

文章编号: 1671- 251X(2010) 10- 0032- 04

基于粒子群优化的支持向量机在瓦斯浓度预测中的应用研究

张剑英¹, 许徽¹, 陈娟², 曹新德¹

(1. 中国矿业大学信电学院, 2. 中国矿业大学材料学院, 江苏 徐州 221008)

摘要:为了准确预测煤矿瓦斯浓度, 基于从芦岭煤矿KJ98监控系统中提取的生产现场瓦斯浓度时间序列数据, 对基于粒子群优化的支持向量机理论在瓦斯浓度短期预测中的应用进行了研究。首先对瓦斯浓度时间序列进行小波软阈值去噪和相空间重构等预处理, 然后采用粒子群优化算法对支持向量机的惩罚因子、损失函数、核函数参数进行了优化, 并基于最优参数建立了瓦斯浓度预测的支持向量机模型。仿真结果表明, 采用粒子群优化的支持向量机理论进行煤矿瓦斯浓度预测, 极大地提高了预测的准确性和精确度; 误差分析结果表明, 该方法预测结果的误差很小, 且测试样本越小, 误差越小。

关键词:煤矿; 瓦斯浓度; 预测; 支持向量机; 粒子群优化; 建模

中图分类号:TD712. 5 **文献标识码:**A

Application Research of Support Vector Machine Based on Particle Swarm Optimization in Gas Concentration Forecasting

ZHANG Jiayi¹, XU Hui¹, CHEN Juan², CAO Xinde¹

(1. School of Information and Electrical Engineering of CUMT., Xuzhou 221008, China.

2. Shool of Material Science and Engineering of CUMT., Xuzhou 221008, China)

Abstract: In order to forecast gas concentration of coal mine, the paper studied application of support vector machine based on particle swarm optimization in gas concentration short-term forecasting according to original gas concentration time series data in KJ98 monitoring and control system of Luling coal mine. In the application, gas concentration time series are pre-processed firstly by use of methods of soft-threshold de-noising of wavelet and phase-space reconstructing, then parameters of penalty factor, loss function and kernel function of support vector machine are optimized by use of particle swarm optimization algorithm, and a model of support vector machine of gas concentration forecasting is built based on the optimal parameters. The simulation result showed that the forecasting method of support vector machine based on particle swarm optimization greatly improves correctness and accuracy of gas concentration forecasting. Error analysis result showed that forecasting errors are small by use of the method and the less forecasting sample is, the smaller error is.

Key words: coal mine, gas concentration, forecasting, support vector machine, particle swarm optimization, modeling

0 引言

矿井瓦斯事故是煤矿五大自然灾害之一^[1]。瓦

收稿日期: 2010- 06- 22

作者简介: 张剑英(1963-), 女, 江苏徐州人, 教授, 博士, 主要研究方向为机器学习、模式识别、信号处理。E-mail: xhfc1985@sina.com

© 1994-2010 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

斯爆炸事故将严重危及矿井生产和井下工人的生命安全, 其产生的有害气体也将直接影响井下工人的身体健康, 甚至使人窒息死亡^[2]。因此, 准确预测矿井瓦斯浓度能有效防止瓦斯聚积超限和瓦斯爆炸等煤矿恶性事件的发生, 进而保障煤矿的生产安全。

矿井瓦斯浓度预测是以煤层瓦斯含量及其分布

规律, 或以煤层中瓦斯浓度的变化规律为基础, 并结合地质因素、开采因素来选取合理参数, 以一定方法预测瓦斯浓度的工作过程。本文以瓦斯浓度这一影响因素为出发点, 从淮北矿业集团公司芦岭煤矿的监控系统中提取相关的瓦斯浓度原始值, 采用基于粒子群优化(Particle Swarm Optimization, PSO) 的支持向量(Support Vector Machine, SVM) 回归理论对其进行有效预测, 以便工作人员及时采取必要的措施, 避免瓦斯爆炸事故的发生。

