

基于 PCA-LWCA-LS-SVM 回采工作面瓦斯涌出量预测

张 瑞 李雅梅

(辽宁工程技术大学电气与控制工程学院 辽宁 葫芦岛 125105)

摘 要 回采工作面瓦斯涌出量各影响因素之间存在多重共线性,这会对预测造成不利影响。为降低这种影响并提高预测精度,提出将主成分分析法(PCA)与双层狼群算法(LWCA)优化的最小二乘支持向量机(LS-SVM)相耦合。该方法利用 PCA 对数据进行降维处理,再利用 LS-SVM 泛化能力强的特点来进行瓦斯涌出量的预测。为了提升 LS-SVM 的性能,将 LWCA 与 LS-SVM 结合,利用 LWCA 来优化 LS-SVM 的参数。采用该模型对实际中的回采工作面瓦斯涌出量进行预测,并与 LS-SVM 预测模型、PCA 与遗传算法优化的 LS-SVM 相耦合的预测模型进行比较,验证该模型的性能。实验表明:该模型预测结果中最大相对误差为 2.35%,最小相对误差为 0.26%,平均相对误差为 1.22%,预测精度得到显著提升。

关键词 主成分分析 最小二乘支持向量机 领导狼群算法 数据降维 回采工作面

中图分类号 TP391 TP712 **文献标识码** A **DOI**:10.3969/j.issn.1000-386x.2018.12.013

PREDICTION OF GAS EMISSION IN WORKING FACE BASED ON PCA-LWCA-LS-SVM

Zhang Rui Li Yamei

(Faculty of Electrical and Control Engineering, Liaoning Technical University, Huludao 125105, Liaoning, China)

Abstract There are multiple collinearities among the influencing factors of the gas emission in the working face, and it has a negative effect on the prediction. In order to reduce the effect and improve the accuracy of prediction, the principal component analysis(PCA) was coupled with the least square support vector machine(LS-SVM) optimized by the two-layer wolf colony algorithm(LWCA). The method used PCA to reduce the dimension of data. The volume of gas emission was predicted by using the generalization ability of LS-SVM. To improve the performance of LS-SVM, LWCA was combined with LS-SVM, which optimized the parameters of LS-SVM. The model was adopted to forecast the actual gas emission in the working face, and was compared with LS-SVM prediction model, PCA and LS-SVM prediction model optimized by genetic algorithm to verify the performance of the model. The experiment shows that the maximum relative error is 2.35%, the minimum relative error is 0.26%, the average relative error is 1.22% in the predicted results, and prediction accuracy is significantly improved.

Keywords PCA LS-SVM LWCA Data reduction Working face

0 引 言

瓦斯是引发煤矿安全事故的主要因素之一。瓦斯涌出量的精确预测,是预防煤矿瓦斯事故的重要前提。针对此项工作,已有众多学者提出了较为有效的预测方法,如:矿山统计法、瓦斯地质数学模型法、分源预测法等线性瓦斯预测方法,以及卡尔曼滤波法、神经网络

预测法、灰色系统法、主成分回归分析法、聚类分析法等非线性瓦斯预测方法。但上述预测模型也存在着一一定的缺陷,如:神经网络模型需要选择模型及参数,存在着收敛速度慢等缺点^[1];灰色理论预测当原始数据序列波动大并且信息过于分散时,预测精度将会降低^[2];聚类分析法中隶属度的确定受人为因素影响较大^[3]。且上述各种方法都不能很好地解决实际工作中普遍存在的变量之间多重共线性问题^[4]。

针对以上现状,提出基于主成分分析 PCA 与双层狼群算法 LWCA 优化最小二乘支持向量机 LS-SVM 相耦合的预测模型。该模型首先引入 PCA 对数据进行降维处理,保留绝大部分信息的同时,降低了数据的维度。然后利用 LS-SVM 求解速度快、泛化能力强的特点^[5]对瓦斯涌出量进行预测。为了进一步提升预测模型的性能,借鉴文献[6]利用 LWCA 优化 Elman 神经网络 ENN (Elman Neural Network) 参数的思想,采用 LWCA 来优化 LS-SVM 的参数,改善了传统的群体智能算法收敛速度慢,易陷入局部最优解等问题^[7],在简化了模型求解过程的同时提高了模型的预测精度。同时由于利用 LS-SVM 进行瓦斯涌出量的预测,改善了神经网络需要大量训练样本及训练时间长的缺点。

