

王彦彬.基于 PCA-PSO-ELM 的瓦斯涌出量预测[J].湖南科技大学学报(自然科学版),2020,35(4): 1-9. doi:10.13582/j.cnki.1672-9102.2020.04.001

Wang Y B. Prediction of Gas Emission Based on PCA-PSO-ELM[J]. Journal of Hunan University of Science and Technology (Natural Science Edition), 2020,35(4):1-9. doi:10.13582/j.cnki.1672-9102.2020.04.001

基于 PCA-PSO-ELM 的瓦斯涌出量预测

王彦彬*

(辽宁工程技术大学 工商管理学院, 辽宁 葫芦岛 125105)

摘要:为了更加准确有效地预测瓦斯涌出量,提出采用主成分分析结合粒子群算法、极限学习机的瓦斯涌出量预测方法,其中极限学习机中隐含层节点数量及激活函数的类型由粒子群算法进行组合优化.实验综合考虑影响回采工作面瓦斯涌出量的 13 个因素对沈阳某煤矿历史数据进行分析,首先采用主成分分析对数据进行降维,消除指标数据之间的相关性,将降维后的数据划分为训练集和测试集 2 部分,设计了粒子群算法的惯性权重,并由粒子群算法结合十折交叉验证对极限学习机的 2 个参数进行优化,选择最优参数组合建立预测模型,通过对测试集瓦斯涌出量进行预测,其均方误差为 0.108 3,优于采用极限学习机及随机森林的预测结果.

关键词:瓦斯涌出量预测;主成分分析;粒子群算法;极限学习机;十折交叉验证

中图分类号:TD712.5 **文献标志码:**A **文章编号:**1672-9102(2020)04-0001-09

Prediction of Gas Emission Based on PCA-PSO-ELM

Wang Yanbin

(College of Business Administration, Liaoning Technical University, Huludao 125105, China)

Abstract: In order to predict gas emission accurately and effectively, a method of using principal component analysis (PCA), particle swarm optimization (PSO) and extreme learning machine (ELM) was presented, in which the number of hidden layer neurons and the type of excitation functions in ELM were optimized by PSO. In the experiment, 13 main influencing factors were considered and gas emission data of a coal mine in Shenyang were analyzed. First, PCA was used to reduce the data dimensions, eliminate the correlations of the data. Then the result was divided into training set and testing set, and the inertia weight of PSO was designed, meanwhile the two parameters of ELM were optimized by PSO with 10-fold cross-validation. After that, the prediction model was built with the optimized parameters. Finally, the gas emission of testing set samples were predicted, and the mean square error (MSE) was 0.108 3, which was better than the prediction result by ELM and random forest.

Keywords: gas emission prediction; principal component analysis; particle swarm optimization; extreme learning machine; 10-fold cross-validation

瓦斯事故严重威胁着煤矿的安全生产,在煤矿开采的过程中,由瓦斯引起的安全事故会造成大量的人员伤亡和财产损失,而提高瓦斯涌出量预测精度是实现瓦斯有效管理和保障煤矿安全生产的基础工作^[1].

对瓦斯涌出量进行预测的方法主要有矿山统计法和分源瓦斯预测法.矿山统计法受开采条件和地质

收稿日期:2019-03-12

基金项目:国家自然科学基金资助项目(71771111)

*通信作者,E-mail:65256016@qq.com

条件的影响较大,因此,矿山统计法的应用受到一定的限制;而分源预测法涉及的参数较多,部分参数的取值具有随意性,预测结果具有不确定性^[2].此外,随着计算机技术的发展,出现了一些新的瓦斯涌出量预测方法,如基于灰色系统理论的方法^[3]、主成分分析结合多步线性回归法^[4]、EMD-SVM法^[5]、EMD-PSO-SVM法^[1]、LLE结合BA-Elman的方法^[6]、蚁群粒子群混合算法优化LS-SVM的方法^[7]、BP神经网络方法^[8]等.采用上述方法对瓦斯涌出量进行预测取得了大量研究成果,但是由于矿井瓦斯涌出量受到煤层瓦斯含量、煤层厚度、煤层倾角、煤层埋藏深度等诸多因素的影响,且各影响因素之间呈现出复杂的非线性、相关性特征,为瓦斯涌出量的预测带来了较大影响^[6].此外,上述方法没有考虑所建模型的过拟合问题,因此,有必要继续探索新的方法对瓦斯涌出量进行预测.

主成分分析是一种常见的数据简化方法,能有效地减少数据维数及冗杂信息的影响^[9].极限学习机是黄广斌等^[10]提出的一种智能优化算法,具有泛化性能好、学习速度快、设置参数少的优点,目前很多研究集中在采用优化算法对极限学习机的输入权值和隐含层偏差进行优化,然而极限学习机的性能主要受隐含层神经元个数及激活函数的影响,且当隐含层神经元个数较多时,需要优化的参数较多.粒子群算法是一种性能良好的全局随机搜索算法,可以对参数进行组合优化.而十折交叉验证通过把数据集分为10份,轮流选择其中的9份作为训练集,另外1份作为测试集,重复10次实验,可以保证建立的模型具有良好的泛化性能.因此,本文提出采用主成分分析消除数据之间的相关性、降低数据维度,采用粒子群算法结合十折交叉验证对极限学习机中的隐含层神经元个数及激活函数类型进行优化的方法建立瓦斯涌出量预测模型,并验证其有效性.

