

宋金华,岳浩. 基于 PCA-PSO-LSSVM 的综合管廊投资估算方法[J]. 湖南科技大学学报(自然科学版), 2024, 39(1):36-44. doi:10.13582/j.cnki.1672-9102.2024.01.005

SONG J H, YUE H. On Investment Estimation Method of Comprehensive Pipe Gallery Based on PCA-PSO-LSSVM [J]. Journal of Hunan University of Science and Technology (Natural Science Edition), 2024, 39(1):36-44. doi:10.13582/j.cnki.1672-9102.2024.01.005

基于 PCA-PSO-LSSVM 的综合管廊投资估算方法

宋金华*, 岳浩

(河北工业大学 土木与交通学院, 天津 300401)

摘要:为有效地解决现有综合管廊投资估算方法的预测精度不高,且预测精度易受样本量大小、特征参数冗余或缺失等问题,构建一种将主成分分析法(PCA)与粒子群算法(PSO)优化最小二乘支持向量机(LSSVM)结合的综合管廊投资估算预测模型.采用 PCA 对影响综合管廊投资估算的特征参数进行降维,剔除噪声或冗余数据,以贡献率较大的主成分作为 LSSVM 的输入向量,综合管廊单公里造价作为 LSSVM 的输出向量;利用 PSO 对 LSSVM 的核函数参数 σ 与惩罚因子参数 C 进行寻优,建立基于 PCA-PSO-LSSVM 的综合管廊投资估算预测模型,并对测试集样本进行预测.预测结果显示:PCA-PSO-LSSVM 模型平均相对误差为 3.28%,满足投资决策阶段对投资估算预测误差的要求($\pm 10\%$),且与 PCA-LSSVM 模型、PSO-LSSVM 模型、GA-BP 模型和 GA-SVM 模型相比,预测精度分别提高了 67.29%,70.52%,48.13%和 38.60%.PCA-PSO-LSSVM 模型预测精度高,泛化性能优,可作为综合管廊投资估算的有效预测方法.

关键词:综合管廊;投资估算;主成分分析法;粒子群算法;最小二乘支持向量机

中图分类号:TU990.3;TP391 文献标志码:A 文章编号:1672-9102(2024)01-0036-09

On Investment Estimation Method of Comprehensive Pipe Gallery Based on PCA-PSO-LSSVM

SONG Jinhua, YUE Hao

(School of Civil and Communications, Hebei University of Technology, Tianjin 300401, China)

Abstract: In order to effectively solve the problems that the existing comprehensive pipeline corridor investment estimation methods are not high in prediction accuracy, and the prediction accuracy is easily affected by sample size, redundancy or lack of feature parameters, a method of Principal Component Analysis (PCA) and Particle Swarm Optimization (PSO) optimized least square support vector machine (LSSVM) combined with a comprehensive pipeline gallery investment estimation forecast model. It uses PCA to reduce the dimensionality of the characteristic parameters that affect the investment estimation of the comprehensive pipe gallery, eliminate noise or redundant data, use the principal component with a larger contribution rate as the input vector of LSSVM, and use the cost per kilometer of the comprehensive pipe gallery as the output vector of LSSVM. It uses PSO to optimize LSSVM's kernel function parameter σ and penalty factor parameter C , establishes a comprehensive pipeline corridor investment estimation prediction model based on PCA-PSO-LSSVM, and

predicts the test set samples. The prediction results show that the average relative error of the PCA-PSO-LSSVM model is 3.28%, which meets the requirement of investment estimation prediction error ($\pm 10\%$) in the investment decision-making stage, and is compatible with the PCA-LSSVM model, PSO-LSSVM model, and GA-BP model. Compared with the GA-SVM model, the prediction accuracy is improved by 67.29%, 70.52%, 48.13% and 38.60% respectively. The PCA-PSO-LSSVM model has high prediction accuracy and excellent generalization performance. It can be used as an effective prediction method for comprehensive pipeline gallery investment estimation.

Keywords: comprehensive pipeline corridor; investment estimation; principal component analysis method; particle swarm algorithm; least square support vector machine

地下综合管廊作为容纳多种公用管道或电缆的隧道,不仅有助于公用基础设施的集约化管理,而且对实现地下空间资源的可持续化发展具有重要作用^[1].但据相关研究成果^[2-4]显示,综合管廊初始建设成本约为传统直埋式管道铺设方式的1~2倍,高成本严重制约了其大幅度建设与城镇新型基础设施的布局.若能在综合管廊的可行性研究阶段精准预测出投资估算,不仅可为其投资决策提供有力依据,还可有效地控制项目成本与实现精细化造价管理.