本文研究的瓦斯浓度预测基本流程如图 1 所示。

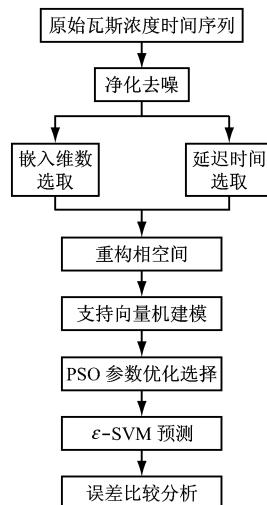


图 1 瓦斯浓度预测基本流程

1 瓦斯浓度时间序列预处理

所采用的瓦斯含量资料来源于芦岭煤矿 KJ98 监控系统所获得的瓦斯浓度数据, 选用 II826 工作面中的 II826- 西延 2 号眼(T2)这个比较有代表性巷道的瓦斯浓度原始数据, 选取 2006 年 2 月的 2 500 个瓦斯浓度值。瓦斯浓度原始数据曲线如图 2 所示。

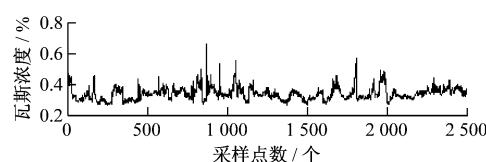


图 2 II826- 西延 2 号眼(T2)瓦斯浓度原始数据曲线

1.1 瓦斯浓度时间序列软阈值小波去噪

瓦斯浓度历史数据的准确性是预测精度的基础, 但由于在实际中瓦斯浓度时间序列必然存在一定的噪声, 因此, 在预测前需要对原始瓦斯浓度数据进行去噪处理。

小波理论能有效区分信号中的突变成分和噪

声, 其多分辨率、去相关性、选基灵活性等特点都能更好地去除噪声并刻画信号的非平稳特征^[3]。常用的小波去噪法为软阈值去噪法, 如式(1)所示。

$$W_{i,j} = \begin{cases} \text{sign } W_{i,j} \cdot (|W_{i,j}| - \lambda) & |W_{i,j}| \geq \lambda \\ 0 & |W_{i,j}| < \lambda \end{cases} \quad (1)$$

式中: $W_{i,j}$ 为小波系数; $\lambda = \sigma \sqrt{2 \ln N}$, 其中 σ 为噪声标准差, N 为分解层次。

经软阈值去噪法处理后的小波系数连续性好, 易于处理。

应用小波软阈值去噪法对瓦斯浓度时间序列进行去噪处理的步骤:

(1) 瓦斯浓度时间序列的小波分解: 选择一个小波并确定分解层次 N , 然后进行 N 层小波分解。

(2) 小波分解高频系数的阈值量化: 对第一层到第 N 层的每一层高频系数选择一个阈值进行软阈值量化处理。

(3) 瓦斯浓度时间序列的小波重构: 进行瓦斯浓度时间序列的小波重构, 从而得到去噪后的信号。

对 II826- 西延 2 号眼(T2) 瓦斯浓度时间序列进行软阈值去噪处理, 阈值为 0.0177, 仿真结果如图 3 所示。

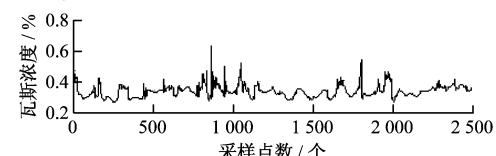


图 3 II826- 西延 2 号眼(T2)瓦斯浓度时间序列软阈值去噪仿真结果

1.2 瓦斯浓度时间序列相空间重构

瓦斯浓度时间序列中包含着大量信息, 因此, 必须将其扩展到三维或者更高维数的相空间, 才能将其中蕴含着的参与运动的全部变量轨迹展现出来, 这就要对瓦斯浓度时间序列进行相空间重构。选用 C-C 法对其进行相空间重构。

