文章编号:1671-4598(2023)08-0238-07 DOI:10.16526/j.cnki.11—4762/tp.2023.08.036 中图分类号:TP389.1 文献标识码:A

基于松鼠觅食算法优化 LSSVM 的泥石流预测

李 璐[1,2],张永强[1],李丽敏[3],马 媛[1],窦婉婷[1],王 悦[4]
(1. 西安思源学院 理工学院,西安 710038;
2. 铜川职业技术学院 机电工程学院,陕西 铜川 727031;
3. 西安工程大学 电子信息学院,西安 710048;
4. 西安交通大学城市学院 传播系,西安 710018)

摘要:针对山区村镇泥石流影响因素多元复杂、LSSVM 算法参数随机导致的精度不佳及陷入局部最优问题,采用核主成分分析 KPCA 降维、SSA 算法参数寻优的方法建立 LSSVM 泥石流灾害预测模型;以山阳县中村镇泥石流为例,分析泥石流全域地形地貌成灾因子,对数据预处理清洗规范,利用 KPCA 主成分贡献率选取出 6 个成灾因子作为 LSSVM 算法的输入数据,泥石流发生概率为输出,建立泥石流预报模型,并用 SSA 算法进行模型参数的优化;将 SSA 寻优后的 LSSVM 预测结果与 GA、GC 参数寻优模型预测结果比对,结果表明 SSA-LSSVM 准确率达到 93.2%,相比其他模型提高[4.8%—1.4%],且 SSA 算法优化的 LSSVM 模型的 MAE、MSE 和 RMSE 最小且接近于零,同时从泥石流发生的预报等级维度进行结果比对分析,结果进一步说明模型预测的精度及稳健性;该研究说明 SSA-LSSVM 算法可用于泥石流灾害发生概率的预测,为此类灾害预测提供了科学依据。

关键词:LSSVM;松鼠觅食 SSA;KPCA;泥石流预测模型

Prediction of Debris Flow Based on Squirrel Foraging Algorithm Optimized LSSVM

LI Lu[1,2], ZHANG Yongqiang[1], LI Limin[3], MA Yuan[1], DOU Wanting[1], WANG Yue[4]
(1. School of Science and Engineering, Xi'an Siyuan University, Xi'an 710038, China;
2. School of Mechanical and Electrical Engineering, Tongchuan Vocational and Technical College, Tongchuan 727031, China;
3. School of Electronics and Information, Xi'an Polytechnic University, Xi'an 710048, China;
4. Department of Communication, City College of Xi'an Jiaotong University, Xi'an 710018, China)

Abstract: In order to solve the problem of poor accuracy and local optimal caused by multiple and complex influencing factors of debris flow in mountainous villages and towns and the random parameters of LSSVM algorithm, the LSSVM debris flow disaster prediction model was established by KPCA dimension reduction and SSA algorithm parameter optimization methods. Mudslides son duong district of villages and towns, for example, global topography by factor analysis of debris flow, wash specification for data preprocessing, 6 by using KPCA principal component contribution rate to select the factors as the input data of LSSVM algorithm, debris flow occurrence probability as output, debris flow forecast model is established, and model parameters are optimized with the SSA algorithm. By comparing the prediction results of LSSVM optimized by SSA with those of GA and GC parameter optimization models, the results show that the accuracy of SSA-LSSVM reaches 93.2%, which is higher than that of other models [4.8%—1.4%]. Moreover, MAE, MSE and RMSE of LSSVM optimized by SSA algorithm are minimum and close to zero. At the same time, the results are compared and analyzed from the prediction grade dimension of debris flow occurrence, and the results further illustrate the accuracy and robustness of the model prediction. This study shows that SSA-LSSVM algorithm can be used to predict the probability of debris flow disasters, and provides a scientific basis for the prediction of such disasters.

Keywords: LSSVM; Squirrels foraging SSA; KPCA. Debris flow prediction model

收稿日期:2023-03-06; **修回日期**:2023-03-21。
基金项目:陕西省教育厅科研计划资助项目(22JK0515);陕西省自然科学基础研究计划项目(2023-JC-YB-464)。
作者简介:李 璐(1994-),女,山西长治人,硕士,助教,主要从事地质灾害预报、人工智能算法方面的研究。
通讯作者:张永强(1984-),男,陕西西安人,硕士,高级工程师,主要从事人工智能算法方向的研究。
引用格式:李 璐,张永强,李丽敏,等.基于松鼠觅食算法优化 LSSVM 的泥石流预测[J].计算机测量与控制,2023,31(8):238-244.