1 核心理论模型

1.1 PCA 算法

在瓦斯涌出量预测过程中,多个影响因素之间常具有多重共线性,此将会对模型的建立及其预测性能造成不利影响。利用 PCA 算法对其进行处理,可改善此问题。同时由于主成分贡献率较小的特征向量往往与噪声有关,因此也可起到一定的去噪效果^[8]。采用 PCA 处理后的数据,既保留了原数据的大部分信息,又能够降低数据的维度,从而降低问题的复杂性。

PCA 降维步骤如下:

将含有 k 个样本,且每个样本具有 n 个特征 x_1, x_2, \dots, x_n 的数据集表示为矩阵形式:

$$X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & & \vdots \\ x_{k1} & x_{k2} & \cdots & x_{kn} \end{bmatrix} = (X_1, X_2, \dots, X_n) \quad (1)$$

Step 1 对式(1)进行标准化处理:

$$x_{ij}^* = \frac{x_{ij} - \bar{x}_j}{\sqrt{\text{Var } x_j}} \quad i=1, 2, \dots, k; j=1, 2, \dots, n \quad (2)$$

式中: $\bar{x}_j = \frac{1}{k} \sum_{i=1}^k x_{ij}$, $\text{Var } x_j = \frac{1}{k-1} \sum_{i=1}^k (x_{ij} - \bar{x}_j)^2$, x_{ij} 为原始样本, x_{ij}^* 为标准值。

Step 2 计算样本相关系数矩阵:

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix} \quad (3)$$

式中: $r_{ij} = \frac{\text{Cov}(x_i^*, x_j^*)}{\sqrt{\text{var } x_i^*} \sqrt{\text{var } x_j^*}}$, $\text{Cov}(x_i^*, x_j^*)$ 为 x_i^* 与 x_j^*

之间的协方差。

Step 3 计算 R 的特征值 $(\lambda_1, \lambda_2, \dots, \lambda_n)$, 特征向量 $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{in})$, $i=1, 2, \dots, n$ 。

Step 4 利用步骤 3 中获取的特征向量 $\alpha_i = (\alpha_{i1}, \alpha_{i2}, \dots, \alpha_{in})$, 求得主成分:

$$F_i = \alpha_{i1}X_1 + \alpha_{i2}X_2 + \cdots + \alpha_{in}X_n \quad i=1, 2, \dots, n \quad (4)$$

Step 5 利用主成分累计贡献率确定需要采用的主成分个数:

$$L(p) = \frac{\sum_{i=1}^p \lambda_i}{\sum_{i=1}^n \lambda_i} \quad (5)$$

由此便可利用以上步骤所获得的主成分代替原始数据进行后续的处理。

1.2 LS-SVM 回归算法

LS-SVM 从损失函数着手,在其优化问题的目标函数中使用二范数,并用等式约束替换不等式约束。从而缩短了 SVM 的学习时间,具有求解速度快,泛化能力强^[9]的优势。优化目标为:

$$\min \frac{1}{2} \omega^T \omega + \frac{c}{2} \sum_{i=1}^n \zeta_i^2 \quad (6)$$

$$\text{s. t. } y_i = \omega^T \varphi(x_i) + b + \zeta_i$$

式中: c 为正则化参数,它可以在模型的复杂程度和训练误差之间做一个折衷选择,便于使所求的模型拥有较好的泛化能力。 ζ_i 为松弛变量。通过引入拉格朗日函数及 KKT 最优条件,得出 LS-SVM 的回归模型:

$$f(x) = \sum_{i=1}^n \alpha_i K(X, X_i) + b \quad (7)$$

式中: $k(x, x_i)$ 为核函数,本文选取学习能力较强的高斯核函数^[10]:

$$k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (8)$$

式中: σ 为核宽度。

1.3 LWCA 算法

当通过交叉验证 CV (Cross Validation) 的方式来取得 LS-SVM 参数 c 与 σ 的值时,不能保证所获取的参数为全局最优,从而不能充分发挥模型的性能。因此本文利用 LWCA 的全局寻优能力及收敛速度快等优点来获取 LS-SVM 回归模型的最优参数。

LWCA 是模拟狼群捕食过程而提出的一种算法,由于其采用胜者为王和强者生存的法则,使其具有良好的全局寻优能力及快速的收敛速度^[11],其规则如下:

1) 初始化狼群。

首先建立由 N 匹狼组成的狼群,令狼群中的个体随机分布在搜索空间内。

$$X_i = (x_{i1}, x_{i2}, \dots, x_{id}) \quad 1 \leq i \leq N, 1 \leq d \leq D$$

$$x_{id} = x_{\min} + \text{rand} \times (x_{\max} - x_{\min}) \quad (9)$$

式中: rand 为均匀分布在 $[0, 1]$ 中的随机数, x_{\max} 、 x_{\min} 为搜索空间的上下界。

2) 首狼的选取。

首先在狼群中选出适应值最优的 q 匹竞选狼,竞选狼在 h 个方向中的第 j 个点第 d 维的位置更新为:

$$y_{jd} = xx_{id} + \text{rand} \times \text{stepa} \quad (10)$$

式中: rand 为均匀分布在 $[-1, 1]$ 内的随机数; stepa 为搜索步长; xx_{id} 为竞选狼, $1 \leq j \leq h$ 。

3) 向首狼移动。

由于首狼最为接近猎物,所以参照首狼位置,其他狼向首狼移动,其他狼的位置更新公式为:

$$z_{id} = x_{id} + \text{rand} \times \text{stepb} \times (x_{id} - x_{id}) \quad (11)$$

式中: rand 为均匀分布于 $[-1, 1]$ 的随机数, stepb 为移动步长, x_{id} 为首狼位置, x_{id} 为其他狼当前的位置。

4) 种群包围。

首狼找到猎物后,通知其他狼对猎物进行包围:

$$x_i^{t+1} = \begin{cases} x_i^t & r_m < \theta \\ x_l + \text{rand} \times ra & r_m > \theta \end{cases} \quad (12)$$

式中: ra 为狼群的包围步长, x_i^t 为个体狼当前所在位置, x_l 为首狼的位置, r_m 为在 $[0, 1]$ 中产生的随机数, θ 为预设阈值。

5) 越界处理。

$$x_{id}^{t+1} = \begin{cases} x_{\max} & x_{id}^{t+1} > x_{\max} \\ x_{\min} & x_{id}^{t+1} < x_{\min} \end{cases} \quad (13)$$

在搜索的初期,为尽快寻找到全局最优的邻域,狼群采用较大的包围步长,在到达最优邻域的附近后,个体减小包围步长,以进行局部的搜索。步长计算公式如下:

$$ra(t) = ra_{\min} \times (x_{\max} - x_{\min}) \times \exp\left(\frac{\ln\left(\frac{ra_{\min}}{ra_{\max}}\right) \times t}{\max t}\right) \quad (14)$$

式中: $\max t$ 为最大迭代次数, ra_{\max} 为最大的包围步长, ra_{\min} 为最小包围步长。

狼群按照以上规则搜索猎物,每轮迭代完成后,采用淘汰适应值最差的 m 个个体,再以随机的方式生成 m 个个体的方式对狼群进行更新。此算法可以精确、快速地搜寻到全局最优解。

2 预测模型建立

首先利用 PCA 对数据进行降维处理。而后通过

LWCA 对 LS-SVM 回归模型的参数进行全局寻优以提升其性能。

2.1 适应度函数选取

以下式作为衡量狼群个体适应度的标准:

$$J(x_i) = -RMSE \quad (15)$$

式中: $RMSE$ 为模型的训练均方根误差,其定义如下误差越小,狼群个体的适应度越好。

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - y_i^d)^2} \quad (16)$$

式中: N 为样本的数量, y_i 为模型的实际输出值, y_i^d 为样本的期望输出值。

2.2 基于 PCA-LWCA-LS-SVM 的预测模型建立

在建立预测模型的过程中,以狼群的个体代表 LS-SVM 的正则化参数 c 与核参数 σ ,根据式(15)确定的适应值来衡量狼群位置的优劣。

模型的建立步骤如下:

Step 1 对狼群进行初始化,令其规模为 N ,最大迭代次数为 $\max t$,竞选狼个数 q ,搜索方向 h ,竞选狼的最大搜索次数 $\max dh$,搜索步长 stepa ,移动步长 stepb ,最大最小包围步长 ra_{\max} 、 ra_{\min} 及最差狼群个数 m ,通过式(9)初始化狼群的位置分布。

Step 2 初始化 LS-SVM 的正则化参数 c 与核参数 σ ,并将其映射至狼群个体。

Step 3 输入经过 PCA 降维的训练样本。

Step 4 利用式(15)计算狼群个体的适应值,狼群根据适应值进行迭代寻优。

Step 5 当模型达到要求的精度或达到最大迭代次数时停止训练。通过适应度最优的狼群的位置,获取 LS-SVM 的参数,从而获得预测模型。

3 实验

3.1 样本选取

选取煤层深度、煤层厚度、煤层倾角、开采层原始瓦斯含量、煤层间距、采高、临近层瓦斯含量、临近层厚度、层间岩性、工作面长度、推进速度、采出率、日产量,共 13 个对瓦斯涌出量影响较大的因素作为模型的输入变量。