将瓦斯浓度时间序列 $\{x(t_i), i=1, 2, \dots, n\}$ 分解成 t 个不相交的时间序列, 当 $t=2$ 时为 $\{x(t_1), x(t_3), \dots, x(t_{n-1})\}$ 和 $\{x(t_2), x(t_4), \dots, x(t_n)\}$, 当 $t=1$ 时为单个时间序列本身, 长度为 $n/2$ 。对于 t 个不相交的瓦斯浓度子序列, 计算每个子序列的统计量 $S(m, n, r, t)$:

$$S(m, n, r, t) = \frac{1}{t} \sum_{s=1}^t C_s(m, n/t, r, t) - C_s^m(1, n/t, r, t) \quad (2)$$

式中: m 为嵌入维数, n 为瓦斯浓度时间序列的

长度; r 为领域半径的大小, $r > 0$; τ 为延迟时间。

C-C 法认为相空间中的延迟时间 τ 和嵌入维数 m 是相关的, 二者之间的关系为 $\tau_w = (m - 1)\tau$, 其中 τ_w 为延迟时间窗口。 τ 依赖于 $m^{[4]}$, 且 τ 及 x_i 个分量之间相互依赖。

计算如下 3 个统计量:

$$\bar{S}(t) = \frac{1}{16} \sum_{m=2}^5 \sum_{j=1}^4 S(m, n, r_j, \tau) \quad (3)$$

$$\Delta\bar{S}(t) = \frac{1}{4} \sum_{m=2}^5 \Delta S(m, n, \tau) \quad (4)$$

$$\tau_w = |\Delta\bar{S}(t)| + |\bar{S}(t)| \quad (5)$$

式中: $\bar{S}(t)$ 为所有子序列的统计量 $S(m, n, r, \tau)$ 的均值; $\Delta\bar{S}(t)$ 的第一极小值对应延迟时间 τ ; τ_w 为最佳嵌入窗宽, 它对应 $S_{cor}(t)$ 最小值。

对经过去噪处理的 II826- 西延 2 号眼(T2)瓦斯浓度时间序列进行相空间重构, 在 Matlab 中编程计算后的结果如表 1、表 2 所示。

表 1 II826- 西延 2 号眼(T2) $\Delta\bar{S}(t)$ 值

t/s	19	20	21	22	23
$\Delta\bar{S}(t)$	0.133 6	0.122 2	0.114 0	0.115 5	0.118 0

表 2 II826- 西延 2 号眼(T2) $S_{cor}(t)$ 值

t/s	81	82	83	84	85
$S_{cor}(t)$	0.017 5	0.006 7	0.012 1	0.014 8	0.019 0

从表 1 可看出, $\Delta\bar{S}(t)$ 第一次取最小值时 $t = 21$ s, 因此, II826- 西延 2 号眼(T2) 瓦斯浓度时间序列相空间重构后的延迟时间 $\tau = 21$ s。而在表 2 中, $S_{cor}(t)$ 取最小值时 $t = 82$ s, 所以延迟时间窗口 $\tau_w = 82$ s。根据公式 $\tau_w = (m - 1)\tau$ 可得嵌入维数 $m \approx 5$ 。

2 支持向量机参数的粒子群优化

在支持向量机瓦斯浓度时间序列预测模型中, 选用不同的参数会对训练效果及瓦斯浓度预测精度造成很大的影响。

(1) 损失函数 ε 的影响

损失函数 ε 的值影响泛化能力大小及支持向量的数量, 取值适当则可以均衡训练模型的泛化能力和拟合准确度。

(2) 惩罚因子 C 的影响

在瓦斯浓度时间序列里, 惩罚因子 C 的取值与经验误差紧密相关。若其取值较大, 则训练模型和实际值的拟合精度较好, 但经验误差较大; 若其取值

较小, 则经验误差较小, 但是会导致风险值较大。

(3) 核函数参数 σ 的影响

核函数参数 σ 的取值将直接影响到映射函数, 进而相应地改变子空间中分布的样本数据的复杂程度, 也影响到最优模型的 VC 维。

可见, 采用合适的参数对于最终的预测精度有着极大的影响, 而粒子群优化算法是从鸟儿寻食这种种群行为中获得启发, 并且在计算机中得到模拟来求解特定优化问题的算法^[5]。在粒子群优化算法中一个重要的概念即是“粒子”(Particle)。粒子是 D 维空间中的一个点, 它对应于优化问题的潜在特定解^[6]。