针对综合管廊投资估算的预测问题,国内专家及学者提出了多种预测方法.传统的方法是以《城市地下综合管廊工程消耗量定额》(建标[2017]131号)和《城市综合管廊工程投资估算指标》(国办发[2014]27号)为依据,采用定额估算法^[5]和指标估算法^[6]进行综合管廊投资估算的预测.但这2种算法对指标的考虑过于笼统且计算繁杂,难以实现综合管廊投资估算的精准预测,也无法揭示综合管廊投资估算与影响因素间复杂的高维度非线性关系.随着人工智能的发展,出现了以机器学习算法为理论,利用计算机技术进行综合管廊投资估算的预测方法.胡庆国等^[7]采用GA-BP神经网络进行综合管廊投资估算预测,提高了预测精度,但在处理小样本、高维非线性与全局最优上仍存在不足;李芊等^[8]利用GA-SVM进行综合管廊投资估算预测,解决了GA-BP神经网络存在的问题,但该方法计算较为复杂且运行效率较低.

基于上述分析,本文将具有运算简单、收敛迅速且预测精度高等优点的最小二乘支持向量机^[9](LSSVM)引入综合管廊投资估算的预测,但考虑到其核函数参数 σ 与惩罚因子参数 C 会直接影响LSSVM的预测精度,因此,采用PSO对核函数参数 σ 与惩罚因子参数 C 进行寻优.此外,鉴于影响综合管廊投资估算的特征因素较多,故采用PCA进行特征因素降维,以剔除噪声或冗余因素;最终,构建基于PCA-PSO-LSSVM的综合管廊投资估算预测模型,并用已建综合管廊造价资料验证其有效性.

1 模型基础原理

1.1 最小二乘支持向量机

最小二乘支持向量机^[10](Least Square Support Vector Machine,简称LSSVM)是一种对标准支持向量机(SVM)进行拓展改进的新型机器学习算法,以等式约束代替以往的不等式约束问题,实现了拉格朗日乘子求解过程的简化,大幅降低了SVM的计算复杂度.

设有训练样本 $\{x_i, y_i\}_{i=1}^l, x_i \in \mathbf{R}^n, y_i \in \mathbf{R}, x_i, y_i$ 分别表示LSSVM的输入向量与输出向量,则可构建预测函数,如式(1)所示.

$$f(x) = \mathbf{w}^T \varphi(x) + b. \quad (1)$$

式中: \mathbf{w}^T 为函数权重变量; $\varphi(x)$ 为将输入样本映射到高维空间的非线性映射函数; b 为偏差系数.

将式(1)与结构风险函数相结合,可转化为如式(2)所示的优化问题.

$$\begin{cases} \min J(\mathbf{w}, \delta) = 1/2 \|\mathbf{w}\|^2 + C(1/2 \sum_{i=1}^l \delta_i^2); \\ \text{s.t. } y_i = \mathbf{w}^T \varphi(x) + b + \delta_i^2, i = 1, 2, \dots, l. \end{cases} \quad (2)$$

式中: C 为惩罚因子参数,且有 $C > 0$; δ_i 为松弛变量.在式(2)中引入拉格朗日乘子 α_i ,即可得式(3):

$$L(w, b, \delta, \alpha) = 1/2 \|w\|^2 + C(1/2 \sum_{i=1}^l \delta_i^2) - 1/2 \sum_{i=1}^l \alpha_i^2 (\mathbf{w}^T \varphi(x) + b + \delta_i - y_i). \quad (3)$$

根据 KKT 条件可知:

$$\begin{cases} w = \sum_{i=1}^l \alpha_i \varphi(x); \\ \sum_{i=1}^l \alpha_i = 0; \\ \alpha_i = C\delta_i; \\ \mathbf{w}^T \varphi(x) + b + \delta_i = y_i. \end{cases} \quad (4)$$

将非线性方程求解问题转化为线性求解问题:

$$\begin{bmatrix} 0 \\ y \end{bmatrix} = \begin{bmatrix} 0 & \mathbf{I}_l^T \\ \mathbf{I}_l & K + C^{-1}\mathbf{I}_l \end{bmatrix} \begin{bmatrix} b \\ \alpha \end{bmatrix}. \quad (5)$$

