

doi:10.11835/j.issn.1000-582X.2019.06.010

# 露天矿行车事故预测方法及应用

白润才<sup>1a</sup>, 柴森霖<sup>1b</sup>, 刘光伟<sup>1b</sup>, 付恩三<sup>2</sup>, 赵景昌<sup>1a</sup>

(1. 辽宁工程技术大学 a. 辽宁省高等学校矿产资源开发利用技术及装备研究院;  
b. 矿业学院, 辽宁 阜新 123000; 2. 中华人民共和国应急管理部信息研究院, 北京 100029)

**摘要:**为有效解决露天矿山行车事故预测模型建模时,易受小样本数据、离群数据规模影响,导致模型精度损失、算法抗噪容差能力及收敛速度下降等问题,提出一种基于二次惩罚项修正(PTS)的改进支持向量回归机模型(WLSSVR)。根据训练样本的数据分布特性,研究了服从露天矿山现实应用场景的二次惩罚项,进一步提高回归机模型的抗噪容差能力;考虑非线性预测模型影响因子选择困难的问题,研究了数据降维及因子分析方法,并将主成分分析方法引入到输入数据预处理算法中,以保证算法可得到理想的输入;针对传统回归机模型易受核参数选择影响,易导致模型早熟和收敛速度慢等问题,研究了粒子群惯性因子、学习因子的自适应迭代形式,提出了一种应用改进粒子群算法优化回归机模型核参数的方法。以露天矿行车事故频次预测为例,进行了预测和对比实验。实验结果表明:引入 PTS 模型的测试集预测结果明显优于不采用 PTS 策略的预测结果。这说明,应用文中提出的二次惩罚策略和参数优化算法对复杂系统的事故预测问题研究是可行且有效的。

**关键词:**露天矿;支持向量回归机;二次惩罚修正;改进粒子群算法;行车事故预测

中图分类号:TD713

文献标志码:A

文章编号:1000-582X(2019)06-088-11

## The prediction method of traffic accident and its application in open-pit mine based on the PTS-WLSSVR model

BAI Runcai<sup>1a</sup>, CHAI Senlin<sup>1b</sup>, LIU Guangwei<sup>1b</sup>, FU Ensan<sup>2</sup>, ZHAO Jingchang<sup>1a</sup>

(1a. Research Center of Coal Resources Safe Mining and Clean Utilization; 1b. School of Mining, Liaoning Technical University, Fuxin 123000, Liaoning, P. R. China; 2. Information Research Institute of the Ministry of Emergency Management of the People's Republic of China, Beijing 100029, P. R. China)

**Abstract:** To effectively solve the problems of the accuracy loss of prediction models for traffic accident prediction of open-pit mines, the decrease of the algorithm's capability of anti-noise tolerance and convergence rate caused by small sample data and outliers, we propose a modified support vector regression

收稿日期:2019-01-05

基金项目:国家自然科学基金资助项目(51304104);辽宁省教育厅基金资助项目(LJYL038);辽宁省煤炭资源安全开采与洁净利用工程研究中心开放基金资助项目(TU15KF07)。

Supported by National Natural Science Foundation of China(51304104), Project Fund of Education Department of Liaoning Province (LJYL038) and the Open Projects of Research Center of Coal Resources Safe Mining and Clean Utilization, Liaoning(LNTU16KF07).

作者简介:白润才(1961—),男,博士,辽宁工程技术大学教授、博导,主要研究方向为露天矿优化开采设计、数字化矿山及矿业系统工程。

通讯作者:柴森霖,辽宁工程技术大学博士研究生,主要从事数字化矿山技术及矿业系统工程方面的研究,  
(E-mail)3560103696@qq.com。

model based on penalized trimmed squares (PTS). According to the data distribution characteristics of the training samples, the penalized trimmed squares submitted to the application scenario of open-pit mine is studied to improve the anti-noise tolerance capability of the regression model. In consideration of the difficulties of the nonlinear prediction model impact factor selection, the method of principal component analysis is introduced into the preprocessing algorithm to reduce the data dimension and ensure that the algorithm can get ideal input data. In view of the problems of premature and slow convergence speed caused by the nuclear parameter selection, the inertial factor and the learning factor of particle swarm are studied and an improved particle swarm algorithm to optimize nuclear parameters regression of the model is proposed. The prediction and comparison experiments are carried out in the case of the accident frequency prediction of open-pit mine. The experimental results show that the test set prediction results of the PTS model are better than those without the PTS policy model. This indicates that the modified penalized trimmed squares strategy and parameter optimization algorithm proposed in this paper is feasible and effective for the study of accident prediction of complex systems.

**Keywords:** open-pit mine; SVR; PTS; MPSO; traffic accident prediction

露天矿行车事故预测是实现露天矿山安全、科学、合理有效决策的基础,对于矿山隐患排查、事故预防和降低事故损失均具有重要意义<sup>[1]</sup>。

目前,典型的事故预测算法主要可以归类为如下几种方法:情景分析法、回归预测法、时间预测法、灰色理论方法、马尔可夫理论方法以及非线性系统分析方法<sup>[2]</sup>,这些方法均在一定程度上增强了安全决策的科学性,对于指导工程和隐患预防均起到了一定积极作用。如李明洋等<sup>[3]</sup>采用最小二乘法拟合灰色模型中微分方程,并将一种改进的GM(1,1)模型应用于对非煤矿山死亡率的预测中,并取得了理想的预测效果;顾晓叶等<sup>[4]</sup>提出构建最佳事故预测模型的基本思路和步骤,并结合实际探讨了国内事故死亡人数随经济发展的趋势;孙轶轩等<sup>[5]</sup>将自回归滑动平均模型和支持向量回归机模型相结合,构建时间序列组合预测模型,最终实现对道路交通事故相关指标的趋势预测。赵玲等<sup>[6]</sup>针对各种单一灰色预测模型存在的局限性,提出一种最优加权的灰色组合预测模型来实现交通事故死亡人数的预测;彭小玲<sup>[7]</sup>将支持向量机技术应用于海上船舶交通事故预测分析中;周荣义等<sup>[2]</sup>将主成分分析方法和人工神经网络方法相结合,对煤矿百万吨死亡率指标进行了合理预测;宋传平等<sup>[8]</sup>将GM(1,1)灰色预测模型应用于军用车辆交通事故频数和死亡人数的短期预测中,验证了此方法对于预测问题的适用性;朱志洁等<sup>[9]</sup>将主成分分析方法与BP神经网络相结合,对煤与瓦斯突出问题进行了有效预测;李建刚等<sup>[10]</sup>针对于露天矿运输系统行车事故发生频次数据少的缺点,采用了灰色系统理论其进行了灰色预测;付华等<sup>[11]</sup>提出带有整定因子的扩展卡尔曼滤波器来优化加权最小二乘模型参数,并应用该算法提出一种瓦斯浓度动态预测的新方法。