1 理论与算法

1.1 主成分分析

主成分分析可以对多个指标进行线性变换,将其转换为线性无关的综合指标,各综合指标即为主成分,根据主成分累计贡献率选取贡献率大的主成分代替原指标实现对数据的降维.

假设样本 \mathbf{X}' 包括 p 项指标,令 $\mathbf{X}' = [\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_p]^T$, $\mathbf{A} = \begin{pmatrix} a_{11} & \dots & a_{1p} \\ \vdots & \vdots & \vdots \\ a_{p1} & \dots & a_{pp} \end{pmatrix}$ 为系数矩阵,则综合指标 $\mathbf{Y} =$

$[y_1, y_2, \dots, y_p]^T = \mathbf{A}\mathbf{X}'$, 且满足以下条件:

- 1) 系数矩阵 \mathbf{A} 中, $a_{i1}^2 + a_{i2}^2 + \dots + a_{ip}^2 = 1 (i = 1, 2, \dots, p)$;
- 2) 综合指标 $y_i (i = 1, 2, \dots, p)$ 之间线性无关;
- 3) 综合指标 $y_i (i = 1, 2, \dots, p)$ 的方差依次减少.

主成分分析的主要步骤如下^[11]:

- 1) 计算样本均值及协方差矩阵;
- 2) 求解协方差矩阵的特征值并按照从大到小的顺序排列,进而求解单位正交特征向量;
- 3) 根据特征向量写出主成分表达式,计算各主成分的方差贡献率并根据累计方差贡献率大于85%的原则选择 $l (l < p)$ 个主成分,实现对原始数据的降维.

1.2 极限学习机

极限学习机是根据摩尔-彭罗斯(MP)广义逆矩阵理论提出的一种采用典型的单隐含层前馈神经网络结构的学习算法,其基本原理如下^[10,12]:假设输入层有 m 个神经元、隐含层有 M 个神经元、输出层有 n 个神经元, $\omega_{ji} (1 \leq j \leq M, 1 \leq i \leq m)$ 、 $\beta_{jk} (1 \leq j \leq M, 1 \leq k \leq n)$ 分别为输入层与隐含层及隐含层与输出层的连接权值, $b_j (1 \leq j \leq M)$ 为隐含层神经元的阈值,为极限学习机的激活函数.设有 N 个样本 $(\mathbf{x}_i, \mathbf{t}_i)$, $1 \leq i \leq N$, 其中 $\mathbf{x}_i = [x_{i1}, x_{i2}, \dots, x_{im}]^T \in \mathbf{R}^m$ 为输入数据, $\mathbf{t}_i = [t_{i1}, t_{i2}, \dots, t_{in}]^T \in \mathbf{R}^n$ 为输出数据,极限学习机的网络结构如图1所示.

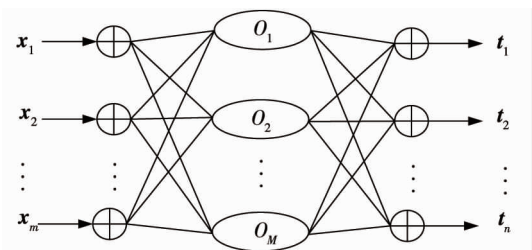


图1 极限学习机的网络结构

极限学习机可以用数学表达式表达如式(1)所示.

$$\sum_{j=1}^M \beta_j^T g(\omega_j x_i + b_j) = o_i, i = 1, \dots, N. \quad (1)$$

式中: $\beta_j = [\beta_{j1}, \beta_{j2}, \dots, \beta_{jm}]$; $\omega_j = [\omega_{j1}, \omega_{j2}, \dots, \omega_{jm}]$; o_i 为第 i 个样本的计算输出.

当隐含层节点数和训练集样本数接近时,极限学习机的代价函数为

$$E(S, \beta) = \sum_{i=1}^N o_i - t_i < \varepsilon, (\varepsilon > 0). \quad (2)$$

式中: S 为包含输入层与隐含层的权值及隐含层的阈值, $S = (\omega_j, b_j), j = 1, 2, \dots, M$; $\beta = [\beta_1, \beta_2, \dots, \beta_M]^T$; t_i 为第 i 个样本对应的真实值.

极限学习机就是为了找到最优组合 S, β , 使 $E(S, \beta)$ 的值最小.

