

基于 GA - LSSVR 算法的回采工作面瓦斯涌出量预测

曹庆奎, 商娜欣

(河北工程大学 经济管理学院, 河北 邯郸 056038)

摘要: 针对回采工作面瓦斯涌出量问题的小样本、非线性、影响因素关系复杂等特点, 采用遗传 - 最小二乘支持向量回归算法对瓦斯涌出量进行预测, 利用定量方法进行分析, 避免了定性分析的局限性, 有效提高了预测的精度。该模型首先利用遗传算法对最小二乘支持向量回归机中的参数进行训练和优化, 然后运用遗传 - 最小二乘支持向量回归模型对测试样本进行了回采工作面瓦斯涌出量测试。测试结果表明: 与支持向量回归机以及最小二乘支持向量回归机的预测值相比, 遗传 - 最小二乘支持向量回归的回采工作面瓦斯涌出量预测可靠性和精确性更高。

关键词: 瓦斯涌出量; 回采工作面; 预测; 最小二乘支持向量回归机; 遗传算法

中图分类号: TD712.5

文献标识码: A

Prediction of gas emission quantity of the working face based on genetic algorithm - least squares support vector regression

CAO Qing - kui, SHANG Na - xin

(School of Economics and Management, Hebei University of Engineering, Hebei Handan 056038, China)

Abstract: The problem of the gas emission of the working face was characterized by small samples, nonlinear, and it was affected by complex factors. Using the genetic - least squares support vector regression algorithm to predict the gas emission could avoid the qualitative analysis limitations and effectively improve the accuracy of the forecast, because it was a quantitative method for analysis. First, the model used the genetic algorithms to train and optimize the least squares support vector regression parameters, and then used the genetic - least squares support vector regression model to predict the amount of gas emission of test samples. The test results show that: the genetic algorithm - least squares support vector regression model has a higher reliability and accuracy, compared to the predicted values of the support vector regression and the least squares support vector regression.

Key words: gas emission quantity; working face; prediction; least squares support vector regression; genetic algorithm

瓦斯是煤矿开采过程中主要危险因素之一, 同时也是一种新型的清洁能源, 随着煤炭工业的迅速发展, 煤矿开采中发生瓦斯涌出的几率也逐渐增大, 煤矿瓦斯事故发生率也逐年提高, 因此, 对煤矿回采工作面瓦斯涌出量进行有效的预测有非常重大的理论以及现实意义, 不仅可以加强煤矿的生产安全防治工作, 还能够有效的减少煤矿瓦斯重大灾害的发生。

针对瓦斯涌出量的预测现已形成了多种理论和方法^[1]; 吕伏等^[2]利用多元统计学中的主成分分析方法对回采工作面的瓦斯涌出量进行了多步线性回归预测, 结果表明该方法具有较好的精确度; 王生全等^[3]和朱红青等^[4]针对瓦斯涌出量建立了 BP 神经网络预测模型, 并与遗传算法相结合进行了瓦斯涌出量的预测; 邵良杉等^[5-6]将小波框架理论引入到支持向量机预测函数中, 利用小

收稿日期: 2014 - 03 - 13

基金项目: 国家自然科学基金项目(61240050); 河北省科技计划项目(13455406D)

特约专稿

作者简介: 曹庆奎(1963 -), 男, 河北乐亭人, 博士研究生, 教授, 从事管理与决策的理论与方法研究。

波变换将输入向量映射到一个高维特征空间从而利用支持向量机进行预测; 王晓路等^[7]提出了一种基于虚拟状态变量的卡尔曼滤波瓦斯涌出量预测模型, 该模型有较好的跟踪能力和反应速度; 孙林等^[8]基于结构风险最小化的原则提出了一种基于最小二乘支持向量机的回采工作面瓦斯涌出量的预测方法; 陶云奇等^[9-10]建立了改进的灰色马尔柯夫模型, 对工作面绝对瓦斯涌出量进行了预测; 谷松等^[11]和伍爱友等^[12]将灰色系统理论应用到瓦斯涌出量的预测当中, 结果表明预测结果较为理想; 曹庆奎等^[13]针对矿井工作面瓦斯涌出量进行了未确知聚类研究, 较好的解决了矿井瓦斯涌出量的分类问题, 对瓦斯涌出量的预测提供了分类方法。本文将遗传算法和最小二乘支持向量回归机相结合, 利用遗传算法对最小二乘支持向量回归的参数进行优化, 建立回采工作面瓦斯涌出量的预测模型, 并以实例验证其有效性。

