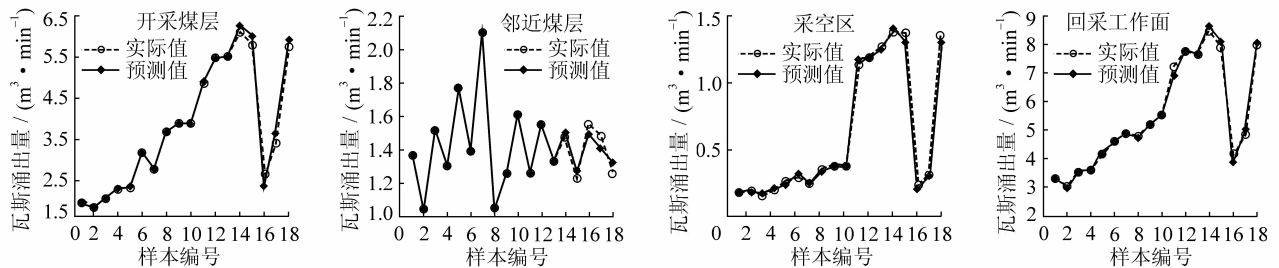


图 1 LS-SVM 模型预测

Fig. 1 Prediction results of LS-SVM model

3 实验结果分析

从图 1 可以直观地看出, LS-SVM 预测模型对学习样本具有很高的拟合精度, 其均方差平均在 $10^{-4} \sim 10^{-7}$ 之间 (表 1); 对测试样本也具有较高的预测精度, 对各区瓦斯涌出量的预测误差平均为 3.8% 左右, 对回采工作面的瓦斯涌出量预测误差为 2.8% 左右, 能够满足煤矿对于瓦斯涌出量预测精度的要求; LS-SVM 模型稳定, 算法简单, CPU 耗时少, 符合对瓦斯涌出量实时在线预测的要求。

4 结 语

回采工作面瓦斯涌出量由多种因素共同影响决定, 且各种因素之间存在复杂的非线性关系, 很难用一般的算法逼近。本文提出了一种基于 LS-SVM 的瓦斯涌出量预测新方法, 充分考虑影响瓦斯涌出量的各种因素, 分别对回采工作面开采层、邻近层和采空层瓦斯涌出量建立 LS-SVM 预测模型, 从而实现对回采工作面瓦斯涌出量的精确预测。实验表明, 该方法具有精度高、速度快、模型稳定等优点, 适合对煤矿瓦斯涌出量的在线实时预测。

参考文献:

- [1] 戴广龙, 汪有清, 张纯如, 等. 保护层开采工作面瓦斯涌出量预测 [J]. 煤炭学报, 2007, 32 (4): 382-385.
Dai Guanglong, Wang Youqing, Zhang Chunru, et al. Forecast of the gas effused from the face in protective seam [J]. Journal of China Coal Society, 2007, 32 (4): 382-385.
- [2] 包剑影, 苏 煊, 李贵坚. 阳泉煤矿瓦斯治理技术 [M]. 北京: 煤炭工业出版社, 1996.
Bao Jianying, Su Xuan, Li Guijian. Technique of gas control of Yangquan Coal Mine [M]. Beijing: China Coal Industry Publishing House, 1996.
- [3] 朱红青, 常文杰, 张 彬. 回采工作面瓦斯涌出 BP 神经网络分源预测模型及应用 [J]. 煤炭学报, 2007, 32 (5): 504-508.
Zhu Hongqing, Chang Wenjie, Zhang Bin. Different-source gas emission prediction model of working face based on BP artificial neural network and its application [J]. Journal of China Coal Society, 2007, 32 (5): 504-508.
- [4] 林柏泉, 崔恒信. 矿井瓦斯防治理论与技术 [M]. 徐州: 中国矿业大学出版社, 1998.
Lin Baiquan, Cui Hengxin. The theory and technique of methane prevention in coal mine [M]. Xuzhou: China University of Mining and Technology Press, 1998.
- [5] Vapnik V N. The nature of statistical learning theory [M]. New York: Springer-Verlag, 1999.
- [6] Vapnik V N. An overview of statistical learning theory [J]. IEEE Transaction Neural Networks, 1999, 10 (5): 988-999.
- [7] Wu Dehui. Prediction method for machining quality based on weighted least squares support vector machine [A]. Proceedings of the 6th World Congress on Control and Automation [C]. Dalian: 2006.
- +++++

第六届国际矿业科学与技术大会征稿启事

1. 论文征集 由中国矿业大学主办的“第六届国际矿业科学与技术大会”将于 2009 年 10 月 18~20 日在江苏省徐州市中国矿业大学召开, 现已开始征集论文. 欢迎国内外矿业科学与技术领域的专家、学者踊跃投稿. 本次大会以“矿业与可持续发展”为主题, 探讨、交流矿业科学与技术领域的最新研究进展、面临的机遇和挑战以及该领域的发展前景.

2. 大会主要议题 (1) 现代矿山开采与安全; (2) 矿山岩石力学与地下工程; (3) 煤炭加工转化与资源综合利用 (矿业材料); (4) 矿山地质与测绘科学; (5) 矿山装备与数字化; (6) 矿山环境保护与节能减排; (7) 能源战略及其可持续发展.

3. 论文摘要具体要求 论文作者需提交一篇 300 字左右的信息型中英文论文摘要 (包括目的、方法、内容、结果和结论), 通过 E-mail (icmst2009@cumt.edu.cn) 以 Word 附件形式发送或邮寄 2 份论文摘要复印件 (A4 纸) 到:

江苏省徐州市中国矿业大学 (南湖校区) 党政办公楼 A317 室 (邮编: 221116)

第六届国际矿业科学与技术大会秘书处收

请同时提供一个 200 字左右的作者简介 (中文即可), 注明年龄、职称、工作单位、联系方式 (通讯地址、电话、手机、传真、邮编、E-mail 等).

4. 大会网站 有关大会的最新消息将及时在大会网站 (<http://icmst2009.cumt.edu.cn>) 上发布, 大会征文通知等可在该网站下载.

5. 论文摘要投稿截止时间 2009 年 1 月 15 日.

6. 大会工作语言 大会工作语言为英语, 所有论文报告和发表必须使用英语.