综合上述文献可知,这些算法各有特定的适应领域,并对各自问题所应用的具体场景均起到了积极的作用,但这些算法也存在一定的局限性,主要体现在:对于现实工程中小样本数据的处理能力、在广泛存在离群点的情况下的算法抗噪容差能力等几方面,并且这些局限性因素也是造成预测精度损失的最根本原因。为此,在前人的研究成果基础上,针对现实工程中采样数据具有小样本特性、且离群点规模不定,易导致预测模型精度损失等问题展开研究,提出一种带有二次惩罚修正的加权最小二乘支持向量回归机预测模型。模型设计的基本思想是利用统计学习方法对于小样本数据的学习能力,来提高算法对小样本数据处理的拟合精度;对于离群点过滤问题,模型在加权最小二乘过滤的基础上,引入二次惩罚修正项,通过计算每一个点的损失代价,来剔除异常点,进一步提高模型精度。最终,经仿真对比实验验证,论证了所述算法的在求解事故预测问题、提高预测精度等方面的有效性。

## 1 WLS-SVR 回归模型

支持向量回归机技术<sup>[12-13]</sup>(support vector regression, SVR)是20世纪末由AT&T贝尔实验室的Vapnic提出,此模型是对SVM模型(support vector machine)在回归问题中的一个推广应用。由于该算法

解的稀疏性、对传统神经网络模型的局部极值和过拟合问题的独特处理能力,被广泛应用于非线性预测、拟合函数参数估计等场景<sup>[13]</sup>。加权最小二乘支持向量机(weighted least square support vector regression, WLSSVR)是在 SVR 及 LS-SVR(least square support vector regression)两模型的基础上发展起来的,该算法保留了 LS-SVR 算法模型中的平方误差损失以及等式约束等特性,在有效降低传统 SVR 模型的计算成本的同时,消除了模型异方差性,具有更好的学习能力、泛化能力,预测准确率更高<sup>[13]</sup>。

对于给定样本集合  $\{(x_k, y_k) | k=1, 2, \dots, N\}$ , 其中  $x_k \in \mathbf{R}^p$  为  $p$  维的参数样本,  $y_k \in \mathbf{R}$  为样本输出集, 基于 WLSSVR 的函数拟合问题的可表述为如下的凸二次优化模型:

$$\begin{cases} \min J(\boldsymbol{\omega}, \boldsymbol{\delta}) = \frac{1}{2} \boldsymbol{\omega}^T \boldsymbol{\omega} + \frac{1}{2} C \sum_{i=1}^N \lambda_i \delta_i^2 \\ \text{s.t. } \delta_i = \boldsymbol{\omega}^T \varphi(\mathbf{x}_i) + b - y_i, i=1, 2, \dots, N \end{cases}, \quad (1)$$

式中:  $\boldsymbol{\omega}$  为权系数向量;  $\varphi(\mathbf{x}_i)$  为输入参数到希尔伯特空间的映射;  $C$  为惩罚因子;  $b$  为模型偏差;  $\lambda_i$  为误差权值。

根据式(1)中优化模型采用拉格朗日算子法进行求解, 建立二次凸优化对偶模型, 对偶模型的拉格朗日方程可表示为:

$$L(\boldsymbol{\omega}, b, \boldsymbol{\delta}, \mathbf{a}) = J(\boldsymbol{\omega}, \boldsymbol{\delta}) - \sum_{i=1}^N a_i [\boldsymbol{\omega}^T \cdot \varphi(\mathbf{x}_i) + b + \delta_i - y_i]. \quad (2)$$

对式(2)采用函数极值求解方法进行求解, 其偏导数方程组如式(3)。

$$\begin{cases} \frac{\partial L(\boldsymbol{\omega}, b, \boldsymbol{\delta}, \mathbf{a})}{\partial \boldsymbol{\omega}} = 0 \Rightarrow \boldsymbol{\omega} - \sum_{i=1}^N a_i \varphi(\mathbf{x}_i), \\ \frac{\partial L(\boldsymbol{\omega}, b, \boldsymbol{\delta}, \mathbf{a})}{\partial b} = 0 \Rightarrow \sum_{i=1}^N a_i = 0, \\ \frac{\partial L(\boldsymbol{\omega}, b, \boldsymbol{\delta}, \mathbf{a})}{\partial \delta_i} = 0 \Rightarrow C \lambda_i \delta_i - a_i = 0, \\ \frac{\partial L(\boldsymbol{\omega}, b, \boldsymbol{\delta}, \mathbf{a})}{\partial a_i} = 0 \Rightarrow \boldsymbol{\omega}^T \cdot \varphi(\mathbf{x}_i) + b + \delta_i - y_i = 0. \end{cases} \quad (3)$$

化简为线性方程组式(3), 消除方程组中的无关变量  $\boldsymbol{\omega}$  和  $\delta_i$ , 其等价问题即为求解线性方程组中的未知变量  $\mathbf{a}$  和  $b$ , 整理方程组后可将其表述为矩阵形式如式(4)。

$$\begin{bmatrix} 0 & \mathbf{I}_\lambda^T \\ \mathbf{I}_\lambda & \boldsymbol{\Omega} + \frac{1}{\gamma} \text{diag}\left(\frac{1}{\lambda_1}, \frac{1}{\lambda_2}, \dots, \frac{1}{\lambda_N}\right) \end{bmatrix} \begin{bmatrix} b^* \\ \mathbf{a}^* \end{bmatrix} = \begin{bmatrix} 0 \\ \mathbf{y} \end{bmatrix}, \quad (4)$$

式中:  $\mathbf{y} = [y_1, y_2, \dots, y_N]^T$ ,  $\mathbf{a}^* = [a_1^*, a_2^*, \dots, a_N^*]$ ,  $\mathbf{I}$  为单位矩阵,  $\boldsymbol{\Omega} = \{K(\mathbf{x}_i, \mathbf{x}_j)\}_{i,j=1}^N$ 。