式中: $\mathbf{I}_l = [1, 2, \dots, l]^T$; K 为核函数, 本文采用径向基核函数作为 LSSVM 的核函数, 简称 RBF 核函数, 如式(6)所示.

$$K(x, x_j) = \exp\left(-\frac{\|x - x_j\|^2}{2\sigma^2}\right). \quad (6)$$

式中: σ 为 RBF 核函数的宽度参数. 最终 LSSVM 的预测模型可变换为

$$y(x) = \sum_{i=1}^l \alpha_i K(x, x_i) + b. \quad (7)$$

1.2 主成分分析法

主成分分析法^[11-12] (Principal Component Analysis, 简称 PCA) 是一种将原始高维且线性相关的变量通过线性变换生成一组低维、线性无关且包含绝大部分原始信息变量的方法, 其具体步骤为

1) 构造样本矩阵. 设样本量为 m , 每个样本均包含 n 个变量, 则可构造一个 $m \times n$ 型的矩阵, 记作 \mathbf{X}_{mn} :

$$\mathbf{X}_{mn} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{bmatrix}. \quad (8)$$

式中: $X_{ij} (i = 1, 2, \dots, m; j = 1, 2, \dots, n)$ 为第 i 个样本的第 j 个特征指标参数.

2) 特征参数标准化. 为克服特征参数间不同量纲的影响, 本文利用正态标准化法对其进行标准化.

$$\bar{X}_j = \frac{1}{m} \sum_{i=1}^m X_{ij}; \quad (9)$$

$$\delta_j = \sqrt{\frac{1}{m-1} \sum_{i=1}^m (X_{ij} - \bar{X}_j)^2}; \quad (10)$$

$$X_{ij}^* = \frac{X_{ij} - \bar{X}_j}{\delta_j}. \quad (11)$$

式中: X_{ij}^* 为标准化后的特征指标; \bar{X}_j 为特征指标 j 的均值; δ_j 为特征指标 j 的标准差.

3) 计算特征参数相关系数矩阵 $\mathbf{R}_{n \times n}$.

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

式中: $r_{ij} (i, j = 1, 2, \dots, n)$ 为经标准化处理后特征指标的相关系数, 且有 $r_{ij} = r_{ji}$.

4) 计算主成分累计贡献率.在求得特征参数相关系数矩阵 $R_{n \times n}$ 特征值 $\lambda_i (i = 1, 2, \dots, n)$ 的基础上,利用式(13)计算前 q 个主成分的累计贡献率 φ .

$$\varphi = \frac{\sum_{i=1}^q \lambda_i}{\sum_{i=1}^n \lambda_i}. \quad (13)$$

通常而言,当 $\varphi > 85\%$ 时,则取前 q 个主成分作为 LSSVM 的输入向量.

1.3 粒子群算法

粒子群算法^[13-15](Particle Swarm Optimization,简称 PSO)是一种基于鸟群觅食行为来寻求最优解空间的算法,具有运算简单、收敛迅速且参数少等优点.其基本思路是将优化问题的可能解视为空间中的一个粒子,每个粒子皆具有位置和速度 2 个特征,且拥有一个由被优化函数决定的适应度值,对初始化粒子群进行不断迭代直至产生最优解.PSO 算法的粒子速度与位置更新公式如式(14)所示.

$$V_{s+1} = \omega V_s + C_1 \text{rand}[P_{bs} - P_s] + C_2 \text{rand}[g_b - P_s]. \quad (14)$$

式中: V_{s+1} 为第 $s+1$ 个粒子的速度; ω 为惯性权重; C_1, C_2 为学习因子; $\text{rand} \in [0, 1]$; P_{bs} 为第 s 个粒子的最优位置; P_s 为第 s 个粒子的当前位置; g_b 为搜索到的整体最优值.

鉴于 LSSVM 的惩罚因子参数 C 与核函数参数 σ 为影响其模型性能和预测精度的主要参数,故本文以训练集样本的预测值的平均相对误差作为 PSO 的适应度函数,对惩罚因子参数 C 与核函数参数 σ 进行寻优,其寻优目标函数为

$$\min f(C, \sigma) = \frac{1}{t} \sum_{i=1}^t \frac{|T_i - Y_i|}{T_i}; \quad (15)$$

s.t. $C \in [C_{\min}, C_{\max}]$, $\sigma \in [\sigma_{\min}, \sigma_{\max}]$.