求解过程可以改写为 $H\beta = T$, 其中 H 为隐含层输出矩阵, T 为极限学习机的输出矩阵, H, β, T 的定义如式(3)~式(5)所示.

$$H(\omega_1, \dots, \omega_M, b_1, \dots, b_M, x_1, \dots, x_N) = \begin{bmatrix} g(\omega_1 x_1 + b_1) & \dots & g(\omega_M x_1 + b_M) \\ \dots & \dots & \dots \\ g(\omega_1 x_N + b_1) & \dots & g(\omega_M x_N + b_M) \end{bmatrix}_{N \times M}; \quad (3)$$

$$\beta = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_M \end{bmatrix}_{M \times n}; \quad (4)$$

$$T = \begin{bmatrix} t_1^T \\ \vdots \\ t_N^T \end{bmatrix}_{N \times n}. \quad (5)$$

当激活函数 $g(x)$ 无限可微时, H 中的参数 ω_j 和 $b_j (j=1, 2, \dots, M)$ 可以在训练初始时刻根据隐含层神经元的个数随机生成, 训练过程中不需要继续进行调整, 而连接权值 β 可以通过求解 $\min_{\beta} H\beta - T$ 获得, 其解为 $\hat{\beta} = H^+ T$, 其中 H^+ 是输出矩阵 H 的 Moore-Penrose 广义逆矩阵.

1.3 PSO 优化 ELM 参数

极限学习机的效果主要受隐含层神经元个数及激活函数的影响, 为此, 引入粒子群算法对隐含层神经元个数及激活函数的类型进行优化.

粒子群算法是 1995 年 Kennedy 与 Eberhart 受鸟类觅食行为启发而提出的一种全局优化算法^[13-14], 其中粒子为待优化问题的潜在解, 每个粒子具有位置信息和速度信息, 设在 t' 时刻粒子的位置信息为 $x'_i(t')$ 、粒子速度为 $v_i(t')$, 则粒子按照式(6)进行位置和速度的更新.

$$\begin{cases} v_i(t' + 1) = \omega' v_i(t') + c_1 r_1 (f(t') - x'_i(t')) + c_2 r_2 (g'(t') - x'_i(t')); \\ x'_i(t' + 1) = x'_i(t') + v_i(t' + 1). \end{cases} \quad (6)$$

式中: ω' 为惯性权重; c_1 和 c_2 为学习因子; r_1 和 r_2 为 $[0, 1]$ 范围内的随机数; $f(t')$ 为 t' 时刻粒子的个体极值; $g'(t')$ 为 t' 时刻种群的全局极值.

粒子群算法中的惯性权重 ω' 对算法的搜索能力起着重要的作用, 惯性权重较大时粒子群算法具有较强的全局搜索能力, 惯性权重较小时粒子群算法具有较强的局部搜索能力. 为了提高算法的搜索速度, 同时避免陷入早熟, 本文采用可变惯性权重的粒子群算法对极限学习机的参数进行寻优, 结合极限学习机在训练过程中的决定系数, 提出惯性权重表达式如式(7)所示.

$$\omega'(t') = \cos\left(\frac{\pi}{2p_{\text{avg}}(t')g'(t')}\right). \quad (7)$$

式中: $p_{\text{avg}}(t')$ 为 t' 时刻粒子的个体极值的平均值, 即根据每个粒子训练得到的决定系数的平均值; $g'(t')$

为 t' 时刻训练得到的全局最优决定系数.

起始时刻算法具有较大的全局搜索能力,随着算法的执行,惯性权重呈现出多样性的取值并最终趋于稳定.

1.4 瓦斯涌出量的预测流程

采用 PCA-PSO-ELM 方法对瓦斯涌出量进行预测的流程如图 2 所示.