其中标准的 WLSSVR 模型中给出误差权向量  $\boldsymbol{\lambda}$  的定义形式如式(5)。

$$\lambda_i = \begin{cases} 1 & |\delta_i/\hat{s}| \leq c_1, \\ \frac{c_2 - |\delta_i/\hat{s}|}{c_2 - c_1} & c_1 \leq |\delta_i/\hat{s}| \leq c_2, \\ 10^{-4} & \text{otherwise,} \end{cases} \quad (5)$$

式中:  $\hat{s}$  为误差项的鲁棒估计值,  $\hat{s} = 1.483 \cdot \text{MAD}(\boldsymbol{\delta})$ ,  $c_1, c_2 \in [2.5, 3]$ 。

根据优化条件引入核函数  $K(\mathbf{x}_i, \mathbf{x}_j) = \langle \varphi(\mathbf{x}_i)^T \varphi(\mathbf{x}_j) \rangle$ , 综合考虑各核函数的适应性, 文中选择径向基核函数, 最终通过求解式(4)即可获得的预测函数表达, 如式(6)。

$$y(\mathbf{x}_i) = \sum_{k=1}^N \alpha_j^* K(\mathbf{x}_i, \mathbf{x}_j) + b^*. \quad (6)$$

## 2 PTS-WLSSVR 行车事故预测模型

采用 WLSSVR 技术进行回归学习是有效解决复杂系统输出值预测问题的重要手段之一。但随着工程上的广泛应用,这类统计学习方法也暴露出自身的缺点,如预测模型精度质量受限于正则化参数  $c$  和核参数  $\sigma$  的选择、训练样本中离群点的过滤能力差、预测模型精度受限于离群点规模等几方面问题。为此,在结合行车事故预测问题的工程实际的基础上,将改进的二次惩罚修正策略 (penalized trimmed squares, PTS) 引入到 WLSSVR 算法模型中,实现对传统模型离群点的控制,并将改进的粒子群算法引入 PTS-WLSSVR 模型参数寻优问题中,从而提高整个预测模型的鲁棒特性。

### 2.1 引入 PTS 改进 WLSSVR 算法模型

对 WLSSVR 模型引入惩罚项的作用在于过滤掉回归模型中的离群点,以增强算法的抗噪能力。目前,在应用统计学领域,对于这类离群点的处理也提出了很多种改进方法<sup>[14-17]</sup>,较为有效的如 M 估计、GM 估计以及最小二次修正 (LTS) 估计等。这类修正方法的特点是在处理带有小规模离群点回归问题时,能有效提高算法的稳健性,但精度易受离群点规模限制。然而,对于文中这样一个预测问题,在训练前期是无法判定离群点的具体规模的。因此,此种规模限制对于文中的预测结果将具有显著影响。为从本质上改进这种限制,增进算法的稳健性,作者在文献<sup>[15-16]</sup>提出的 PTS 修正策略的启发下,对 WLSSVR 目标函数进行了修正,其修正的基本思路是对每一个样本进行惩罚,并不断的计算每个样本被作为异常点清理后的惩罚代价,其原理如图 1 所示。

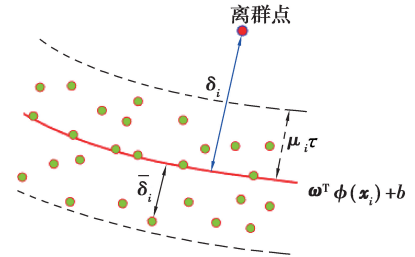


图 1 离群点的过滤原理

Fig. 1 Filtering principle of outliers

基于此种基本思路,首先采用最小二乘建立离群点识别模型,其目标函数如下:

$$\min [\mathbf{y} - (\boldsymbol{\omega}^T \cdot \boldsymbol{\varphi}(\mathbf{x}_i) + b)]^2 = \sum_{i=1}^n \max [e_i^2, \rho_i (c_i \sigma)^2]. \quad (7)$$

通过式(7)识别后,将识别结果采用 0-1 决策变量  $\epsilon_i$  的形式嵌入到回归模型中,故其目标函数模型可被改写为:

$$\begin{aligned} \min J(\boldsymbol{\omega}, \boldsymbol{\delta}) &= \frac{1}{2} \boldsymbol{\omega}^T \boldsymbol{\omega} + \frac{1}{2} C \sum_{i=1}^N (\lambda_i \bar{\delta}_i^2 + \epsilon_i (\mu_i \boldsymbol{\tau})^2), \\ \text{s.t. } \boldsymbol{\omega}^T \boldsymbol{\varphi}(\mathbf{x}_i) + b - y_i &\leq \bar{\delta}_i + \epsilon_i Q - \boldsymbol{\omega}^T \boldsymbol{\varphi}(\mathbf{x}_i) - b + y_i \leq \bar{\delta}_i + \epsilon_i Q, \\ \epsilon_i &\in \{0, 1\} \\ \bar{\delta}_i &\geq 0 \quad i = 1, 2, \dots, N, \end{aligned} \quad (8)$$

式中:  $Q$  是一个常数,表示残差  $\delta_i$  的上限;  $\epsilon_i$  为是否删除离散点的决策变量;  $\mu_i \boldsymbol{\tau}$  为第  $i$  个样本的惩罚项;  $\boldsymbol{\tau}$  是一个镇定系数;  $\mu_i$  为稳健性和效率调节参数,根据文中常见此参数的值被定义为式(10)形式。

$$\mu_i = \mu \sqrt{1 - h_i^*}. \quad (10)$$

式中:  $h_i^*$  为 Hessian 阵的第  $i$  个元素, Hessian 矩阵元素值的计算见式(11);  $\mu$  为一个常数,根据文中模型训练场景中发现建议取值范围为  $\mu \in [2.7, 3.0]$ 。

$$h_i^* = \begin{cases} \mathbf{x}_i^T (\mathbf{X}_{k+1}^T \mathbf{X}_{k+1})^{-1} \mathbf{x}_i, & i = k + 1, \dots, N, \\ \mathbf{x}_i^T (\mathbf{X}_k^T \mathbf{X}_k)^{-1} \mathbf{x}_i, & i = 1, 2, \dots, K. \end{cases} \quad (11)$$

