

基于支持向量机模型的瓦斯涌出量预测*

王剑平¹, 黄红霞², 李宏彪¹, 张云生¹

1. 昆明理工大学 信息工程与自动化学院, 昆明 650051

E-mail: kmustwjp@126.com

2. 西南林学院 交通机械与土木工程学院, 昆明 650224

E-mail: hhx001@gmail.com

摘要: 针对现阶段瓦斯涌出量影响因素样本量少的具体状况, 提出一种支持向量机(SVM)理论的煤矿瓦斯涌出量的预测方法。从非线性预测分析入手, 建立瓦斯涌出量预测非线性支持向量机模型, 并与其他预测方法进行比较, 验证结果表明采用支持向量机用于瓦斯涌出量预测具有较高精度。

关键词: 支持向量机, 瓦斯涌出量, 预测

Prediction of Mine Gas Emission Rate Based on the SVM Model

Wang Jianping¹, Huang Hongxia², Li Hongbiao¹, Zhang Yunsheng¹

1. Faculty of Information Engineering and Automation, Kunming University of Science and Technology, Kunming 650051, P. R. China

E-mail: kmustwjp@126.com

2. Faculty of Communication, Machinery and Civil Engineering, Southwest Forestry College, Kunming 650224, P. R. China

E-mail: hhx001@gmail.com

Abstract: In allusion to the actuality that the samples of influence factors on gas emission rate are few now, this paper gives out a predicting method for mine gas emission based on support vector machine (SVM) theory, building a SVM model of mine gas emission from nonlinear forecasting. Comparing to the other forecasting methods, the result shows that this model has more precision.

Key Words: Support Vector Machine, Gas Emission Rate, Forecast

1 引言(Introduction)

瓦斯涌出量对矿井安全、设计、建设和开采都有重要影响^[6]。影响瓦斯涌出量因素众多, 如: 地质结构、煤质结构、煤层厚度、开采方式等。所有这些因素之间的非线性关系错综复杂, 难以控制, 对矿井瓦斯涌出量预测存在着较大的困难。

国内外采用的传统的瓦斯涌出量预测方法主要有: 统计分析法、瓦斯含量法、分源算法、类比法、灰色理论^[7]等。这些方法都存在一定的局限性, 主要表现为: 一是这些预测模型表示线性关系, 对瓦斯涌出量与预测影响因素的高度非线性无法辨识, 对于一些模糊等不确定性因素无法处理, 造成预测结果的严重失真; 二是在实际预测时, 由于受到各种客观条件的影响, 各种标准尚未完善, 统计口径还不统一, 资料不太齐全, 无法取得足够多的样本, 使得样本较少, 无法保证预测的精度。

针对第一种局限性目前广泛采用神经网络方法来弥补其非线性与不确定性, 但是神经网络方法要求有足够的样本, 使得问题最终归结为寻找到小样本预

测上来。支持向量机 (Support Vector Machine, SVM) 是由Vapnik等在20世纪90年代中期提出的一种新的机器学习方法^[1]。它采用结构风险(SRM)原则, 较好地解决了小样本的学习问题, 具有全局唯一最优解、泛化能力强、模型结构由算法自动确定等一系列优点。本文利用支持向量机模型建立瓦斯涌出量时间序列非线性预测模型, 实现矿井开采过程中的瓦斯涌出量动态非线性预测。

2 支持向量机非线性预测模型(SVM Nonlinear Forecasting Model)

2.1 瓦斯涌出量映射模型(Mapping Model of Gas Emission Rate)

SVM就是通过内积函数定义的非线性变换将输入空间转化到一个高维空间, 从而在高维空间中寻找输入向量和输出向量之间的非线性关系, 实现在有限样本情况下的机器学习方法^[2]。

与人工神经网络相比, SVM不存在陷入局部最优等问题, 泛化能力较强; SVM有严格的理论基础, 是基于结构风险最小化原则的方法, 明显优于传统的基于经验风险最小化原则的常规神经网络方法; SVM算法是一个凸二次优化问题, 保证找到的解是全局最优解, 能较好地解决小样本、非线性、高维数等实际问题。据于此, 可建立瓦斯涌出量预测映射模型。