1 最小二乘支持向量回归机

最小二乘支持向量回归机(least squares support vector regression, 简称 LSSVR) 通过构造回归函数来解决预测问题, 通过一个非线性映射函数将输入样本数据映射到高维特征空间, 然后在此空间进行回归预测^[14], 与支持向量回归机(support vector regression, 简称 SVR) 有所不同, LSSVR 构造了一个新的二次损失函数, 将 SVR 的二次规划问题转变为求解线性方程问题, 可以有效的增加回归精度和计算速度^[15]。

训练样本集为 $D = \{ (x_i, y_i) \mid i = 1, 2, \dots, n \}$, 其中 x_i 为输入数据, y_i 为输出数据。

LSSVR 的优化问题为

$$\min_{\omega, b, \varepsilon} J(\omega, \varepsilon) = \frac{1}{2} \|\omega\|^2 + \frac{1}{2} C \sum_{i=1}^n \varepsilon_i^2 \quad (1)$$

约束条件为

$$y_i = \omega^T \Phi(x_i) + b + \varepsilon_i \quad (2)$$

式中 C - 惩罚参数; ω - 权向量; b - 阈值; ε_i - 误差。

引入拉格朗日乘子 a_i , 将问题转化为

$$L(\omega, b, \varepsilon, a) = J(\omega, \varepsilon) - \sum_{i=1}^n a_i (\omega^T \Phi(x_i) + b + \varepsilon_i - y_i) \quad (3)$$

对(3)式中 ω 、 b 、 ε_i 和 a_i 分别求偏导数, 可以得到

$$\begin{cases} \frac{\partial L}{\partial \omega} = 0 \\ \frac{\partial L}{\partial b} = 0 \\ \frac{\partial L}{\partial \varepsilon_i} = 0 \\ \frac{\partial L}{\partial a_i} = 0 \end{cases} \Rightarrow \begin{cases} \omega = \sum_{i=1}^n a_i \Phi(x_i) \\ \sum_{i=1}^n a_i = 0 \\ a_i = C \varepsilon_i \\ \omega^T \Phi(x_i) + b + \varepsilon_i = y_i \end{cases} \quad (4)$$

消除 ω 和 ε_i , 得到线性方程组

$$\begin{bmatrix} 0 & I^T \\ I & Z Z^T + c^{-1} I \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \quad (5)$$

其中 $\vec{y} = [y_1, y_2, \dots, y_n]^T$, $I = [1, 1, \dots, 1]^T$, $a = [a_1, a_2, \dots, a_n]^T$, $Z = [\Phi(x_1), \Phi(x_2), \dots, \Phi(x_n)]^T$, 引入核函数 $k(x_i, x)$ 替代 $\Phi(x_i) \cdot \Phi(x)$, 本文核函数采用径向基函数 $k(x_i, x) = e^{-\frac{\|x_i - x\|^2}{\delta^2}}$, $\delta > 0$ 。最终获得的 LSSVR 回归函数为

$$y = \sum_{i=1}^n a_i k(x_i, x) + b \quad (6)$$

2 遗传—最小二乘支持向量回归模型

为了使 LSSVR 的预测性能更好, 就需要获得较优的 C, δ, ε 参数组合, 这样才能使预测结果更加靠近实际真实值^[16], 因此需要利用遗传算法(Genetic Algorithm, GA)对参数 C, δ, ε 进行优化, 因为 GA 具有非常强的全局搜索能力, 能够有效的模拟生物进化过程中的自然选择和遗传变异, 可以有效的优化参数, 因此可以利用这一优点对 LSSVR 中的参数进行优化, 从而构造遗传—最小二乘支持向量回归模型(GA - LSSVR), 具体的操作步骤如图 1 所示。