若一个群落由 m 个粒子组成, 一个 D 维的向量表示其中的第 i 个粒子, 如式(6)所示。

$$X_i = (x_{i1}, x_{i2}, \dots, x_{iD})^T \quad i = 1, 2, \dots, m \quad (6)$$

记第 i 个粒子到目前为止搜索到的最优位置为 $p_{besti} = (p_{besti1}, p_{besti2}, \dots, p_{bestiD})$, 整个粒子群到目前为止搜索到的最优位置为 $g_{best} = (p_{best1}, p_{best2}, \dots, p_{bestD})$ 。

对 II826- 西延 2 号眼(T2) 瓦斯浓度数据进行小波软阈值去噪和相空间重构后, 应用粒子群优化算法对支持向量机预测建模参数进行优化, 以取得最佳预测精度和效果, 其中惩罚因子 C 、损失函数 ε 和核函数参数 σ 为待优化的参数。在 Matlab 中编程计算后仿真结果如图 4 所示。

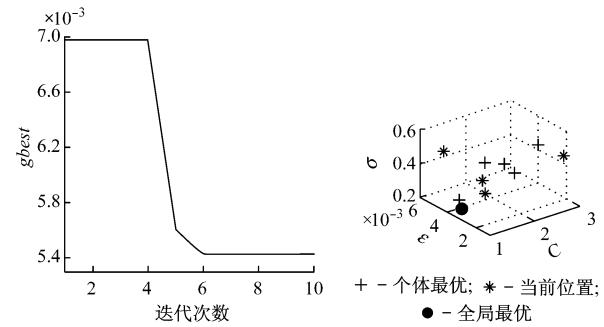


图 4 II826- 西延 2 号眼(T2) 支持向量机预测建模参数的粒子群优化算法仿真结果

根据图 4 得出最优参数, 即 $\varepsilon = 0.010 38$, $C = 1.1$, $\sigma = 0.153$ 。

3 瓦斯浓度预测的支持向量机建模

支持向量机是在统计学习理论基础上发展来的一种新的机器学习方法, 它通过核函数将样本映射到了高维空间中, 并在高维空间中构造最优分类

超平面^[7-8]。支持向量机模型的求解最终转化成为一个凸二次优化问题,从而有效地避免了陷入局部极值等问题的发生。

3.1 支持向量机建模

假设给定一个特定的训练数据样本集 $(x_1, y_1), (x_2, y_2), \dots, (x_l, y_l) \in \mathbb{R}^n \times \mathbb{R}$, 用一个线性函数 $f(x) = w \cdot x + b$ 来对数据进行拟合, 用 ε 不敏感损失函数作为损失函数, 这时就产生了对回归的支持向量估计^[9]。支持向量回归估计的优化问题表示为

$$\text{Minimize } \frac{1}{2} \|w\|^2 + C \sum_{i=1}^l (\xi_i + \xi_i^*)$$

$$w \cdot x_i + b - y_i \leq \xi_i^* + \xi_i, i = 1, \dots, l$$

Subject to $y_i - (w \cdot x_i) - b \leq \xi_i^* + \xi_i, i = 1, \dots, l$
 $\xi_i, \xi_i^* \geq 0$

使用经粒子群优化后的最优参数对II826-西延2号眼(T2)瓦斯浓度数据进行支持向量机建模预测, 将前2400个瓦斯浓度值作为训练样本, 后100个瓦斯浓度值作为测试样本, 在Matlab中仿真后结果如图5所示。