式中:  $\mathbf{X}_{k+1}$  表示  $\mathbf{x}_i$  的最小协方差行列式矩阵。

模型中主要控制的一个参数即为  $\bar{\delta}_i$ , 该参数的取值主要依靠惩罚项进行确定,具体函数形式如式(12)所示。

$$\bar{\delta}_i = \begin{cases} \delta_i, & |\delta_i| \leq \mu_i \boldsymbol{\tau}, \\ \mu_i \boldsymbol{\tau}, & \text{otherwise.} \end{cases} \quad (12)$$

## 2.2 WLS-SVR 核参数优化

为提高非线性预测模型的拟合精度及收敛速度,笔者采用改进的粒子群算法(MPSO)对正则化参数和核参数进行优化。具体做法是将参数  $c, \sigma$  为平面粒子位置坐标,建立优化模型如式(13),将式(14)作为优化粒子群算法的适应值函数。

$$\begin{cases} \min_{c, \sigma} f(c, \sigma) = \frac{\sum_{k=1}^N \lambda_k \epsilon_k}{\sum_{k=1}^N \lambda_k} \ll \frac{\sum_{k=1}^N \epsilon_k}{n}, \\ \epsilon_k = y_k - y_k^*, \end{cases} \quad (13)$$

$$F(c, \sigma) = M / \frac{\sum_{k=1}^N \lambda_k \epsilon_k}{\sum_{k=1}^N \lambda_k} = \frac{\sum_{k=1}^N \lambda_k}{\sum_{k=1}^N \lambda_k \epsilon_k}. \quad (14)$$

式中  $M$  表示全局目标的镇定系数,  $M = \max(\sum_{k=1}^N \lambda_k \epsilon_k / \sum_{k=1}^N \lambda_k)$ 。

考虑模型易受训练数据规模局限性以及其他多种综合因素限制,且上述影响对粒子群算法极为敏感,严重制约粒子群算法种群的搜索范围,导致随着迭代进行会促进种群丧失多样性,极易造成迭代初期算法便出现早熟现象。为使算法在迭代初期具有较强的空间搜索能力,在迭代后期具有较强的寻优能力,对粒子群状态转移模型中的惯性因子( $\omega$ )、学习因子( $c_1, c_2$ )进行了如式(15)的修正。此种修正的作用在于能有效的保证算法在迭代初期以较大的范围搜索最优解,保持种群多样性;在迭代后期能实现更精细的局部搜索,以提高整体优化建模精度。

$$\begin{cases} \omega = \left[ \left( 1 - \left( \frac{k}{G} \right)^{\frac{1}{3}} \right) \omega_s + \left( \frac{k}{G} \right)^{\frac{1}{3}} \omega_e \right] / 2, \\ c_1 = c_{1s} - (c_{1s} - c_{1e}) \left( \frac{k}{G} \right)^{\frac{1}{3}}, \\ c_2 = c_{2s} + (c_{2e} - c_{2s}) \left( \frac{k}{G} \right)^{\frac{1}{3}}, \end{cases} \quad (15)$$

式中:  $\omega_s$  和  $\omega_e$  分别表示惯性因子  $\omega$  的初始值和上一期迭代结果值;  $c_{1s}$  和  $c_{1e}$  分别表示学习因子  $c_1$  的初始值和上一期迭代结果值;  $k$  表示迭代次数;  $G$  表示总的迭代次数。

为进一步扩大粒子搜索空间,减少个体寻优过程中易陷入局部最优状态的可能,在每次迭代产生的搜索空间内,算法会采用轮盘赌法随机调整部分粒子位置,来加大搜索空间内寻优策略的随机性,其轮盘赌事件概率如式(16)所示。

$$P(p_s) = \frac{\sum_i^p F_i(c, \sigma)}{\sum_k^N F_k(c, \sigma)}, \quad s \in [0, \text{最大粒子数}]. \quad (16)$$

位置更新模型如式(17)所示。

$$p_i = p_i + \left[ (p_{\max} - p_i) \frac{p_i}{(p_{\max} - p_{\min})} + (p_i - p_{\min}) \frac{(p_{\max} - p_{\min}) - p_i}{(p_{\max} - p_{\min})} \right] \frac{k\gamma}{G}, \quad (17)$$

式中:  $\gamma$  为  $[-1, 1]$  区间上的随机数,  $k$  为粒子种群更新的次数;  $p_{\max}, p_{\min}$  为粒子位置的区间边界;其他参数同上。

综合上述修正策略,基于 PSO 算法的核参数优化流程如算法 1 所示。

算法 1: 基于 MPSO 的核参数优化算法

```

Begin
SVR模型参数初始化;
while(终止条件)
  构建或修正PTS-WLSSVR模型;
  拟合残差 $\delta$ ;
  If( $\delta > \varepsilon$ )
    按照式(12)重新构建参数优化的目标函数;
    初始化种群粒子 $x_i(c, \sigma), i=1, 2, \dots, n$ ;
    设置粒子算法参数;
    while(终止条件)
      计算适应度;
      更新极值;
      按照式(15)更新 $w$ 和 $c$ ;
      采用 $w$ 、 $c$ 和传统速度公式更新粒子速度;
      计算粒子位置更新概率 $P(p_i)$ ;
      计算位置更新粒子规模;
      对选出的粒子,按照式(17)更新位置;
    End;
    修正正则化参数 $c$ 和核参数 $\sigma$ ;
    break;
  End
else
  理想输出;
  break;
End
End
End

```

图 2 基于 MPSO 的核参数优化算法

Fig. 2 Kernel parameter optimization algorithm based on MPSO

### 3 应用实例

车辆事故频次是评价露天矿行车安全性的重要指标,也是露天矿行车事故预测的一个重要指标。为论证文中模型的可行性及求解预测问题的有效性,采用笔者提出的算法模型对某露天矿山车辆事故频次进行预测。依据该露天矿历年矿山行车事故的因素分析,构建该矿行车安全预测影响指标体系结构<sup>[18-21]</sup>,如图 3 所示。并依据行车安全影响指标体系中所列安全因素,收集该矿安全指标数据,数据主要采集自该矿 2000—2017 年运营的数据。考虑统计指标为该矿全年综合安全指标的评价,为此,统计数据  $X_{ij}$  为指标分项打分的平均值,此项打分来自于 10 位现场专家对该矿全年指标体系分项状态的综合评价,其中打分标准为 10 分制,分数越高因素体系分项的综合质量越好;另外,事故频次( $Y$ )采用历年的事故频次的统计数据。对于模型训练,将这 18 年的统计数据分为 2 组,其中 2000—2009 年采样数据作为模型的训练样本,2010—2017 年采样数据作为模型的测试样本。