(1) 选择 LSSVR 的训练样本数据集, 形成初始化 LSSVR 数据集。

(2) 利用 GA 对参数进行编码。在编码过程中采取实数编码方式, 可以有效的提高 GA 的求解速度和性能, 能够有效的求解复杂的连续参数优化问题。

(3) 对初始参数种群进行训练, 训练过程中调用 LSSVR 算法。

(4) 计算 GA 适应度。适应度函数定义为训练数据上的交叉验证后的均方误差平均值为:

$$F = \frac{1}{l} \sum_{i=1}^l (y_i - f_i)^2 \quad (7)$$

式中 l - 训练集的样本数; y_i - 第 i 个样本的实际值; f_i - 第 i 个样本的预测值。

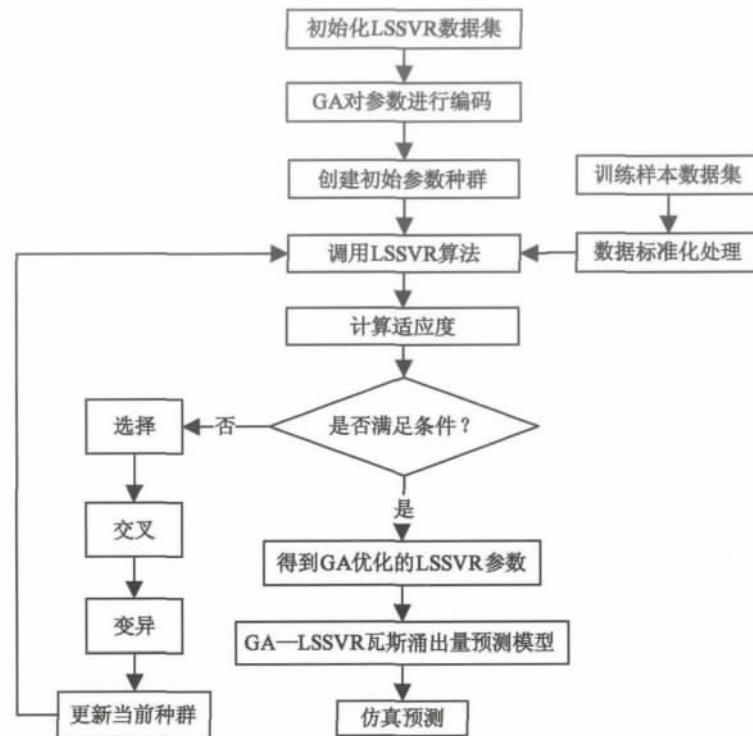


图1 GA—LSSVR流程图

Fig.1 GA—LSSVR process

(5) 进行遗传操作。根据计算出的 GA 的适应度对群体进行遗传操作,即选择、交叉、变异操作,对种群进行更新,然后进行步骤(3),调用 LSSVR 算法进行操作。

(6) 停止训练。如果满足 GA 训练的停止条件则停止训练,即得到了最优的 LSSVR 参数,如果不满足停止条件,则进行步骤(3),继续对参数进行优化和训练。

(7) 进行预测和仿真。将经过 GA 训练以后已经优化了的 LSSVR 参数进行预测,利用 GA-LSSVR 预测模型进行预测,并进行仿真预测,测验结果。

3 案例分析

3.1 样本数据的选择

在文献[2]中选取 15 个典型的回采工作面瓦斯涌出资料,其中前 12 个用于 GA 训练样本,另外 3 个作为 GA-LSSVR 算法的测试样本。原始煤层数据参数有 10 个^[2]: 瓦斯含量(m^3/t)、深度(m)、厚度(m)、工作面长度(m)、推进速度(m/min)、采出率(%)、邻近层瓦斯含量(m^3/t)、邻近层厚度(m)、开采深度(m)、瓦斯涌出量(m^3/t)。

15 个典型的回采工作面煤层赋存条件以及瓦斯涌出情况见表 1。

在进行预测时需要将原始数据进行标准化处理,可免除不同量纲对数据产生的不利影响,在进行标准化处理时采用统计学中的比例转换法,正向指标转换公式为 $x' = (x - x_{\min}) / (x_{\max} - x_{\min})$,逆向指标转换公式为 $x' = (x_{\max} - x) / (x_{\max} - x_{\min})$ 。标准化以后的样本数据见表 2。

3.2 参数选择与优化

利用 GA 训练样本参数,确定 GA 中种群规模、进化代数、交叉变异概率的数值以及各种参数的取值范围见表 3。

利用 matlab7.0 优化工具软件^[17],将表 1 中的 GA 训练样本中前 12 组数据中的瓦斯含量、深度、厚度、工作面长度、推进速度、采出率、邻近层瓦斯含量、邻近层厚度、层间距、开采深度 11 个参数向量作为训练函数的输入向量,瓦斯涌出量作为目标向量,输入到 GA 中进行训练,结束训练后即可得到 GA-LSSVR 预测模型的最优的参数,惩罚参数 C 为 20.309,径向基核函数参数 δ 为 10.02,不敏感损失函数 ε 为 0.135。