式中: Y_i, T_i 为第 i 个样本的预测值与真实值; t 为预测样本数; $f(C, \sigma)$ 为寻优目标函数; C_{\min}, C_{\max} 为惩罚因子参数的最小值与最大值; $\sigma_{\min}, \sigma_{\max}$ 为核函数参数的最小值与最大值.

2 模型的构建及预测

综合管廊投资估算的预测建立在 PCA 和 PSO-LSSVM 的基础上,基于 PCA-PSO-LSSVM 预测模型的构建步骤(具体如图 1 所示):

1) 数据降维与归一化.采用 PCA 对初始数据及特征指标进行降维与归一化处理,将满足累计方差贡献率 $\varphi > 85\%$ 的前 q 个主成分作为 LSSVM 的输入向量.

2) 参数寻优.利用训练样本对 LSSVM 进行训练,并结合 PSO 算法对其惩罚因子 C 和核函数参数 σ 进行寻优,以实现最优 PSO-LSSVM 模型.

3) 模型预测及效果评价.运用测试样本对构建的 PCA-PSO-LSSVM 预测模型试验,同时将其试验结果与 LSSVM, PSO-LSSVM, GA-BP 等其他模型的预测结果进行对比.

2.1 初始数据准备及特征指标量化

本文通过分析现有研究成果^[7,16]及已建综合管廊造价资料,共收集了 29 个样本.按表 1 中特征指标量化的方式对非数值型指标进行量化,量化后 29 个样本的原始数据如表 2 所示,随机选取 6 个作为测试样本,其余的作为训练样本.

2.2 提取特征指标的主成分

采用 PCA 对 13 个原始特征指标进行降维,13 个原始特征指标的相关系数矩阵如表 3 所示,各成分归一化后的特征值、贡献率及累计方差贡献率如表 4 所示.由表 4 可知:前 7 个主成分的累计方差贡献率为 $89.30\% > 85\%$,故取前 7 个主成分作为 LSSVM 的输入向量;依据主成分方程原理,采用提取的前 7 个主成分代替影响综合管廊投资估算的 13 个原始特征指标,数据如表 5 所示.

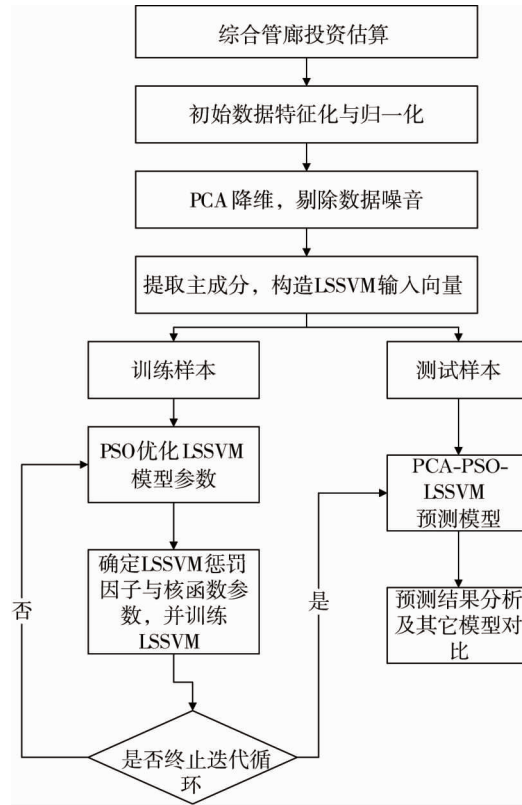


图1 基于PCA-PSO-LSSVM的综合管廊投资估算预测流程

表1 非数值型指标量化方式

非数值型特征指标	量化方式说明
建设地点	1:结合新道路;2:老城区结合道路;3:结合相关管道
土质特征	1:砂性土;2:湿陷性土;3:黄土;4:黏土或软黏土;5:淤泥质土
截面形状	1:矩形;2:圆形;3:异形;4:1+2;5:1+3;6:1+2+3
施工方式	1:现浇;2:预制装配式;3:浅埋暗挖;4:盾构施工法;5:2+3
支护方式	1:无支护;2:土钉墙或复合土钉墙;3:喷锚支护或喷锚+钻孔桩;4:钢板桩+钻孔桩;5:挡土墙
特殊管道是否入廊	1:雨污管道入廊;2:燃气管道入廊;3:皆不入廊;4:1+2