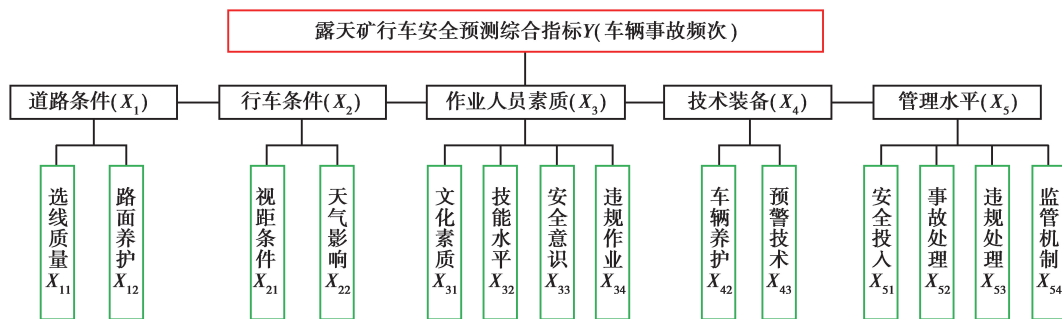


图 3 露天矿行车安全影响指标体系

Fig. 3 Impact indicator system of traffic safety in open-pit mine

#### 3.1 基于 PCA 的数据预处理

露天矿车辆事故频次是衡量露天矿行车安全的一项综合性指标,因露天矿运输系统庞杂、影响因素复杂、关联性指标众多,并且各指标间存在相互的耦合关系。鉴于此特点,文中采用主成分分析法<sup>[22-23]</sup>对原始影响指标进行预处理,分析出其主成分以降低原始指标数据维度,从而消除多种关联因素间的相互影响。

应用 Matlab 科学计算软件系统对行车安全影响因素数据进行主成分分析,并按照如下步骤进行主成分

提取:

1) 相关系数矩阵  $R$  计算: 由于指标数据均源于打分系统, 因为其评分标准一致。因此, 在 PCA 处理过程中无需进行标准化处理, 为此, 文中直接采用公式  $r_{jk} = 1/n \sum_i X_{ij} X_{ik}$  计算相关系数, 计算结果如式(18)所示。

$$R = \begin{bmatrix} 1.000 & 0.622 & 0.164 & 0.357 & -0.109 & 0.107 & 0.107 & 0.376 & 0.424 & 0.422 & 0.429 & 0.325 & 0.232 & 0.540 \\ 0.622 & 1.000 & 0.427 & 0.180 & 0.420 & 0.391 & 0.27 & 0.323 & 0.445 & 0.262 & 0.161 & 0.617 & -0.161 & 0.399 \\ 0.164 & 0.427 & 1.000 & 0.036 & 0.410 & 0.610 & 0.469 & 0.373 & 0.369 & 0.430 & 0.390 & 0.344 & -0.165 & -0.270 \\ 0.357 & 0.18 & 0.036 & 1.000 & 0.179 & 0.383 & 0.524 & 0.618 & 0.684 & -0.042 & 0.089 & 0.085 & 0.756 & 0.018 \\ -0.109 & 0.42 & 0.410 & 0.179 & 1.000 & 0.684 & 0.447 & 0.069 & 0.289 & 0.105 & -0.233 & 0.437 & -0.069 & -0.205 \\ 0.107 & 0.391 & 0.610 & 0.383 & 0.684 & 1.000 & 0.619 & 0.328 & 0.431 & 0.116 & -0.114 & 0.357 & 0.249 & -0.364 \\ 0.017 & 0.27 & 0.469 & 0.524 & 0.447 & 0.619 & 1.000 & 0.517 & 0.412 & 0 & 0.127 & 0.223 & 0.233 & -0.102 \\ 0.376 & 0.323 & 0.373 & 0.618 & 0.069 & 0.328 & 0.517 & 1.000 & 0.650 & -0.185 & 0.597 & -0.240 & 0.202 & -0.048 \\ 0.424 & 0.445 & 0.369 & 0.684 & 0.289 & 0.431 & 0.412 & 0.650 & 1.000 & -0.008 & 0.273 & 0.108 & 0.255 & 0.100 \\ 0.422 & 0.262 & 0.430 & -0.042 & 0.105 & 0.116 & 0.000 & -0.185 & -0.008 & 1.000 & 0.249 & 0.595 & 0.165 & 0.151 \\ 0.429 & 0.161 & 0.390 & 0.089 & -0.233 & -0.114 & 0.127 & 0.597 & 0.273 & 0.249 & 1.000 & -0.199 & 0.027 & 0.078 \\ 0.325 & 0.617 & 0.344 & 0.085 & 0.437 & 0.357 & 0.223 & -0.240 & 0.108 & 0.595 & -0.199 & 1.000 & 0.095 & 0.362 \\ 0.232 & -0.161 & -0.165 & 0.756 & -0.069 & 0.249 & 0.233 & 0.202 & 0.225 & 0.165 & 0.027 & 0.095 & 1.000 & -0.098 \\ 0.540 & 0.399 & -0.270 & 0.018 & -0.205 & -0.364 & -0.102 & -0.048 & 0.100 & 0.151 & 0.078 & 0.362 & -0.098 & 1.000 \end{bmatrix} \quad (18)$$

2) 计算矩阵  $R$  的特征值、特征向量及贡献率: 采用 Matlab 中内置的函数计算矩阵  $R$  的特征值和特征向量。按照特征值的大小对各成分进行排序, 并按照特征值所占特征值总体的比例来计算贡献率, 计算结果如表 1 所示。

表 1 载荷及贡献率

Table 1 Burden and contribution rates

成分	特征值	贡献率/%	累计贡献率/%
1	4.477	31.98	31.98
2	2.429	17.35	49.33
3	2.330	16.64	65.97
4	1.646	11.75	77.73
5	1.241	8.87	86.59
6	0.590	4.22	90.81
7	0.390	2.78	93.59
8	0.325	2.32	95.91
9	0.208	1.49	97.40
10	0.154	1.10	98.50
11	0.106	0.76	99.26
12	0.063	0.45	99.71
13	0.034	0.24	99.96
14	0.007	0.05	100.00