采用沈阳某煤矿 2015 年间瓦斯涌出量的检测数据来验证本文提出的模型的性能。共选取 30 组数据作为样本集,其中前 20 组数据作为本文模型的训练样

本集,其余 10 组作为测试样本。

3.2 数据降维

利用 spss 软件对现场获取的数据进行 PCA 降维处理,将所得数据列于表 1、表 2。

表 1 特征值累积贡献率

成分	合计	贡献率/%	累计/%
1	7.828	60.213	60.213
2	2.150	16.542	76.755
3	1.226	9.432	86.187
4	0.731	5.620	91.807
5	0.465	3.576	95.383
6	0.233	1.793	97.176
7	0.148	1.139	98.315
8	0.097	0.742	99.057
9	0.069	0.535	99.592
10	0.034	0.258	99.850
11	0.01	0.074	99.923
12	0.006	0.046	99.970
13	0.004	0.030	100.000

表 2 成分矩阵

成分	1	2	3
V_1	0.821	0.492	0.098
V_2	0.988	0.048	0.028
V_3	-0.310	0.199	0.859
V_4	0.875	0.403	0.000
V_5	-0.061	-0.940	0.026
V_6	0.992	0.054	0.031
V_7	0.733	0.147	0.256
V_8	0.266	-0.091	0.836
V_9	-0.379	-0.499	-0.540
V_{10}	0.376	0.635	0.440
V_{11}	-0.844	-0.383	-0.151
V_{12}	-0.983	-0.007	-0.041
V_{13}	0.955	0.173	0.069

注: $V_1 \sim V_{13}$ 分别代表 $X_1 \sim X_{13}$ 各变量所对应提取的成分矩阵将所得降维后的样本集列于表 3

由于前三个主成分的累积贡献率为 86.187%, 大于 85%, 根据主成分选取原则^[12], 选取前三个主成分。

表 3 降维后样本集

序号	F1	F2	F3	瓦斯涌出量
1	-3.414 5	-1.769 3	0.935 7	2.91
2	-2.165 6	1.095 8	-0.209 9	3.52
3	-2.670 2	-0.948 1	-0.776 2	3.62
4	-2.214 3	0.801 3	2.389 5	4.13
5	-0.295 1	1.507 2	-0.498 1	4.60
6	-0.087 8	3.125 6	1.195 9	4.94
7	-0.889 2	-0.665 0	-2.096 8	4.78
8	-0.525 1	0.000 7	-0.304 8	5.25
9	1.145 1	0.963 3	-1.790 7	5.56
10	3.405 4	-1.756 2	-0.060 3	7.26
11	4.212 2	-0.873 1	0.580 0	7.80
12	-3.606 4	-1.206 2	0.190 6	3.34
13	-3.775 2	-2.504 3	-0.496 7	2.97
14	-2.147 7	0.712 4	0.299 7	3.56
15	-2.776 1	-1.188 3	-0.239 6	3.62
16	-1.800 2	1.530 1	1.525 3	4.17
17	-0.135 3	1.625 6	-0.838 2	4.60
18	0.336 2	3.754 0	-0.206 5	4.92
19	-0.834 8	-0.673 2	-1.991 4	4.78
20	-0.712 2	-0.008 9	-0.713 3	5.23
21	0.915 3	1.018 0	-1.772 5	5.56
22	3.385 8	-2.094 0	-0.071 9	7.24
23	4.073 7	-0.614 1	0.460 4	7.80
24	3.772 6	-1.106 5	0.573 8	7.68
25	4.943 9	0.379 8	0.991 8	8.51
26	4.225 0	-0.178 0	0.073 9	7.95
27	-2.327 2	0.524 0	0.132 8	4.06
28	-0.620 7	0.531 9	-0.403 4	4.92
29	4.242 9	-0.538 4	0.896 1	8.04
30	-3.660 6	-1.446 1	2.225 0	3.38

3.3 仿真实验

将测试样本应用于 PCA-LWCA-LS-SVM 预测模型中。初始化狼群,经多次实验,最终狼群算法的参数的设置如表 4 所示。

表 4 狼群算法参数设置

N	$\max t$	q	h	$\max dh$	$stepa$	$stepb$	m	θ
200	800	5	4	15	1.5	0.9	5	0.2

利用 MATLAB 软件对本文提出的模型进行仿真实验,表 3 中测试样本的 $\{F_1, F_2, F_3\}$ 对应模型的输入,将所获得预测结果列于表 5。

表 5 PCA-LWCA-LS-SVM 预测结果

序号	实际值	预测值	相对误差/%
1	5.56	5.64	1.439
2	7.24	7.35	1.519
3	7.80	7.70	1.282
4	7.68	7.66	0.26
5	8.51	8.71	2.35
6	7.95	8.02	0.881
7	4.06	4.00	1.478
8	4.92	4.99	1.422
9	8.04	7.96	0.995
10	3.38	3.40	0.592