0　引言

泥石流灾害的发生是在自然演变或人为因素的影响下，一种复杂的非线性动力学演化过程。我国山区较多，泥石流灾害是山区常见的一种自然灾害，由于它本身高频发生、分布区域广泛及破坏力极强，对山区人民生命、财产有着极大的威胁，对防灾减灾工作提出严峻的考验。泥石流的早期预报可以有效减少灾害的损失，泥石流形成主要有三大条件，分别是地形地貌、松散物源、水源[1]。近年来国家对地质灾害的防灾减灾比较重视，陆陆续续出台政策，随着灾害的频繁发生，对泥石流灾害的研究一直都是热度较高的课题，相关学者针对泥石流的研究主要有：1）通过灾害区域地面调查结合相关遥感技术，观察并分析泥石流灾害全域的地形地貌，从而分析其成灾机理[2]；2）对物源动储量、泥沙补给、流量等影响因素通过力学及流变学的理论建立相关泥石流的运动方程[3-4]；3）通过实时监测收集雨量信息，对降雨强度与临界雨量阈值分析并建立雨量模型[5]；4）通过实时监测采集成灾因子，对泥石流发生的概率及等级进行预报，从而达到提前预报预警提示，减少灾害重大损失[6]。随着机器学习理论的不断发展，非线性模型也被广泛应用在泥石流灾害预测的理论中，文献［7］融合泥石流的多个影响因子，通过遗传规划法建立临界降雨指数智能预测模型；文献［8］基于 PCA（principal component analysis）筛选泥石流灾害成灾因子并使用 BP（back propagation）神经网络对泥石流发生的危险性进行预测，此方法选用有效成灾因子的方法结合预测模型极大提升了泥石流危险性的预测，但是使用 PCA 筛选因子处理非线性关系有一定缺陷。文献［9］使用混合核函数改进了 KPCA 筛选因子算法，预测等级达到一定的提升。基于这一思想，学者[10-12]将成灾因子筛选、模型参数寻优等优化模型方式使得预测模型精度更加稳健。本文借鉴这一思想分析泥石流全域地形地貌成灾机理并筛选因子，构造出泥石流灾害预测模型。

为进一步提升泥石流预测的精度，本文以陕西省山阳县中村镇泥石流全域为研究对象，首先分析灾害区域地形地貌选出成灾因子，避免因使用单因子预测导致的精度低问题；其次采用核主成分分析法（KPCA，Kernel principal component analysis）进行高维度影响因子的筛选；另外构造最小二乘支持向量机（LSSVM，least squares support vector machines）模型对泥石流发生概率模型的建立，相对于支持向量机将非线性问题转换为线性问题求解方式极大的简化，同时使用多算法进行 LSSVM 中的超参数优化，经过优化后的模型解决了过早收敛导致陷入局部最优的问题以及参数随机选取导致的精度不佳问题，通过模型优化提高了泥石流预测的精度。最后通过与其他寻优预测算法进行比对，比对出最佳预测模型，为泥石流地质灾害研究带来活力及新思路。

1　算法理论

1.1　KPCA 方法

主成分分析方法[13]是一种无监督降维算法，针对线性数据效果较好，但是其对于数据之间存在非线性关系时 PCA 降维效果比较差，为了弥补这一缺陷，在 PCA 计算协方差矩阵时加入核函数，用来解决非线性映射问题。KPCA 在高维特征空间对原数据映射，经过 PCA 对高维数据计算特征向量及特征值来确定主成分因子。

假设样本集：$X=\{x_1, x_2, \cdots, x_M\}$，$x_k \in R^N$ 为列矢量，M 为样本总数，φ 为满足 $\sum_{k=1}^{M}\varphi(x_k)=0$ 的非线性的映射，F 为相应空间。协方差矩阵表达：

$$C = 1/M \cdot \sum_{j=1}^{M}\varphi(x_j)(x_j)^T \tag{1}$$

通过特征分解 C 值得出：

$$\lambda v = Cv \tag{2}$$

当所有特征值 $\lambda \geqslant 0$，v 为由 $\varphi(x_1)$，$\varphi(x_2)$，…，$\varphi(x_M)$ 组成的空间，所以式（2）等于：

$$\lambda(\varphi(x_k), v^r) = (\varphi(x_k), Cv^r) \qquad k = 1,2,\cdots,M \tag{3}$$

由于 v^T 是 $\varphi(x)$ 的线性组合，所以得出：

$$v^T = \sum_{n=1}^{M}\varphi(x_i)C_i^r \tag{4}$$

将式（1）、（4）代入式（3），并令 $K_{ij}=(\varphi(x_i), \varphi(x_j))$ i，$j=1$，2，…，M，代入得出：

$$M\lambda^r \boldsymbol{c}^r = Kc^r \tag{5}$$

$M\lambda^r$ 为特征值，$\boldsymbol{c}^r$ 为特征向量，当满足 $\boldsymbol{c}^r>0$ 条件：$\boldsymbol{c}^p$，$\boldsymbol{c}^{p+1}$，…，$\boldsymbol{c}^M$，进行归一化后得出：

$$M\lambda^r(\boldsymbol{c}^r, \boldsymbol{c}^r) = 1 \tag{6}$$

求得 $\varphi(x)$ 在 $\boldsymbol{c}^r$ 特征向量的投影：

$$g_r(x) = (v^r, \varphi(x)) = \sum_{i=1}^{M}\boldsymbol{c}^r(\varphi(x_i), \varphi(x_j))$$
$$r = p, p+1, \cdots, M \tag{7}$$