3) 主成分因子的得分矩阵确定: 由表 1 可知, 第 5 个主成分的累计贡献率为 86.59%。考虑保证系统数据的不确定性和真实性, 采用累计贡献率大于 85% 的选择原则, 选择前 5 成分为主成分。并根据这 5 组主成分的特征向量, 整理出主成分因子的得分矩阵形式如式(19)所示。

$$\mathbf{P} = \begin{bmatrix} 0.047 6 & 0.017 9 & 0.276 7 & -0.370 1 & -0.017 9 \\ -0.072 7 & -0.340 & 0.150 1 & -0.289 5 & -0.105 9 \\ -0.427 6 & 0.233 7 & -0.004 1 & 0.071 2 & 0.390 2 \\ 0.375 9 & 0.007 4 & -0.002 3 & 0.133 0 & -0.035 5 \\ 0.046 6 & -0.193 6 & 0.113 9 & 0.273 1 & -0.701 3 \\ 0.045 8 & 0.039 8 & 0.189 0 & -0.392 8 & 0.142 8 \\ 0.002 8 & -0.050 8 & -0.745 1 & -0.086 5 & 0.084 8 \\ -0.251 3 & -0.129 9 & -0.046 4 & -0.151 1 & -0.104 3 \\ 0.009 5 & -0.207 6 & 0.324 2 & 0.623 3 & 0.337 1 \\ 0.012 9 & 0.589 4 & -0.010 9 & 0.245 7 & -0.081 6 \\ -0.478 9 & 0.269 0 & -0.149 8 & 0.088 3 & -0.395 2 \\ 0.264 4 & 0.051 5 & -0.167 3 & 0.041 1 & 0.121 8 \\ 0.519 1 & 0.422 6 & 0.018 1 & -0.118 1 & -0.109 8 \\ 0.178 5 & -0.356 0 & -0.372 7 & 0.158 2 & 0.036 0 \end{bmatrix} \quad (19)$$

4) 将所有的原始采样数据按照公式(20)进行 PCA 数据转换,经转换后的主成分载荷数据,即数据预处理结果如表 2 所示。

$$\hat{\mathbf{X}}_i = \mathbf{P}^T \mathbf{X}_{ij}, i = 1, 2, \dots, 5. \quad (20)$$

表 2 PCA 计算后的数据样本  
Table 2 Data sample calculated by PCA

年限	$\hat{\mathbf{X}}_1$	$\hat{\mathbf{X}}_2$	$\hat{\mathbf{X}}_3$	$\hat{\mathbf{X}}_4$	$\hat{\mathbf{X}}_5$	年限	$\hat{\mathbf{X}}_1$	$\hat{\mathbf{X}}_2$	$\hat{\mathbf{X}}_3$	$\hat{\mathbf{X}}_4$	$\hat{\mathbf{X}}_5$
2000	1.247	5.771	-3.527	1.984	-4.619	2009	4.580	3.261	-4.096	1.710	-3.913
2001	2.786	4.294	-4.775	2.39	-4.993	2010	2.063	2.823	-3.886	2.643	-4.838
2002	3.808	2.195	-3.701	1.443	-4.153	2011	3.351	2.775	-4.524	2.698	-4.101
2003	2.555	2.072	-4.274	2.014	-4.223	2012	-0.899	1.745	-4.913	2.832	-4.071
2004	3.544	3.053	-3.955	2.456	-3.763	2013	-1.161	1.123	-3.705	0.836	-4.540
2005	1.479	3.584	-5.246	1.853	-3.677	2014	1.295	4.281	-4.886	1.115	-4.888
2006	2.578	3.861	-4.469	1.821	-4.224	2015	2.611	3.239	-4.865	1.162	-6.095
2007	0.122	5.190	-2.815	2.013	-4.497	2016	2.819	2.996	-4.030	1.878	-3.745
2008	3.454	2.142	-4.021	1.585	-4.333	2017	2.481	3.450	-4.611	1.558	-4.385

### 3.2 模型训练

为论证算法的有效性,并进一步说明应用 PTS 方法修正预测模型的优越性,笔者将上述分解出 5 种主成分作为输入参数,并采用训练数据分别训练 PTS-WLSSVR+MPSO、WLSSVR+MPSO、WLSSVR 以及 WLSSVR+PSO 这 4 组模型。为消除平台及参数差异,4 组算法模型均在一台 I5 2.9GHz 8G 内存的 win7 主机上编译和运行,4 组模型中组合算法的初始化参数如下: $c_0=3.5, \sigma_0=0.5$ ;粒子群算法种群规模  $N=50$ ,最大迭代次数  $G=150, c_1=2.5, c_2=3, \omega=1$ 。4 组模型逐代的平均误差对数和训练的收敛效果如图 4 所示,其中为满足各组模型间在同一坐标系下对比分析,图 4 中误差曲线为逐代平均误差的对数曲线。



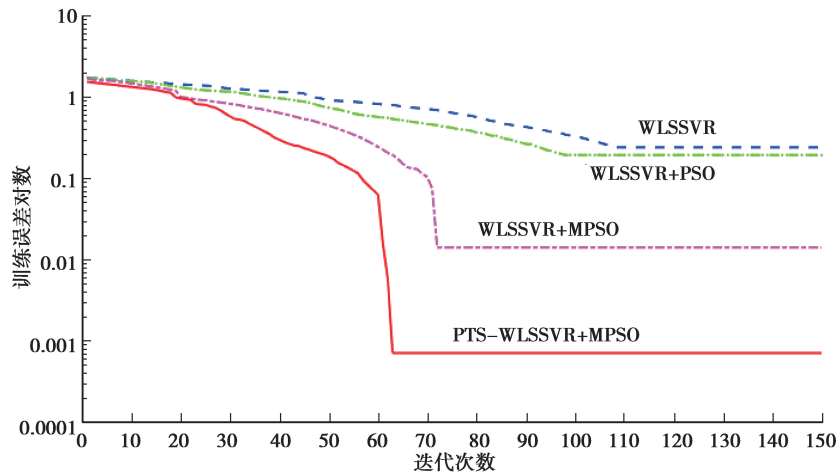


图 4 训练误差对比图

Fig. 4 Error contrast diagram

由图 4 中训练结果可以看出, WLSSVR 模型训练效果最差, 不但精度受限而且收敛速度也相对较慢; 应用 PSO 对支持向量回归模型的参数进行优化后, 其收敛速度有明显的提升, 且文中给出的自适应参数调节机制, 更能体现出收敛速度的提升效果; 对比 WLSSVR+MPSO 算法可发现, 引入改进的 PTS 修正策略后, 能明显降低模型迭代过程中的训练误差, 并能有效改善算法的迭代速度。综合上述分析, 文中改进的算法模型较之传统算法, 在此优化场景下适应性更强, 能极好的满足当前的预测需求。