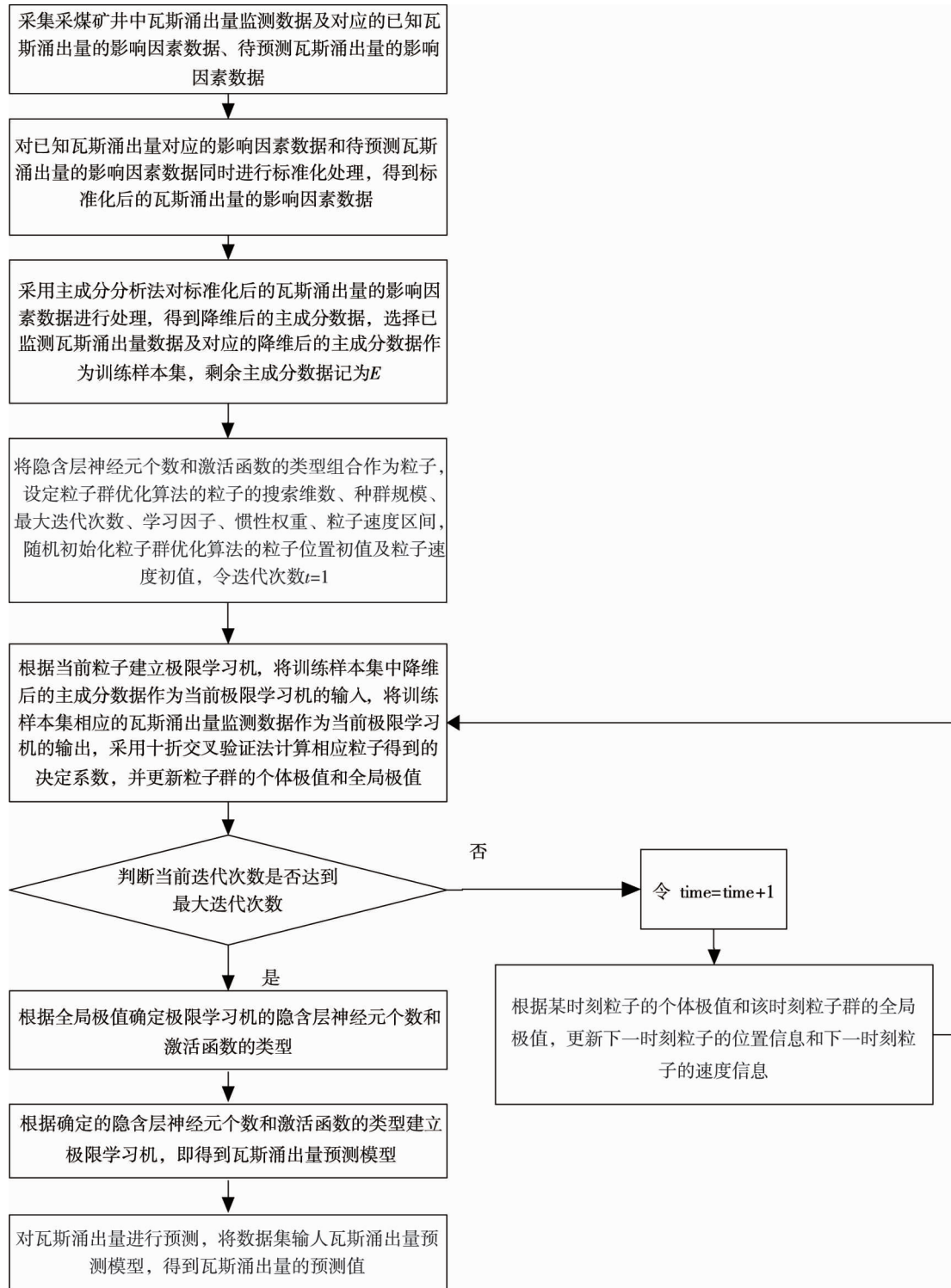


图 2 瓦斯涌出量预测流程

在上述过程中,首先对各指标数据进行标准化处理,继而对数据进行主成分分析,消除不同量纲对数据的影响,然后选择累计贡献率大于 85%的各主成分建立新的瓦斯涌出量样本集合,并将新集合划分为

训练集和测试集 2 部分,其中训练集用于对极限学习机的参数进行训练,训练过程中根据决定系数最大原则,由粒子群算法结合十折交叉验证对隐含层神经元个数及激活函数类型进行组合优化,针对每组参数组合,对随机生成的输入层与隐含层的权值及隐含层神经元的阈值加以固定以保证十折交叉验证结果的唯一性,选择决定系数最大的参数组合建立瓦斯涌出量预测模型,由该模型对测试集数据进行预测,最后输出预测结果.

2 实验

2.1 数据来源及指标选择

参考前人研究成果,实验选取文献[15]中沈阳某煤矿的 30 组数据进行分析,其中瓦斯涌出量的影响因素包括煤层深度 X_1 (m)、煤层厚度 X_2 (m)、煤层倾角 X_3 ($^\circ$)、开采层原始瓦斯含量 X_4 (m^3/t)、煤层间距 X_5 (m)、采高 X_6 (m)、邻近层瓦斯含量 X_7 (m^3/t)、邻近层厚度 X_8 (m)、层间岩性 X_9 、工作面长度 X_{10} (m)、推进速度 X_{11} (m/d)、采出率 X_{12} (%)、日产量 X_{13} (t/d),数据如表 1 所示.