* 此项工作得到云南省教育厅科学研究基金项目资助, 项目名称: 煤矿瓦斯事故征兆智能诊断技术研究, 项目批准号: 07Y41416。

此项工作得到云南省科技厅信息专项项目资助, 项目名称: 煤矿安全预警信息系统研究及应用示范, 项目批准号: 2005it02。

此项工作得到昆明理工大学青年基金项目资助, 项目名称: 煤矿瓦斯灾害的智能诊断技术研究, 项目批准号: 校青2006-52。

设 $\{x_i\}$ 是影响瓦斯涌出量的相关因子, 影响因子序列构造映射关系 $f: R^m \rightarrow R$, (m 为因子数), 将此映射关系表示成系统输出 y_i 与系统输入 x_i 之间的关系, 即 $y_i = f(x_i)$ $y_i \in R, x_i \in R^m$ 。Taklins^[2] 定理保证在一定条件下, 存在光滑映射 $f: R^m \rightarrow R$ 。对瓦斯涌出量的预测就是基于这一基本映射。

根据 l 个独立同分布观测样本 (x_i, y_i) ($i=1,2,\dots,m$), 在函数集 $\{f(x, \omega)\}$ 寻求一个最优函数 $f(x, \omega_0)$, 使期望风险 $R(\omega) = \int L(y, f(x, \omega)) dF(x, y)$ 最小, 其中 $\{f(x, \omega)\}$ 称为预测函数集, ω 为函数的广义参数。 $L(y, f(x, \omega))$ 为由于 $f(x, \omega)$ 对 y 进行预测而造成的损失。由于系统中的不确定因素, 将预测问题表示为一个回归问题: $y_i = f(x_i) + \varepsilon_i$, 式中 $y_i \in R, x_i \in R^m$, ε_i 为噪声, 与 x_i 相独立。在保证经验风险尽可能小的情况下, 极小化置信风险上界, 通过一个核函数将特征函数映射到高维空间, 并构造一个最优分类超平面, 实现预测, 该预测问题可归纳为一个带约束的二次规划问题^[1],

$$\begin{aligned} \text{即 } \min_{\omega, b, \xi} & \left(\frac{1}{2} \omega^T \omega + C \sum_{i=1}^l \xi_i \right) \\ \text{约束: } & y_i (< \omega, \varphi(x_i) > -b) \geq 1 - \xi_i \end{aligned} \quad (1)$$

该二次规划问题用偶规划进行求解得到

$$\max_{\alpha} \omega(Q) = \sum_{i=1}^l Q_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l a_i a_j y_i y_j k(x_i, x_j)$$

$$\begin{aligned} \text{约束: } & y^T a = 0, 0 \leq a_i \leq C \\ & i = 1, 2, \dots, l \end{aligned} \quad (2)$$

式中 $a = (a_1, a_2, \dots, a_l)^T$, a_j 为式 (1) 中不等式约束对应的拉格朗日乘子; $k(x_i, x_j)$ 为核函数。根据文献 [9] 推导, 对于瓦斯涌出量这一非线性预测问题, 可得寻找回归函数:

$$\hat{y} = \langle \omega^*, \varphi(x_i) \rangle + b^* = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x, x_i) + b^* = f(x_i) \quad (3)$$

其中 a_j^* 和 b^* 分别为 a_j 和 b 的最优值。

2.2 核函数(Core Functions)

在映射模型中的核函数是一种非线性映射函数, 它可以将低维输入空间数据映射到特征空间, 使得原来输入空间中的线性不可分问题, 转化为特征空间中的线性可分问题。核函数的引入可以将输入空间中非线性可分的问题转变成高维特征空间中线性可分的问题, 然后在特征空间中使用学习机建立优化超平面, 从而解决问题。作为研究本文主要采用以下核函数:

1) 多项式核函数:

$$K(x, x_i) = [(x \cdot x_i) + 1]^q \quad (4)$$

2) 径向基函数 (RBF):

$$K(x, x_i) = \exp \left\{ -\frac{\|x - y\|^2}{2\delta^2} \right\} \quad (5)$$

3) 采用Sigmoid函数作为内积:

$$K(x, x_i) = \tanh(v(x \cdot x_i) + c) \quad (6)$$

2.3 计算步骤^[4] (Computing Steps)

模型求解算法是一个经典的二次规划问题, 可用现成的优化软件求解。具体步骤为:

解二次规划:

$$\begin{aligned} \text{Max } W(\alpha, \alpha_i^*)|_{w, b, \xi, \varepsilon_i} &= -\frac{1}{2} \sum_{i,j=1}^l (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) - \sum_{i=1}^l (\alpha_i + \alpha_i^*) \varepsilon + \sum_{i=1}^l (\alpha_i - \alpha_i^*) y_i \\ \text{s.t. } & \sum_{i=1}^l (\alpha_i - \alpha_i^*) = 0 \end{aligned}$$

$$0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, 2, \dots, l \quad (7)$$