### 3.3 试验结果分析

应用测试集数据对文中提出的 PTS-WLSSVR 模型进行仿真, 测试集所得出的 8 组仿真结果, 如图 5 所示。

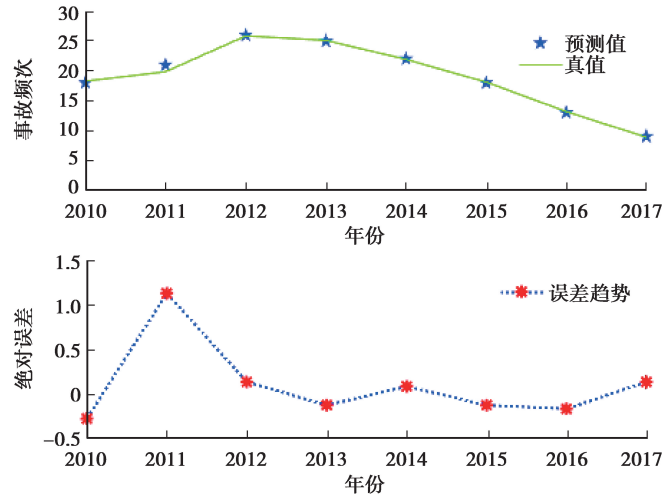


图 5 PTS-WLSSVR 模型预测结果

Fig. 5 PTS-WLSSVR model prediction results.

为进一步说明 PTS 引入后的平差效果, 笔者应用测试数据对未引入文中 PTS 策略的模型进行仿真对比, 两组模型绝对误差对比结果如图 6 所示。

通过图 6 中的对比曲线可以看出, 经过一层的 PTS 过滤能有效控制误差的全局传播, 对于控制模型的误差水平具有极为明显的效果。

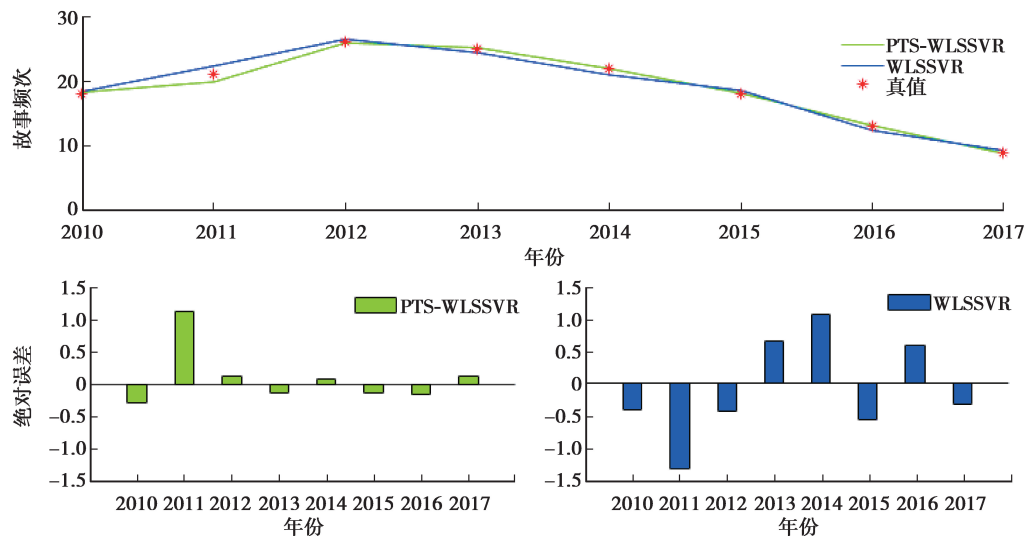


图6 模型优化效果对比

Fig. 6 Comparison of model optimization results

## 4 结 论

1) 在传统 WLSSVR 模型中引入二次惩罚修正增强了算法模型的鲁棒性,提高了算法抗噪能力,经过改进粒子群算法优化核参数后,有效提高了文中预测算法的效率和预测准确率。经多组算法模型的实例验证,论证了文中改进算法的有效性;

2) 采用 PCA 分析方法对原始样本数据进行了预处理,消除了多因素间的耦合联系,提高了主观性影响因素分析的精度。通过实例验证,论证了 PCA 方法在此类预测问题中应用的适用性和有效性;