表 1 瓦斯涌出量样本集

序号	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}	涌出量
1	412	2.5	8	2.12	24	2.0	2.10	1.53	4.78	140	4.16	0.960	1 528	2.91
2	423	1.5	11	2.14	17	1.4	2.55	1.62	4.75	180	4.14	0.950	1 751	3.52
3	436	2.3	10	2.53	14	2.2	2.40	1.48	4.91	145	4.67	0.945	274	3.62
4	459	2.4	15	2.45	24	2.3	2.42	1.78	4.75	155	4.57	0.944	2 104	4.13
5	511	2.8	13	3.24	14	2.4	2.21	1.72	4.78	180	3.45	0.930	2 241	4.60
6	515	2.3	17	2.85	17	2.5	2.77	1.87	4.51	170	3.25	0.940	1 973	4.94
7	556	2.7	9	3.37	13	2.5	1.88	1.42	4.85	165	3.68	0.932	2 287	4.78
8	550	3.1	12	3.67	15	2.9	2.32	1.65	4.83	155	4.01	0.920	2 352	5.25
9	590	3.0	11	3.68	12	3.6	3.11	1.46	4.53	175	3.53	0.940	2 410	5.56
10	581	5.2	8	4.31	17	5.9	3.47	1.57	4.76	170	2.80	0.797	3 131	7.26
11	611	6.7	9	4.05	16	6.7	3.15	1.80	4.70	175	2.64	0.812	3 354	7.80
12	408	2.0	10	1.92	20	2.0	2.02	1.50	5.03	155	4.42	0.960	1 825	3.34
13	411	2.0	8	2.15	22	2.0	2.10	1.21	4.87	140	4.16	0.950	1 527	2.97
14	420	1.8	11	2.14	19	1.8	2.64	1.62	4.75	175	4.13	0.950	1 751	3.56
15	432	2.3	10	2.58	17	2.3	2.40	1.48	4.91	145	4.67	0.950	278	3.62
16	456	2.2	15	2.40	20	2.2	2.55	1.75	4.63	160	4.51	0.940	2 104	4.17
17	516	2.8	13	3.22	12	2.8	2.21	1.72	4.78	180	3.45	0.930	2 242	4.60
18	527	2.5	17	2.80	11	2.5	2.81	1.81	4.51	180	3.28	0.940	1 979	4.92
19	531	2.9	9	3.35	13	2.9	1.88	1.42	4.82	165	3.68	0.930	2 288	4.78
20	550	2.9	12	3.61	14	2.9	2.12	1.60	4.83	155	4.02	0.920	2 352	5.23
21	563	3.0	11	3.68	12	3.0	3.11	1.46	4.53	175	3.53	0.940	2 410	5.56
22	590	5.9	8	4.21	18	5.9	3.40	1.50	4.77	170	2.85	0.795	3 139	7.24
23	604	6.2	9	4.03	16	6.2	3.15	1.80	4.70	180	2.64	0.812	3 354	7.80
24	607	6.1	9	4.34	17	6.1	3.02	1.74	4.62	165	2.77	0.785	387	7.68
25	634	6.5	12	4.80	15	6.5	2.98	1.92	4.55	175	2.92	0.773	3 620	8.51
26	640	6.3	11	4.67	15	6.3	2.56	1.75	4.60	175	2.75	0.802	3 412	7.95
27	450	2.2	12	2.43	16	2.2	2.00	1.70	4.84	160	4.32	0.950	1 996	4.06
28	544	2.7	11	3.16	13	2.7	2.30	1.80	4.90	165	3.81	0.930	2 207	4.92
29	629	6.4	13	4.62	19	6.4	3.35	1.61	4.63	170	2.80	0.803	3 456	8.04
30	401	2.0	10	1.87	25	2.4	2.14	1.78	5.12	150	4.52	0.950	1 855	3.38

2.2 数据的主成分分析

在进行主成分分析之前,首先借助 MATLAB 对原始数据采用 z-score 方法进行无量纲化处理,消除数据间由于指标单位的不同而存在的较大差异,然后对瓦斯涌出量的 13 个指标进行相关性分析,得到各指标间的相关系数矩阵如表 2 所示.

表2 样本相关系数矩阵

指标	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}	X_{11}	X_{12}	X_{13}
X_1	1.000	0.842	-0.025	0.972	-0.521	0.846	0.621	0.281	-0.587	0.586	-0.878	-0.800	0.666
X_2	0.842	1.000	-0.261	0.887	-0.123	0.991	0.686	0.299	-0.406	0.397	-0.854	-0.975	0.612
X_3	-0.025	-0.261	1.000	-0.130	-0.152	-0.260	-0.008	0.556	-0.441	0.277	0.091	0.262	0.039
X_4	0.972	0.887	-0.130	1.000	-0.449	0.888	0.630	0.210	-0.515	0.502	-0.855	-0.865	0.631
X_5	-0.521	-0.123	-0.152	-0.449	1.000	-0.126	-0.109	-0.029	0.363	-0.557	0.365	0.085	-0.155
X_6	0.846	0.991	-0.260	0.888	-0.126	1.000	0.726	0.294	-0.409	0.409	-0.858	-0.976	0.621
X_7	0.621	0.686	-0.008	0.630	-0.109	0.726	1.000	0.226	-0.628	0.513	-0.694	-0.701	0.436
X_8	0.281	0.299	0.556	0.210	-0.029	0.294	0.226	1.000	-0.318	0.435	-0.309	-0.310	0.300
X_9	-0.587	-0.406	-0.441	-0.515	0.363	-0.409	-0.628	-0.318	1.000	-0.607	0.597	0.375	-0.356
X_{10}	0.586	0.397	0.277	0.502	-0.557	0.409	0.513	0.435	-0.607	1.000	-0.685	-0.399	0.567
X_{11}	-0.878	-0.854	0.091	-0.855	0.365	-0.858	-0.694	-0.309	0.597	-0.685	1.000	0.836	-0.651
X_{12}	-0.800	-0.975	0.262	-0.865	0.085	-0.976	-0.701	-0.310	0.375	-0.399	0.836	1.000	-0.576
X_{13}	0.666	0.612	0.039	0.631	-0.155	0.621	0.436	0.300	-0.356	0.567	-0.651	-0.576	1.000

从表2中可见:各指标之间存在较强的相关性,其中煤层深度 X_1 与开采层原始瓦斯含量 X_4 的相关系数达到0.972,煤层厚度 X_2 与采高 X_6 的相关系数达到0.991.直接使用上述指标对瓦斯涌出量进行预测必然会增加预测模型的复杂度且不能保证预测结果的精度,因此,采用主成分分析消除指标间的相关性,并降低数据的维度,能够从理论上减少预测模型复杂度,提高预测精度.