表 1 回采工作面煤层赋存条件及瓦斯涌出量原始数据

Tab. 1 The raw data of the gas emission quantity of the working face

编号	瓦斯含量 $/\text{m}^3 \cdot \text{t}^{-1}$	深度 $/\text{m}$	厚度 $/\text{m}$	工作面 长度 $/\text{m}$	推进速度 $/\text{m} \cdot \text{min}^{-1}$	采出率 $\%$	邻近层 瓦斯含 量 $/\text{m}^3 \cdot \text{t}^{-1}$	邻近层 厚度/ m	开采深 度/ m	瓦斯涌 出量 $\text{m}^3 \cdot \text{t}^{-1}$
1	1.92	408	2.0	155	4.42	0.96	2.02	1.50	1 825	3.34
2	2.15	411	2.0	140	4.16	0.95	2.10	1.21	1 527	2.97
3	2.14	420	1.8	175	4.13	0.95	2.64	1.62	1 751	3.56
4	2.58	432	2.3	145	4.67	0.95	2.40	1.48	2 078	3.62
5	2.40	456	2.2	160	4.51	0.94	2.55	1.75	2 104	4.17
6	3.22	516	2.8	180	3.45	0.93	2.21	1.72	2 242	4.60
7	2.80	527	2.5	180	3.28	0.94	2.81	1.81	1 979	4.92
8	3.35	531	2.9	165	3.68	0.93	1.88	1.42	2 288	4.78
9	3.61	550	2.9	155	4.02	0.92	2.12	1.60	2 325	5.23
10	3.68	563	3.0	175	3.53	0.94	3.11	1.46	2 410	5.56
11	4.21	590	5.9	170	2.85	0.79	3.40	1.50	3 139	7.24
12	4.03	604	6.2	180	2.64	0.81	3.15	1.80	3 354	7.80
13	2.43	450	2.2	160	4.32	0.95	2.00	1.70	1 966	4.06
14	3.16	544	2.7	165	3.81	0.93	2.30	1.80	2 207	4.92
15	4.62	629	6.4	170	2.80	0.80	3.35	1.61	3 456	8.04

表 2 回采工作面瓦斯涌出量原始数据标准化

Tab. 2 The raw data of the gas emission quantity of the working face

编号	瓦斯含量	深度	厚度	工作面 长度	推进 速度	采出率	邻近层瓦 斯含量	邻近层 厚度	开采 深度	瓦斯涌 出量
1	0.00	0.00	0.04	0.38	0.88	1.00	0.09	0.41	0.14	0.07
2	0.08	0.01	0.04	0.00	0.75	0.95	0.14	0.00	0.00	0.00
3	0.08	0.05	0.00	0.88	0.73	0.95	0.50	0.58	0.11	0.11
4	0.23	0.10	0.11	0.13	1.00	0.95	0.34	0.38	0.26	0.12
5	0.17	0.21	0.09	0.50	0.92	0.89	0.44	0.76	0.28	0.22
6	0.45	0.47	0.22	1.00	0.40	0.84	0.22	0.72	0.34	0.29
7	0.31	0.51	0.15	1.00	0.32	0.89	0.61	0.85	0.22	0.35
8	0.50	0.53	0.24	0.63	0.51	0.84	0.00	0.30	0.36	0.33
9	0.59	0.61	0.24	0.38	0.68	0.79	0.16	0.55	0.38	0.41
10	0.61	0.67	0.26	0.88	0.44	0.89	0.81	0.35	0.42	0.47
11	0.80	0.78	0.89	0.75	0.10	0.12	1.00	0.41	0.77	0.77
12	0.73	0.84	0.96	1.00	0.00	0.21	0.84	0.83	0.87	0.87
13	0.18	0.18	0.09	0.50	0.83	0.95	0.08	0.69	0.21	0.20
14	0.43	0.59	0.20	0.63	0.58	0.84	0.28	0.83	0.32	0.35
15	0.94	0.95	1.00	0.75	0.08	0.16	0.97	0.56	0.92	0.92

表 3 GA 参数选择

Tab. 3 The data of the GA

种群规模	进化代数	交叉概率	变异概率	惩罚参数 C 范围	径向基核函数 参数 δ 范围	不敏感损失 函数 ε 范围
20	200	0.4	0.01	[0, 100]	[0, 1 000]	[0, 1]