3) 文中所建立的影响指标体系是根据特定实例建立,后续工作有必要进一步完善此体系结构以保证可适用于更多的行车事故指标参数的预测场景。

### 参考文献:

- [1] 骆中洲, 吴亚平, 温思龙. 露天矿事故分析与预测及安全信息管理[J]. 化工矿山技术, 1992(5): 4-8.  
LUO Zhongzhou, WU Yaping, WEN Silong. Accident analysis and prediction and information system of safety management in surface mine[J]. Chemical Mining Technology, 1992(5): 4-8. (in Chinese)
- [2] 周荣义, 钟岸, 任竞舟, 等. 基于主成分分析和神经网络的事事故预测方法及应用[J]. 中国安全科学学报, 2013, 23(7): 55-60.  
ZHOU Rongyi, ZHONG An, REN Jingzhou, et al. An accident forecasting method of ANN based on PCA and its application[J]. China Safety Science Journal, 2013, 23(7): 55-60. (in Chinese)
- [3] 李明洋, 姜福川. 基于最小二乘法的灰色 GM(1,1)改进模型在非煤矿山事故预测中的应用[J]. 中国安全生产科学技术, 2013, 9(11): 83-90.  
LI Mingyang, JIANG Fuchuan. Improved gray GM(1, 1) model in the non-coal mine accident prediction based on the method of least square[J]. Journal of Safety Science and Technology, 2013, 9(11): 83-90. (in Chinese)
- [4] 顾晓叶, 于殿宝, 朱玉生, 等. 事故预测数学模型的研究与实践[J]. 中国安全科学学报, 2009, 19(12): 23-27, 204.  
GU Xiaoye, YU Dianbao, ZHU Yusheng, et al. Research and practice of accident prediction mathematical models[J]. China Safety Science Journal, 2009, 19(12): 23-27, 204. (in Chinese)
- [5] 孙轶轩, 邵春福, 计寻, 等. 基于 ARIMA 与信息粒化 SVR 组合模型的道路交通事故时序预测[J]. 清华大学学报(自然科学版), 2014, 54(3): 348-353, 359.  
SUN Yixuan, SHAO Chunfu, JI Xun, et al. Urban traffic accident time series prediction model based on combination of ARIMA and information granulation SVR[J]. Journal of Tsinghua University (Science and Technology), 2014, 54(3): 348-353, 359. (in Chinese)
- [6] 赵玲, 许宏科, 程鸿亮. 基于最优加权组合模型的道路交通事故预测[J]. 计算机工程与应用, 2013, 49(24): 11-15.

- ZHAO Ling, XU Hongke, CHENG Hongliang. Road traffic accidents prediction based on optimal weighted combined model[J]. Computer Engineering and Applications, 2013, 49(24): 11-15. (in Chinese)
- [7] 彭小玲. 支持向量机技术在海上船舶交通事故预测中的应用[J]. 舰船科学技术, 2018, 40(2): 55-57.  
PENG Xiaoling. Research on ship traffic accident prediction using state vector machine[J]. Ship Science and Technology, 2018, 40(2): 55-57. (in Chinese)
- [8] 宋传平, 王鹏飞, 李涛. 基于 GM(1, 1)模型的军用车辆交通事故预测研究[J]. 中国安全科学学报, 2010, 20(11): 52-55.  
SONG Chuanping, WANG Pengfei, LI Tao. Forecast of military vehicle traffic accidents based on GM(1, 1) model[J]. China Safety Science Journal, 2010, 20(11): 52-55. (in Chinese)
- [9] 朱志洁, 张宏伟, 韩军, 等. 基于 PCA-BP 神经网络的煤与瓦斯突出预测研究[J]. 中国安全科学学报, 2013, 23(4): 45-50.  
ZHU Zhijie, ZHANG Hongwei, HAN Jun, et al. Prediction of coal and gas outburst based on PCA-BP neural network[J]. China Safety Science Journal, 2013, 23(4): 45-50. (in Chinese)
- [10] 李建刚, 白润才, 刘光伟, 等. 露天矿行车事故频次的灰色预测[J]. 辽宁工程技术大学学报(自然科学版), 2005, 24(z1): 6-8.  
LI Jiangan, BAI Runcai, LIU Guangwei, et al. Grey forecast of travel accident frequency in open-pit mine[J]. Journal of Liaoning Technical University(Natural Science Edition), 2005, 24(z1): 6-8. (in Chinese)
- [11] 付华, 瞿海, 孟祥云, 等. 一种 EKF-WLS-SVR 与混沌时间序列分析的瓦斯动态预测新方法[J]. 传感技术学报, 2015, 28(1): 126-131.  
FU Hua, ZI Hai, MENG Xiangyun, et al. A new method of mine gas dynamic prediction based on EKF-WLS-SVR and chaotic time series analysis[J]. Chinese Journal of Sensors and Actuators, 2015, 28(1): 126-131. (in Chinese)
- [12] Schuldt C, Laptev I, Caputo B. Recognizing human actions: a local SVM approach[C] // Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004, 26-26 Aug 2004. Cambridge, UK, 2004: 32-36.
- [13] Cherkassky V, Ma Y Q. Practical selection of SVM parameters and noise estimation for SVM regression[J]. Neural Networks, 2004, 17(1): 113-126.
- [14] Zioutas G, Avramidis A. Deleting outliers in robust regression with mixed integer programming[J]. Acta Mathematicae Applicatae Sinica, English Series, 2005, 21(2): 323-334.
- [15] Zioutas G, Avramidis A, Pitsoulis L. Penalized trimmed squares and a modification of support vectors for unmasking outliers in linear regression[J]. REVSTAT, 2007, 5(1): 115-136.
- [16] Zioutas G, Pitsoulis L, Avramidis A. Quadratic mixed integer programming and support vectors for deleting outliers in robust regression[J]. Annals of Operations Research, 2009, 166(1): 339-353.
- [17] Avramidis A, Zioutas G. Locally and globally robust penalized trimmed squares regression[J]. Simulation Modelling Practice and Theory, 2011, 19(1): 148-160.
- [18] 王少群, 张晓明. 高速公路交通事故概率预测方法研究[J]. 价值工程, 2014, 33(13): 67-68.  
WANG Shaoqun, ZHANG Xiaoming. Probability forecast method for freeway traffic accidents[J]. Value Engineering, 2014, 33(13): 67-68. (in Chinese)
- [19] Tsuang M T, Boor M, Fleming J A. Psychiatric aspects of traffic accidents[J]. American Journal of Psychiatry, 1985, 142(5): 538-546.
- [20] 张军, 杨维平, 杨立功, 等. 模糊综合评判在露天矿安全评价中的应用[J]. 安全与环境学报, 2011, 11(6): 259-263.  
ZHANG Jun, YANG Weiping, YANG Ligong, et al. Comprehensive fuzzy evaluation method and its application to the open-air mine safety forecasting[J]. Journal of Safety and Environment, 2011, 11(6): 259-263. (in Chinese)
- [21] 贾水库. 大型露天矿运输安全控制与预警系统[J]. 北京科技大学学报, 2009, 31(5): 542-546.  
JIA Shuike. Pre-warning and control system of transport safety in large-scale surface mine[J]. Journal of University of Science and Technology Beijing, 2009, 31(5): 542-546. (in Chinese)
- [22] Brown S D. Book reviews: introduction to multivariate statistical analysis in chemometrics[J]. Applied Spectroscopy, 2010, 64(4): 112A.
- [23] Egeren L F. Multivariate statistical analysis[J]. Psychophysiology, 1973, 10(5): 517-532.