表2 初始特征指标数据量化表

编号	建设地点	土质类型	截面类型	截面面积/m ²	舱数/个	施工方法	支护方式	埋置深度/m	特殊管道	交叉口数/(个/km)	投料口数/(个/km)	通风口数/(个/km)	造价指数	造价/(万/km)
1	1	1	4	32.40	2	5	2	3.00	3	0	8	33	1.26	7 726.64
2	1	1	3	17.36	3	4	2	3.00	4	0	2	6	1.33	9 553.64
3	1	3	4	29.64	3	4	3	2.50	4	0	5	15	1.38	9 553.52
4	1	1	1	34.66	3	4	2	3.00	4	1	1	2	1.40	10 148.94
5	1	1	1	15.00	2	5	5	2.50	4	0	1	3	1.39	7 579.33
6	1	3	1	8.10	1	4	3	2.00	3	0	1	2	1.41	3 132.75
7	1	4	5	37.20	3	4	2	3.00	4	2	4	8	1.41	10 526.85
8	1	1	6	19.27	2	4	5	3.00	4	2	6	16	1.39	11 443.39
9	1	2	5	16.24	2	4	3	4.80	2	0	3	9	1.39	8 991.48
10	1	1	4	9.90	1	5	4	3.00	4	1	5	12	1.41	9 445.61
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
26	1	4	1	38.22	3	1	5	7.10	3	1	12	35	1.38	8 847.06
27	1	4	1	28.29	2	1	3	7.00	4	1	28	53	1.41	8 571.81
28	2	2	1	42.57	3	5	4	7.40	2	1	11	34	1.41	8 650.36
29	1	1	5	22.03	2	5	1	6.76	2	1	15	28	1.39	9 099.84

表 3 原始指标相关系数矩阵

原始特征指标	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	1.00	-0.05	-0.20	0.19	0.09	0.10	0.07	0.21	-0.26	0.15	-0.06	-0.05	0.09
X_2	-0.05	1.00	-0.21	-0.07	-0.11	-0.39	0.31	0.11	0.09	0.12	0.22	0.06	0.27
X_3	-0.20	-0.21	1.00	-0.30	-0.22	0.66	-0.14	-0.55	0.01	0.16	-0.15	-0.15	-0.26
X_4	0.19	-0.07	-0.30	1.00	0.76	-0.38	-0.20	0.67	0.15	0.37	0.53	0.68	0.10
X_5	0.09	-0.11	-0.22	0.76	1.00	-0.31	-0.24	0.41	0.15	0.14	0.19	0.32	0.00
X_6	0.10	-0.39	0.66	-0.38	-0.31	1.00	-0.07	-0.75	-0.23	-0.07	-0.28	-0.26	-0.26
X_7	0.07	0.31	-0.14	-0.20	-0.24	-0.07	1.00	-0.06	-0.19	0.04	0.10	0.04	0.16
X_8	0.21	0.11	-0.55	0.67	0.41	-0.75	-0.06	1.00	-0.04	0.22	0.51	0.53	0.30
X_9	-0.26	0.09	0.01	0.15	0.15	-0.23	-0.19	-0.04	1.00	0.29	0.07	0.07	0.10
X_{10}	0.15	0.12	0.16	0.37	0.14	-0.07	0.04	0.22	0.29	1.00	0.47	0.46	0.35
X_{11}	-0.06	0.22	-0.15	0.53	0.19	-0.28	0.10	0.51	0.07	0.47	1.00	0.96	0.10
X_{12}	-0.05	0.06	-0.15	0.68	0.32	-0.26	0.04	0.53	0.07	0.46	0.96	1.00	0.04
X_{13}	0.09	0.27	-0.26	0.10	0.00	-0.26	0.16	0.30	0.10	0.35	0.10	0.04	1.00

表 4 归一化特征值、方差贡献率及累计方差贡献率

成分	归一化特征值	方差贡献率/%	累计方差贡献率/%
Y_1	4.11	31.62	31.62
Y_2	1.92	14.80	46.41
Y_3	1.67	12.83	59.24
Y_4	1.41	10.85	70.09
Y_5	1.18	9.09	79.18
Y_6	0.71	5.43	84.61
Y_7	0.61	4.69	89.30
Y_8	0.57	4.41	93.71
Y_9	0.40	3.05	96.76
Y_{10}	0.23	1.80	98.56
Y_{11}	0.13	0.98	99.54
Y_{12}	0.05	0.37	99.91
Y_{13}	0.01	0.09	100.00