经过主成分分析,得到前4个主成分的方差贡献率分别为0.570 143 162, 0.162 605 184, 0.094 412 701及0.058 019 067,累计方差贡献率达到88.5%,大于85%的累计方差贡献率,因此,选取前4个主成分进行后继分析,各主成分的系数见表3.

表3 主成分系数

指标	主成分			
	y_1	y_2	y_3	y_4
X_1	0.345 9	0.007 2	-0.176 4	0.101 7
X_2	0.337 0	-0.222 4	0.131 7	0.038 2
X_3	-0.010 9	0.616 3	0.247 4	-0.036 2
X_4	0.343 2	-0.088 8	-0.156 9	0.073 6
X_5	-0.135 7	-0.284 9	0.691 6	-0.182 5
X_6	0.339 8	-0.220 4	0.130 1	0.010 4
X_7	0.283 3	-0.009 1	0.134 0	-0.575 0
X_8	0.140 8	0.360 2	0.535 5	0.367 1
X_9	-0.233 6	-0.353 4	0.029 2	0.550 8
X_{10}	0.247 6	0.353 2	-0.161 3	0.100 9
X_{11}	-0.346 9	0.013 3	0.062 0	0.003 8
X_{12}	-0.330 4	0.229 0	-0.165 8	-0.017 8
X_{13}	0.263 6	0.030 2	0.087 0	0.410 8

根据各指标系数得到4个主成分的表达式如式(8)~式(11)所示.

$$y_1 = 0.345 9X_1 + 0.337 0X_2 - 0.010 9X_3 + 0.343 2X_4 - 0.135 7X_5 + 0.339 8X_6 + 0.283 3X_7 + 0.140 8X_8 - 0.233 6X_9 + 0.247 6X_{10} - 0.346 9X_{11} - 0.330 4X_{12} + 0.263 6X_{13}; \quad (8)$$

$$y_2 = 0.007 2X_1 - 0.222 4X_2 + 0.616 3X_3 - 0.088 8X_4 - 0.284 9X_5 - 0.220 4X_6 - 0.009 1X_7 + 0.360 2X_8 - 0.353 4X_9 + 0.353 2X_{10} + 0.013 3X_{11} + 0.229 0X_{12} + 0.030 2X_{13}; \quad (9)$$

$$y_3 = -0.176 4X_1 + 0.131 7X_2 + 0.247 4X_3 - 0.156 9X_4 + 0.691 6X_5 + 0.130 1X_6 + 0.134 0X_7 + 0.535 5X_8 + 0.029 2X_9 - 0.161 3X_{10} + 0.062 0X_{11} - 0.165 8X_{12} + 0.087 0X_{13}; \quad (10)$$

$$y_4 = 0.1017X_1 + 0.0382X_2 - 0.0362X_3 + 0.0736X_4 - 0.1825X_5 + 0.0104X_6 - 0.5750X_7 + 0.3671X_8 + 0.5508X_9 + 0.1009X_{10} + 0.0038X_{11} - 0.0178X_{12} + 0.4108X_{13}. \quad (11)$$

其中第1个主成分 y_1 主要受煤层深度 X_1 、煤层厚度 X_2 、开采层原始瓦斯含量 X_4 、采高 X_6 、邻近层瓦斯含量 X_7 的影响较大,可以将其命名为开采条件因子;第2个主成分 y_2 主要受煤层倾角 X_3 、工作面长度 X_{10} 、采出率 X_{12} 的影响较大,其中受煤层倾角 X_3 的影响最大,可以将该主成分命名为工作面信息因子;第3个主成分 y_3 主要受煤层间距 X_5 、邻近层厚度 X_8 的影响较大,可以将该主成分命名为邻近层因子;第4个主成分 y_4 主要受层间岩性 X_9 、日产量 X_{13} 的影响较大,其中,层间岩性 X_9 的系数最大,可以将其命名为层间岩性及日产量因子.上述结果为主成分分析结果,主要是为了减少影响因素之间的相关性,各个主成分为瓦斯涌出量各影响因素的线性组合,不改变原有各影响因素的客观存在.