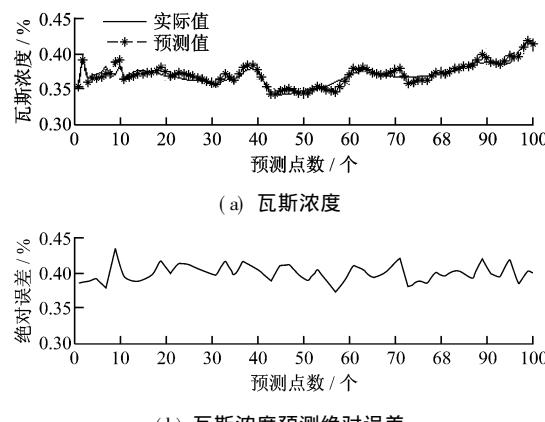


图5 II826-西延2号眼(T2)瓦斯浓度的支持向量机预测建模仿真结果

3.2 误差分析

对得到的预测模型效果进行误差分析。采用预测值和实际值的均方差 E_{ss} 作为判断预测效果的指标, 如式(7)所示。

$$E_{ss} = \left[\frac{1}{L} \sum_{i=1}^L (x(N+i) - \hat{x}(N+i))^2 \right]^{\frac{1}{2}} \quad (7)$$

式中: L 为时间序列长度; x 为实际序列值; \hat{x} 为预测值。

E_{ss} 小则说明预测值偏离实际值的程度小, 预测效果较好。笔者又对150个和50个测试样本进行了预测建模, 统计结果后的 E_{ss} 值如表3所示。

表3 E_{ss} 值

测试样本数/个	150	100	50
E_{ss}	0.008 79	0.005 43	0.003 52

从表3可看出, 测试样本越小, 则 E_{ss} 值越小, 预测的精度就越高。如果要求预测的置信度高, 测试样本就应该小一些, 反之则可以大一些。

4 结语

支持向量机有着优良的泛化能力, 并且在回归预测方面的精度也很高。由于其选取的参数对于泛化的精度和准确度有着巨大影响, 且很多训练模型是建立在试凑或者经验的基础上, 这使得支持向量机的众多优势无法完全实现, 因此, 选择最优的参数就变得极为重要。采用粒子群优化算法对支持向量机的3个参数同时进行粒子寻优, 仿真结果表明, 基于粒子群优化的支持向量机参数选择法可靠、精确, 在此优化参数基础上构造的煤矿瓦斯浓度预测模型具有很好的泛化能力及学习精度。

参考文献:

- [1] 王雨虹, 郝毅魁. 有指导学习在建立瓦斯灾害信息提取模型中的应用[J]. 工矿自动化, 2008(2): 8-12.
- [2] 姚宝魁, 孙广忠. 煤与瓦斯突出的区域性预测[M]. 北京: 中国科学技术出版社, 1993.
- [3] 文鸿雁, 张正禄. 非线性小波变换阈值法去噪改进[J]. 测绘通报, 2006(3): 18-21.
- [4] 修春波, 刘向东, 张宇河. 相空间重构延迟时间与嵌入维数的选择[J]. 北京理工大学学报, 2003, 23(2): 219-224.
- [5] 崔红梅, 朱庆保. 微粒群算法的参数选择及收敛性分析[J]. 计算机工程与应用, 2007, 43(23): 89-91, 131.
- [6] CLARE M, KENNEDY J. The Particle Swarm: Explosion Stability and Convergence in a Multi-dimensional Complex Space [J]. IEEE Transactions on Evolutionary Computation, 2002, 6(1): 58-73.
- [7] 孟祥国, 马军, 段昕. 基于统计学习理论的支持向量机算法研究[J]. 计算机科学, 2006, 33(8): 252-255.
- [8] MITRA P, MURTHY C A, PAL S K. Data Condensation in Large Databases by Incremental Learning with Support Vector Machines [C]// Proceedings of 15th ICPR, 2000, Barcelona: 708-711.
- [9] 张学工. 关于统计理论与支持向量机[J]. 自动化学报, 2000, 26(1): 32-42.