为进一步检验文中所提模型性能,将其与 LS-SVM 预测模型、PCA 与遗传算法优化的 LS-SVM 相耦合的预测模型进行对比。各模型获得的预测结果相对误差见图 1。

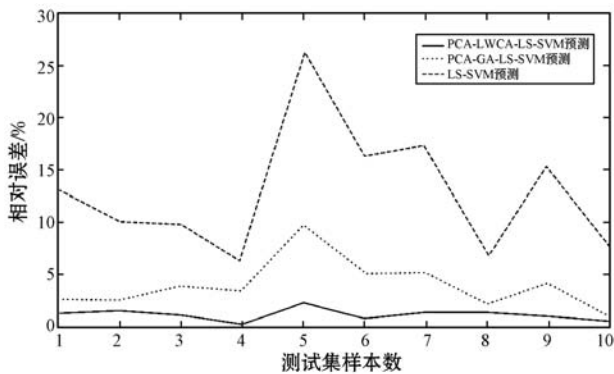


图 1 预测结果相对误差

取三种模型预测的最大相对误差、最小相对误差、平均相对误差,列于表 6。

表 6 预测结果比较

预测模型	最大相对误差	最小相对误差	平均相对误差
LS-SVM	26.05%	6.32%	13.07%
PCA-GA-LS-SVM	9.83%	0.92%	4.08%
PCA-LWCA-LS-SVM	2.35%	0.26%	1.22%

以上结果表明,PCA-LWCA-LS-SVM 预测模型预测精度高、泛化能力强,可以有效地预测回采工作面瓦斯涌出量。

4 结 语

本文提出的基于 PCA-LWCA-LS-SVM 的瓦斯预测模型,利用主成分分析法对高维的原始数据进行降维处理,提取出数据的主要信息,同时缓解了瓦斯涌出量影响因素间的多重共线性对模型带来的不利影响。然后利用 LWCA 对 LS-SVM 的参数进行全局寻优。该方法在简化模型求解过程的同时,又提高了模型的性能。采用实际工程中获取的数据对该模型进行验证,结果显示该模型具有良好的泛化能力及较高的预测精度,可有效地对瓦斯涌出量进行预测。

参 考 文 献

- [1] 程加堂,艾莉,熊燕. 基于 IQPSO-BP 算法的煤矿瓦斯涌出量预测[J]. 矿业安全与环保, 2016, 43(4):38-41.
- [2] 吴龙山. 基于缓冲算子的灰色瓦斯涌出量预测研究[J]. 图书情报导刊, 2011, 21(30):173-176.
- [3] 李艳川,周新华,董子文,等. 基于样本密度的非监督动态改进 FCM 算法[J]. 世界科技研究与发展, 2016, 38(1):20-25.
- [4] 皮子坤,贾宝山,李宗翔. 基于 SPSS 的采煤工作面瓦斯涌出量预测模型研究[J]. 计算机工程与应用, 2016, 52(15):228-232.
- [5] Lu X, Zou W, Huang M. Robust spatiotemporal ls-svm modeling for nonlinear distributed parameter system with disturbance[J]. IEEE Transactions on Industrial Electronics, 2017, 64(10):8003-8012.
- [6] 付华,司南楠,鲁俊杰,等. 基于 bi-LWCA-ENN 煤与瓦斯突出危险性预测[J]. 传感技术学报, 2016, 29(8):1222-1228.
- [7] 陈超,宣士斌,雷红轩. 基于狼群算法与二维最大熵的图像分割[J]. 计算机工程, 2018, 44(1):233-237.
- [8] Lee J, Choe Y. Robust pca based on incoherence with geometrical interpretation[J]. IEEE Transactions on Image Processing, 2018, 27(4):1939-1950.
- [9] 李灵光,王爱民,王小龙. 航空铝合金三维端铣表面粗糙的 LS-SVM 控制研究[J]. 机械设计与制造, 2015(3):256-259.
- [10] 王华忠,俞金寿. 核函数方法及其模型选择[J]. 江南大学学报(自然科学版), 2006, 5(4):500-504.
- [11] 焦钰,王建群,贾洋洋. 基于狼群算法的水电站优化调度模型参数优选[J]. 南水北调与水利科技, 2017, 15(2):58-64.
- [12] 常玉清,王姝,王福利,等. 基于多 PCA 模型的过程监测方法[J]. 仪器仪表学报, 2014, 35(4):901-908.