2.3 瓦斯涌出量预测模型

采用降维后的样本数据进行瓦斯涌出量预测,选取其中前25组样本构成训练集,由粒子群算法对极限学习机进行训练,其中粒子群算法中粒子的搜索维数为 $D=2$,种群规模为 $G_M=5$,最大迭代次数 $G_{\max}=20$,学习因子 $c_1=1.5, c_2=1.5$,惯性权重由式(7)动态确定,粒子速度区间为 $[-10, 10]$,极限学习机中隐含层神经元个数为1~20的整数,因此粒子的位置取值设置为 $[1, 20]$,优化过程中将粒子的第一维数据进行四舍五入取整,极限学习机中激活函数的类型分别为sigmoid函数(取值为1)、sin函数(取值为2)、hardlim函数(取值为3),因此,将粒子的第二维数据锁定为1~3的整数,结合十折交叉验证计算所有粒子的决定系数,最终获得最优隐含层神经元个数为11,激活函数的类型为sigmoid函数,最优决定系数为0.9729,决定系数进化图如图3所示.

从图3的进化图可知:极限学习机的性能在第5代后达到最优并趋于稳定.粒子群算法中对应的各时刻惯性权重值如图4所示.

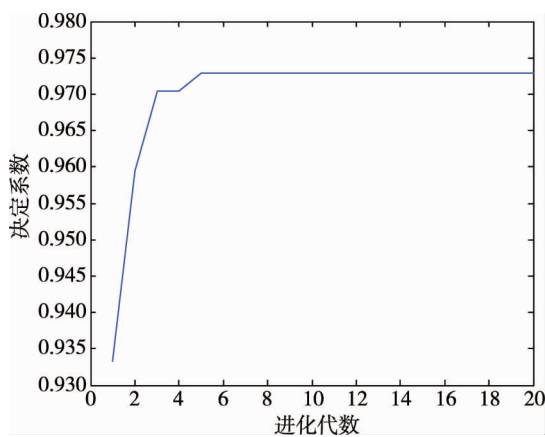


图3 决定系数进化

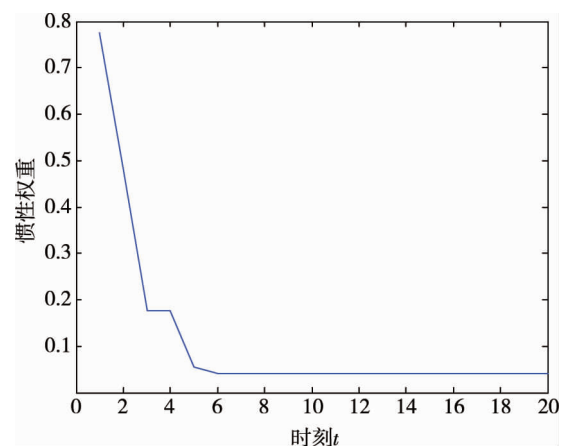


图4 惯性权重变化

其中粒子群算法的初始惯性权重为0.7748,在第5次迭代过程中惯性权重为0.0552,此时,对极限学习机进行优化得到最优决定系数为0.9729,此后各粒子个体极值趋于最优,相应的惯性权重值的取值达到最小值0.0404,并保持稳定不变.

优化后的极限学习机中输入层与隐含层权值 $W_i (1 < i < 4)$ 与隐含层阈值 B 如表4所示.

从表4可知:如果对输入层与隐含层权值及隐含层阈值进行优化,需要优化的参数较多,容易陷入局部极值且优化效率较低.

直接采用极限学习机根据原始数据中前25条数据建立预测模型,其中极限学习机中的激活函数选择常用的sigmoid函数,经过不断尝试,当隐含层节点数设为14时,十折交叉验证结果较优,此时模型的均方误差为3.1769,明显大于PCA-PSO-ELM方法所得结果的均方误差0.08089,表明本文方法能够建立泛化性能良好的预测模型.

表4 权值及阈值

序号	W_1	W_2	W_3	W_4	B
1	0.025 8	-0.057 3	-0.653 5	0.268 7	0.921 3
2	-0.079 0	-0.710 2	-0.827 8	0.724 8	0.542 0
3	-0.299 2	0.435 7	-0.213 3	-0.683 5	0.812 9
4	-0.809 9	0.323 4	0.608 7	0.202 4	0.166 4
5	-0.132 7	-0.136 3	-0.977 8	-0.764 8	0.320 4
6	0.418 5	-0.107 9	-0.533 8	0.252 2	0.657 9
7	-0.768 1	0.016 7	0.867 7	0.670 2	0.001 8
8	-0.843 8	0.056 2	-0.546 4	0.880 8	0.629 0
9	-0.261 5	0.145 8	0.571 9	-0.168 9	0.785 3
10	-0.932 7	-0.278 4	-0.178 5	-0.455 9	0.294 7
11	-0.615 7	-0.327 0	-0.761 2	0.856 0	0.789 7

2.4 瓦斯涌出量预测及评价

选择新样本集中的后5条数据构成测试集,采用建立的瓦斯涌出量预测模型对其进行预测,预测结果见表5.