表 5 样本归一化主成分数值

编号	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7
1	0.00	1.00	0.57	0.39	0.92	0.26	0.69
2	0.00	1.00	0.79	0.78	0.72	0.33	0.59
3	0.00	0.92	0.48	1.00	0.67	0.08	0.59
4	0.20	1.00	0.78	0.91	0.00	0.55	0.25
5	0.00	0.51	0.80	0.31	1.00	0.79	0.23
6	0.00	0.29	0.59	0.80	0.82	1.00	0.90
7	0.50	1.00	0.06	0.98	0.00	0.37	0.87
8	0.26	1.00	0.00	0.64	0.42	0.56	0.35
9	0.00	0.82	0.68	0.48	0.94	1.00	0.73
10	0.00	0.81	0.15	0.72	0.52	1.00	0.53
⋮	⋮	⋮	⋮	⋮	⋮	⋮	
26	1.00	0.00	0.33	0.46	0.79	0.01	0.37
27	1.00	0.12	0.00	0.72	0.59	0.57	0.79
28	1.00	0.79	0.98	0.00	0.48	0.68	0.80
29	0.00	0.98	0.22	0.06	0.36	1.00	0.60

2.3 模型建立

将表 5 中的数据用于 PCA-PSO-LSSVM 模型的建立,以提取的 7 个主成分作为模型的输入变量,综合管廊单公里造价作为输出变量;将初始粒子种群数设为 30,粒子群迭代进化次数设为 100,经反复训练获取粒子群算法的最优适应度曲线,如图 2 所示;经训练样本训练后的 PSO-LSSVM 模型性能如图 3 所示;获得 LSSVM 惩罚因子 C 和核函数参数 σ 的最优值,即 $C=0.3516, \sigma=0.8672$.

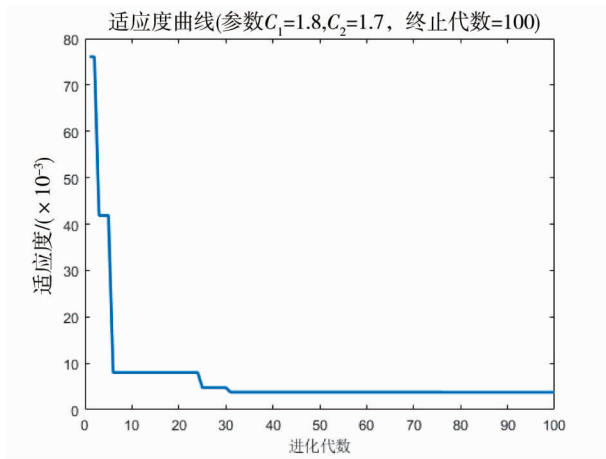


图 2 粒子群算法适应度曲线

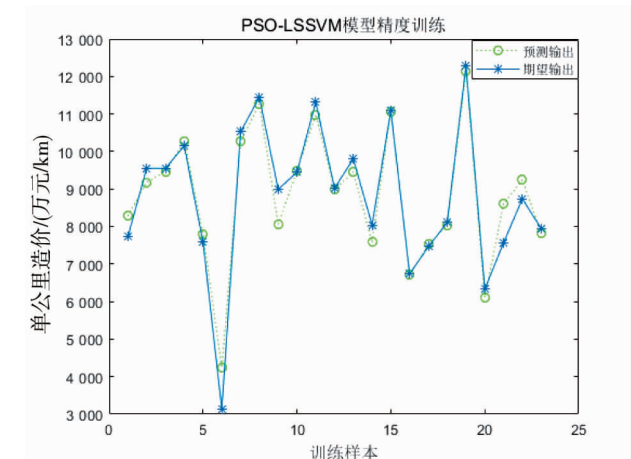


图 3 PSO-LSSVM 模型精度训练

由图 2 可知:随着粒子群进化代数的增加,其适应度值不断趋于稳定,且当进化至 32 代时,粒子群算法的参数寻优与 LSSVM 模型的精度训练结束.此时,结合图 3 可知:经训练样本训练后的 PSO-LSSVM 模型具有较高的线性拟合精度,且在训练样本中预测输出与期望输出的最大相对误差为 5.13%,满足要求,PCA-PSO-LSSVM 模型建立完成.

2.4 模型应用与精度分析

将建立好的 PCA-PSO-LSSVM 模型应用于综合管廊投资估算测试样本的预测,预测结果与 PCA-LSSVM 及 PSO-LSSVM 模型预测结果的对比如图 4 所示,3 种预测模型对测试样本进行预测的相对误差如表 6 所示.