求出 α_i, α_i^* 的值, 以及权值 w 。其中 C 为二次优化参数, ε 为不敏感系数。

解出 b 的值。

把需要预测的影响因素代入式 (8), 即可得出结果。

$$f(x) = \sum_{i=1}^l (\alpha_i - \alpha_i^*) K(x, x_i) + b \quad (8)$$

3 瓦斯涌出量预测 (Forecasting of Gas Emission Rate)

3.1 影响因素与样本的选择 (Effect Factors and the Selection of Samples)

对上述瓦斯涌出量预测SVM模型首先要选择瓦斯涌出量的影响因素。根据文献[3], 影响采煤工作面瓦斯涌出量的主要因素包括煤层埋藏深度、煤层厚度、煤层瓦斯含量、工作面煤层与邻近煤层的层间距、工作面平均日进度、工作面平均日产量等6项指标。论文选取文献[5]中某矿3个工作面18个月的矿井瓦斯涌出量统计资料作为预测样本集 (表1)。

3.2 预测结果比较 (Comparing the Forecast Results)

通过对支持向量机模型, 取SVM核函数分别为多项式和径向基的两种方法作拟合计算, 并用Matlab编程实现, 同时与采用BP神经网络方法的结果进行比较, 这里:

取某煤矿实际记录的煤层深度、厚度、煤层间距、掘进日进度、日产量与相对应的绝对瓦斯涌出量数据为样本参数。其中1~15组的数据共15组样本为训练数据, 16~18组数据为测试数据, 所有数据在训练过程中均进行归一化处理。同时, 对预测效果进行评价的指标采用预测结果的绝对误差AE和平均绝对百分比误差MAPE。

表 1 瓦斯涌出量预测训练样本数据

样本序号	输入值						输出值
	煤层埋藏深度 m	煤层厚度 m	煤层瓦斯含量 m ³ /t	煤层间距 m	日进度 m/d	日产量 t/d	绝对瓦斯涌出量 m ³ /min
1	408	2.0	1.92	20	4.42	1825	3.34
2	411	2.0	2.15	22	4.16	1537	2.97
3	420	1.8	2.14	19	4.13	1751	5.36
4	432	2.3	2.58	17	4.67	2078	3.62
5	456	2.2	2.40	20	4.51	2104	4.17
6	516	2.8	3.22	12	3.45	2242	4.60
7	527	2.5	2.80	11	3.28	1979	4.92
8	531	2.9	3.35	13	4.02	2288	4.78
9	550	2.9	3.61	14	3.53	2325	5.23
10	563	3.0	3.68	12	2.85	2410	5.56
11	590	5.9	4.21	18	2.85	3139	7.24
12	604	6.2	4.03	16	2.64	3354	7.80
13	607	6.1	4.34	17	2.77	3087	7.68
14	634	6.5	4.80	15	2.92	3620	8.51
15	640	6.3	4.67	15	2.75	3412	7.95
16	450	2.2	2.43	16	4.32	1996	4.06
17	544	2.7	3.16	13	3.81	2207	4.92
18	629	6.4	4.62	19	2.80	3456	8.04

表 2 测试数据相对误差比较

样本序号	BP神经网络	SVM	
		多项式	径向基
16	0.0890	0.0142	0.0315
17	0.0040	0.0051	0.0200
18	0.0825	0.1120	0.0130
MAPE	0.0585	0.0438	0.0215

多项式核函数SVM中阶数为 $d=5$, $C=50$, $\varepsilon=0.001$, 经过训练得到训练数据的实际值与预测值的绝对误差, 见图1; 径向基SVM中比较优化选择核函数参数为 $\delta^2=0.5$, $C=5$, $\varepsilon=0.00001$ 经过训练得到训练数据的实际值绝对误差, 见图2。表2为对应瓦斯涌出量数据的实际值与预测值的误差。

从表2三种预测方法相对误差的比较结果, 显然可以得出:

1) BP神经网络和SVM不需要设计任何数学解析模型, 只靠过去的经验来学习, 可处理非线性、模糊与含有噪声的数据。实例表明, 基于神经网络的预测方法具有一定的优点, 但是数据量较少可能会减弱网络的推广能力和预测的精度。