表5 预测结果

序号	Y_1	Y_2	Y_3	Y_4	涌出量监测值	本文方法 预测结果	ELM 预测结果	随机森林 预测结果
1	4.125	-0.212	0.103	0.834	7.95	7.439 182 750	5.309 219 852	7.306
2	-2.210	0.620	0.139	0.835	4.06	3.928 962 138	5.309 219 852	3.803
3	-0.517	0.552	-0.395	1.416	4.92	5.100 203 741	5.309 219 852	4.931
4	4.125	-0.568	0.927	-0.572	8.04	8.071 567 957	5.311 172 977	7.630
5	-3.579	-1.269	2.291	1.196	3.38	2.900 457 544	5.308 243 290	3.963

其中,本文方法所得预测结果的均方误差为0.108 3,为了进行比较,直接采用极限学习机建立预测模型对原始数据集中的后5条数据进行预测,预测结果的均方误差为3.970 1,采用常用具有良好性能的随机森林方法对测试集瓦斯涌出量进行预测,预测结果的均方误差为0.197 8,这2种方法均高于本文方法所得的预测结果,表明所建立的PCA-PSO-ELM模型具有更好的预测性能.

3 结论

1)对影响瓦斯涌出量的指标数据进行标准化处理,消除了不同量纲的影响,经过主成分分析,选取了累计贡献率大于85%的4个主成分数据生成新的样本集合,实现了对数据的降维并消除了数据间的相关性.

2)采用粒子群算法结合十折交叉验证对极限学习机中隐含层神经元的个数及激活函数的类型进行组合优化,设计了粒子群算法的惯性权重,并根据优化结果建立了瓦斯涌出量预测模型,相对输入层与隐含层权值及隐含层阈值进行优化的方法减少了优化参数.

3)采用建立的预测模型对测试集中的瓦斯涌出量进行预测,预测结果的均方误差为0.108 3,低于直接采用极限学习机及采用随机森林对瓦斯涌出量进行预测的结果,表明本文方法的有效性.

4)在对瓦斯涌出量进行预测过程中,指标选择是一个重点,在今后的研究中要更广泛地分析影响瓦斯涌出量的相关指标,如残余瓦斯含量等.此外,本文采用其他文献中公开的数据验证了方法的有效性,未来应该将所提方法应用于实际矿山的瓦斯涌出量预测中,并不断对该方法进行完善.

参考文献:

- [1] 施式亮,李润求,罗文柯.基于 EMD-PSO-SVM 的煤矿瓦斯涌出量预测方法及应用[J].中国安全科学学报,2014,24(7):43-49.
- [2] 章立清,秦玉金,姜文忠,等.我国矿井瓦斯涌出量预测方法研究现状及展望[J].煤矿安全,2007(8):58-60.
- [3] 伍爱友,田云丽,宋译,等.灰色系统理论在矿井瓦斯涌出量预测中的应用[J].煤炭学报,2005,30(5):47-50.
- [4] 吕伏,梁冰,孙维吉,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.
- [5] 刘俊娥,安凤平,林大超,等.采煤工作面瓦斯涌出量的固有模态 SVM 建模预测[J].系统工程理论与实践,2013,33(2):505-511.
- [6] 付华,代巍.基于 LLE 与 BA-Elman 的瓦斯涌出量动态预测研究[J].传感技术学报,2016,29(9):1383-1388.
- [7] 付华,于翔,卢万杰.基于蚁群粒子群混合算法与 LS-SVM 瓦斯涌出量预测[J].传感技术学报,2016,29(3):373-377.
- [8] 高保彬,潘家宇.基于 PLS-BP 神经网络组合模型的回采工作面瓦斯涌出量预测[J].湖南科技大学学报(自然科学版),2015,30(4):14-20.
- [9] 施龙青,谭希鹏,王娟,等.基于 PCA_Fuzzy_PSO_SVC 的底板突水危险性评价[J].煤炭学报,2015,40(1):167-171.
- [10] Huang G B, Zhu Q Y, Siew C K. Extreme Learning Machine : Theory and Applications [J]. Neurocomputing, 2006,70(1/3):489-501.
- [11] 鲁金涛,李夕兵,宫凤强,等.基于主成分分析与 Fisher 判别分析法的矿井突水水源识别方法[J].中国安全科学学报,2012,22(7):109-115.
- [12] 高光勇,蒋国平.采用优化极限学习机的多变量混沌时间序列预测[J].物理学报,2012,61(4):37-45.
- [13] Eberhart R C, Kennedy J. Particle swarm optimization [C]//Proc of IEEE Int Conf on Neural Networks. Nagoya: IEEE Press, 1995: 39-43.
- [14] Kennedy J, Eberhart R C. A new optimizer using particles swarm theory [C]//Proc of 6th Int Symposium on Micro Machine and Human Science. Perth: IEEE Press, 1995:1942-1948.
- [15] 董晓雷,贾进章,白洋,等.基于 SVM 耦合遗传算法的回采工作面瓦斯涌出量预测[J].安全与环境学报,2016,16(2):114-118.