3.3 测试样本的瓦斯涌出量预测

经过训练可以得到最优的参数,利用最优参数可以对瓦斯涌出量的测试样本进行测试,可以

得到相关的预测值,然后对预测值进行适当的处理,将其反归一化。然后将 GA - LSSVR 预测模型得到结果与 SVR 模型进行预测得到的结果进行比较,得出比较的结果见表 4。

表4 不同预测模型下瓦斯涌出量预测值对比

Tab. 4 Comparison of the prediction values of gas emission quantity from different models

序号	实际瓦斯涌出量	SVR 模型		LSSVR 模型		GA - LSSVR 预测模型	
		预测值	预测误差 /%	预测值	预测误差 /%	预测值	预测误差 /%
13	4.06	5.29	30.30	4.90	20.69	4.01	1.23
14	4.92	5.50	11.79	5.30	7.72	5.00	1.63
15	8.04	10.06	25.12	9.06	12.69	7.90	1.74

4 结论

将遗传算法和最小二乘支持向量回归机相结合能够有效的进行瓦斯涌出量的预测,测试的结果可以看出 GA - LSSVR 预测效果是最好的,在模型数据的运行过程中,GA - LSSVR 预测模型计算速度更快,预测的误差更小,更接近于真实值,为以后的回采工作面瓦斯涌出量预测提供一个有效可行的方法。

参考文献:

- [1] 瓦斯通风防火安全研究所. 矿井瓦斯涌出量预测方法的发展与贡献[J]. 煤炭安全, 2003, 34(9): 10 - 13.
- [2] 吕伏, 梁冰, 等. 基于主成分回归分析法的回采工作面瓦斯涌出量预测[J]. 煤炭学报, 2012, 37(1): 113 - 116.
- [3] 王生全, 刘柏根, 张召召, 等. 遗传算法的 BP 网络模型进行瓦斯涌出量预测[J]. 西安科技大学学报, 2012, 32(1): 51 - 56.
- [4] 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.
- [5] 邵良杉, 张宇. 基于小波理论的支持向量机瓦斯涌出量的预测[J]. 煤炭学报, 2011, 36(1): 104 - 107.
- [6] 焦春林, 傅雪海, 葛燕燕, 等. 我国煤矿瓦斯中 H₂S 异常矿井的分布特征[J]. 黑龙江科技学院学报, 2013, 23(4): 375 - 377.
- [7] 王晓路, 刘健, 卢建军. 基于虚拟状态变量的卡尔曼滤波瓦斯涌出量预测[J]. 煤炭学报, 2011, 36(1): 80

- 85.

- [8] 孙林, 杨世元. 基于 LS - SVM 的回采工作面瓦斯涌出量预测[J]. 煤炭学报, 2008, 33(12): 1377 - 1381.
- [9] 陶云奇, 许江, 李树春. 改进的灰色马尔柯夫模型预测采煤工作面瓦斯涌出量[J]. 煤炭学报, 2007, 32(4): 391 - 395.
- [10] 张迎新, 孙浩. 基于 ANSYS 温度场的煤层瓦斯压力模拟可行性[J]. 黑龙江科技学院学报, 2013, 23(4): 333 - 336.
- [11] 谷松, 崔洪庆, 冯文丽. 基于灰色理论的小波神经网络对瓦斯涌出量的预测[J]. 煤炭学报, 2007, 32(9): 964 - 967.
- [12] 王丽华. 基于 GA - LSSVR 的图书馆文献资源需求预测[J]. 计算机应用与软件, 2011, 28(6): 225 - 227.
- [13] 曹庆奎, 任向阳, 刘开第. 矿井工作面瓦斯涌出量的未确知聚类研究[J]. 煤炭学报, 2006, 31(3): 337 - 341.
- [14] 郭志明, 陈立平, 黄文倩, 等. 近红外光谱结合 GA - LSSVR 分析烟草尼古丁含量[J]. 激光与电子学进展, 2012, 49(3): 1 - 5.
- [15] 赵清艳, 张超. 基于 GA - LSSVR 的铣削加工变形预测[J]. 组合机床与自动化加工技术, 2011, 12(4): 57 - 60.
- [16] 薛定宇, 陈阳泉. 高等应用数学问题的 MATLAB 求解[M]. 北京: 清华大学出版社, 2012.
- [17] SHAO LIANG SHAN, FU GUI XIANG. Disaster prediction of coal mine gas based on data mining [J]. Journal of Coal Science & Engineering (China), 2008, 14(3): 458 - 463.

(责任编辑 刘存英)