2) 对于BP神经网络预测模型训练数据的误差比较小, 而推广能力比较弱, 主要原因是神经网络需要大量的数据、隐层神经元个数比较少和存在过拟合现象。

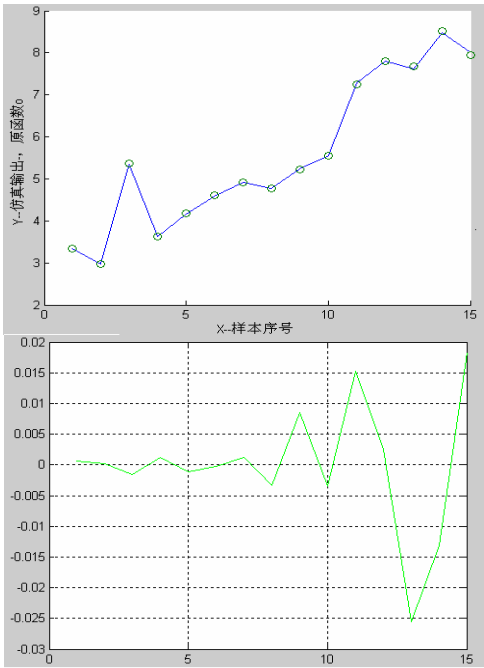


图1 多项式SVM训练误差曲线

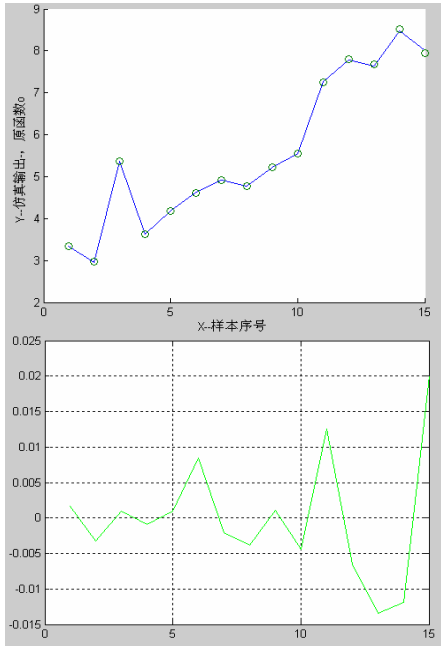


图2 径向基训练误差曲线

3) SVM预测模型, 采用径向基核函数模型比多项式核函数模型具有更好的拟合能力和推广能力, 这是因为核函数参数的选择对于预测的精度和推广能力具有及其重要的影响。

4) 相比较而言, SVM预测模型具有更好的拟合和推广能力, 这是因为瓦斯涌出量预测系统是一个复杂非线性系统。大量影响因子数据样本很难取得。而SVM是针对小样本提出的统计学习算法, 不存在过拟合现象, 可避免神经网络需要大量样本数据的问题。SVM算法的本质是一个求解二次规划问题, 算法简单, 可以推广应用。

4 结论 (Conclusion)

本文建立非线性SVM瓦斯涌出量预测模型,并通过实例验证。通过与BP神经网络进行比较证明采用SVM用于瓦斯涌出量预测比其他方法有更高的预测精度。该预测方法针对在瓦斯涌出量影响因素数据样本量少的情况下,使得对瓦斯涌出量的预测效果有了一定的提高,同时比单纯地使用BP等神经网络模型对瓦斯涌出量预测的效果好,并在实际生产中取得了比较理想的效果。

参考文献(References)

- [1] VAPNIK.The Nature of Statistical Learning Theory. New York, Springer-Verlag , 1995.
- [2] COURANT R, HILBER.Methods of Mathematical Physics[M] . New York,Wiley Inter-Science , 1970.
- [3] 黄炜伟,王强等.基于混沌时间序列的神经网络对瓦斯涌出量预测.煤炭学报,2005,14(5): 7-9.
- [4] 魏连雨,庞明宝.基于神经网络的物流量预测.长安大学学报(自然科学版),2004, 24(6): 55-59.
- [5] 王生全.采煤工作面瓦斯涌出量的建模预测.东北煤炭技术, 1997, 2: 38-41.
- [6] 赵朝义,袁修干,孙金鏢.遗传规划在采煤工作面瓦斯涌出量预测中的应用.应用基础与工程科学学报, 1999.12, 7(4): 385-392.
- [7] 谷松.基于灰色理论的小波神经网络对瓦斯涌出量的预测.煤炭学报,2007.9, 32(9): 964.
- [8] 向玲.基于支持向量机的设备异常状态时间预测.传感技术学报,2007.9, 20(9):2040-2043.
- [9] 史忠植.高级人工智能.北京:科学出版社, 242-P245.