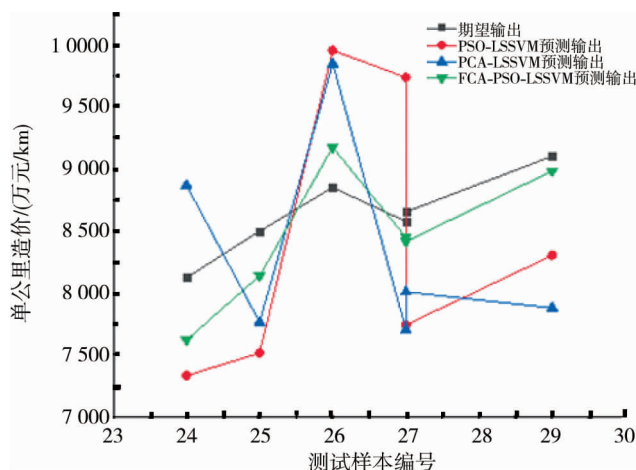


图 4 3 种预测模型预测输出对比

由图 4 与表 6 可知:PCA-PSO-LSSVM 模型的预测输出较 PSO-LSSVM 及 PCA-LSSVM 模型而言,与测试样本期望输出有更高的拟合度,且对测试样本进行预测的平均相对误差与 PSO-LSSVM 及 PCA-LSSVM 相比分别降低了 70.52%与 67.29%.由此可见,本文采用 PCA 对初始数据进行降维处理,剔除数据噪声与冗余,初步提高了 LSSVM 模型预测精度;同时,采用 PSO 对 LSSVM 模型的惩罚因子 C 和核函数参数 σ 进行最优值搜索,再次强化了 LSSVM 模型的预测精度;最终,构建的 PCA-PSO-LSSVM 模型测试样本平均相

对误差为 3.28%,满足综合管廊投资决策阶段投资估算误差容许值(10%),即 PCA-PSO-LSSVM 模型可用于综合管廊投资决策阶段的投资估算.

表 6 3 种预测模型预测输出相对误差

测试样本 编号	测试样本 真实值	PSO-LSSVM		PCA-LSSVM		PCA-PSO-LSSVM	
		预测输出	相对误差/%	预测输出	相对误差/%	预测输出	相对误差/%
24	8 120.82	7 330.67	9.73	8 864.69	9.16	7 618.15	6.19
25	8 494.39	7 512.44	11.56	7 760.48	8.64	8 134.23	4.24
26	8 847.06	9 952.05	12.49	9 844.12	11.27	9 171.74	3.67
27	8 571.81	9 735.01	13.57	7 703.48	10.13	8 444.09	1.49
28	8 650.36	7 738.61	10.54	8 005.91	7.45	8 411.61	2.76
29	9 099.84	8 299.97	8.79	7 875.00	13.46	8 980.63	1.31

为进一步验证本文所提模型预测精度的优越性,同时将其预测输出与 GA-SVM 及 GA-BP 神经网络等模型预测输出进行对比;以上各模型预测输出与测试样本期望输出的相对误差曲线图如图 5 所示.

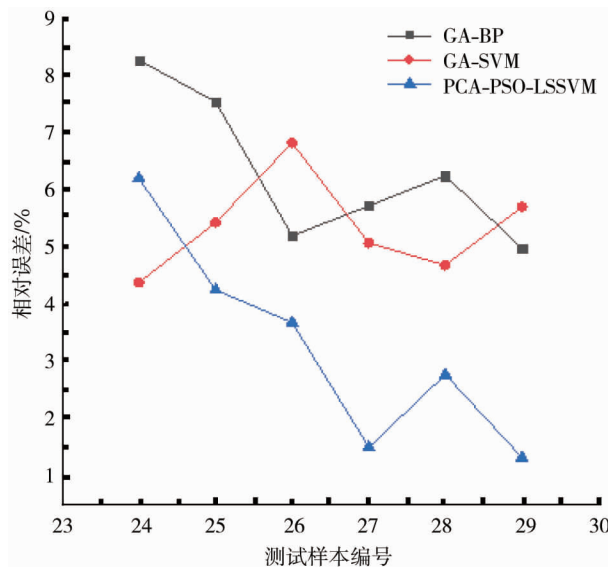


图 5 各模型预测输出相对误差

由图 5 可知:GA-BP 神经网络模型相对误差值位于 4%~8%,GA-SVM 模型相对误差值位于 4%~6%,PCA-PSO-LSSVM 模型相对误差值位于 1%~6%,3 种预测模型相对误差皆满足综合管廊投资决策阶段对投资估算误差的要求,但 PCA-PSO-LSSVM 模型相对误差最小,位于 1%~6%,平均相对误差为 3.28%;与 GA-BP 神经网络模型相比,平均相对误差降低了 48.13%;与 GA-SVM 模型相比,平均相对误差降低了 38.60%.

综上所述,本文所建 PCA-PSO-LSSVM 模型用于综合管廊投资决策阶段投资估算切实可行,且预测精度优于其他模型.

3 结论

1) 采用主成分分析法对初始数据进行降维,将模型维度由 13 维降至 7 维,有效剔除了噪音数据与提高了预测精度.

2) 采用粒子群算法对最小二乘支持向量机的惩罚因子 C 和核函数参数 σ 进行最优值探索,使得最小二乘支持向量机模型性能更优,且预测精度更高.

3) 将所建模型用于工程实例,并将其预测输出结果与 GA-BP 神经网络模型及 GA-SVM 模型预测输出结果相比,本文所建 PCA-PSO-LSSVM 模型平均相对误差为 3.28%,满足综合管廊投资决策阶段投资估算误差容许值(10%);且平均相对误差较 GA-BP 模型与 GA-SVM 模型而言,分别降低了 48.13%,38.60%.

参考文献:

- [1] LUO Y, ALAGHBANDRAD A, GENDER T K, et al. History and recent development of multi-purpose utility tunnels[J]. *Tunnelling and Underground Space Technology*, 2020, 103: 103511.
- [2] CANTO-PERELLO J, CURIEL-ESPARZA J, CALVO V. Strategic decision support system for utility tunnel's planning applying a WOT method[J]. *Tunnelling and Underground Space Technology*, 2016, 55: 146-152.
- [3] HUNT D V L, NASH D, ROGERS C D F. Sustainable utility placement via Multi-Utility Tunnels[J]. *Tunnelling and Underground Space Technology*, 2014, 39: 15-26.
- [4] 蒋承杰. 城市地下综合管廊投融资模式与风险研究[D]. 北京: 对外经济贸易大学, 2017: 24-26.
- [5] 刘杰. 城市综合管廊工程投资估算研究[J]. *科技创新与应用*, 2017(14): 243-245.
- [6] 戴慧奇. 综合管廊估算指标浅析[J]. *水利水电工程造价*, 2020(1): 18-20.
- [7] 胡庆国, 蔡孟龙, 何忠明. 基于 GA-BP 神经网络的综合管廊投资估算研究[J]. *长沙理工大学学报(自然科学版)*, 2020, 17(2): 68-74.
- [8] 李芊, 张悠. 基于遗传支持向量机的综合管廊土建工程造价估算方法研究[J]. *隧道建设(中英文)*, 2018, 38(2): 171-175.
- [9] 李文元, 闫海华, 姚宏杰. 粒子群优化的最小二乘支持向量机在通信装备故障预测中的应用[J]. *微电子学与计算机*, 2013, 30(2): 99-102.
- [10] 於万里, 王艳, 纪志成. 氨糖发酵过程建模与工艺参数优化研究[J]. *系统仿真学报*, 2020, 32(10): 1895-1902.
- [11] TANG P, ZHANG H D, YE T H, et al. A novel method for chemistry tabulation of strained premixed/stratified flames based on principal component analysis[J]. *Applied Mathematics and Mechanics*, 2018, 39(6): 855-866.
- [12] 宋哲, 高建平, 潘龙帅, 等. 基于主成分分析和改进支持向量机的锂离子电池健康状态预测[J]. *汽车技术*, 2020(11): 21-27.
- [13] KENNEDY J, EBERHART R. Particle swarm optimization[C]//*Proceedings of ICNN'95 - International Conference on Neural Networks*. IEEE, 2002: 1942-1948.
- [14] 蔡改贫, 罗茜茜, 刘鑫. 基于 PSO-LSSVM 的 Ni-Cr-W-Mo 合金本构关系研究[J]. *塑性工程学报*, 2020, 27(9): 140-146.
- [15] 宋威, 华子彧. 融入社会影响力的粒子群优化算法[J/OL]. *计算机科学与探索*, 2020: 1-13. (2020-09-25). <https://kns.cnki.net/kcms/detail/11.5602.TP.20200925.1540.012.html>.
- [16] 张悠. 基于支持向量机的综合管廊工程造价估算模型研究[D]. 西安: 西安建筑科技大学, 2018: 41-45.