$g(x)$ 为 $\varphi(x)$ 非线性主元分量，$g(x)$ $[g_1(x), g_2(x), \cdots, g_l(x)]^T$ 为所有投影矢量表示。使用核函数 $K(x_i, x_j)=\langle\varphi(x_i), \varphi(x_j)\rangle$ 求解 $g(x)$ 代替空间的点积运算，核函数变为：

$$g(x) = (v^T, \varphi(x)) = \boldsymbol{K}(x_i, x_j) \tag{8}$$

当 $\varphi(x) \neq 0$ 时，空间样本变换：

$$(x_i) = \varphi(x_i) - 1/M\sum_{i=1}^{M}\varphi(x_i) \tag{9}$$

通过式（8）计算矩阵 $\boldsymbol{K}$，再依据样本变换求取特征向量与特性值，最后依据最大特征值及其对应向量结合输入属性得到主成分。按照式（10）、（11）得出各个成分的贡献率与累计贡献率。

$$\frac{\lambda_i}{\sum_{k=1}^{p}\lambda_k}(i = 1,2,\cdots,p) \tag{10}$$

$$\frac{\sum_{k=1}^{i}\lambda_i}{\sum_{k=1}^{p}\lambda_k}(i = 1,2,\cdots,p) \tag{11}$$

1.2　LSSVM 模型

LSSVM（least squares support vector machines）[14-15]基

于 SVM 将不等式约束转换为等式约束，从而化简 lagrange 乘子 α 求解，对求解 QP 问题转为进行线性方程组的求解。LSSVM 继承了 SVM 的泛化能力和鲁棒性，但其计算效率优于原始的 SVM。给定训练的数据集合（x_i，y_j），$i=1$，2，…，n，分别给出 SVM 及 LSSVM 需求解的问题。

SVM 不等式约束问题：

$$min_{\omega,b,\xi}J(\boldsymbol{\omega},\zeta)=c\cdot\sum_{i=1}^{n}\zeta_1+\frac{1}{2}\boldsymbol{\omega}T\boldsymbol{\omega}$$

$$s.t.\ y_i[\boldsymbol{\omega}^T\cdot\varphi(x_i)+b]\geqslant 1-\zeta_i,i=1,2,\cdots,n \quad (12)$$

LSSVM 等式约束问题：

$$min_{\omega,b,e}J(\boldsymbol{\omega},e)=\frac{1}{2}\gamma\cdot\sum_{i=1}^{n}e_i^2+\frac{1}{2}\boldsymbol{\omega}T\boldsymbol{\omega},\gamma>0$$

$$\text{s.t. } y_i[\boldsymbol{\omega}^T\cdot\varphi(x_i)+b]=1-e_i,i=1,2,\cdots,n \quad (13)$$

ξ 及 e 为松弛变量，用于 SVM 及 LSSVM 中引入离群点，c 及 γ 为平衡寻找最优超平面与偏差量之间最小值，ω 为权重向量，b 为误差，φ（·）为映射函数。

使用 Lagrangea 方法对式（13）优化，转化为单一的参数，求解 α 的极限值，构造出：

$$L(\boldsymbol{\omega},b,e,\alpha)=$$

$$J(\omega,e)-\sum_{i=1}^{n}\alpha_i\{y_i[\boldsymbol{\omega}^T\cdot\varphi(x_i)+b]-1+e_i\} \quad (14)$$

其中：α_i 为拉格朗日乘子。

$\boldsymbol{\omega}$，b，e_i，α_i 分别求导=0：

$$\begin{cases}\dfrac{\partial L}{\partial\omega}=0\rightarrow\boldsymbol{\omega}=\sum\limits_{i=1}^{n}\alpha_i,y_i\varphi(x_i)\\ \dfrac{\partial L}{\partial b}=0\rightarrow\sum\limits_{i=1}^{n}\alpha_i,y_i=0\\ \dfrac{\partial L}{\partial e_i}=0\rightarrow\alpha_i=\gamma e_i,i=1,2,\cdots,n\\ \dfrac{\partial L}{\partial\alpha_i}=0\rightarrow y_i[\omega T\varphi(x_i)+b]-1+e_i=0,i=1,2,\cdots,n\end{cases} \quad (15)$$

φ 依据 4 个求导的条件可列出线性方程组：

$$\begin{bmatrix}0 & \boldsymbol{I}_n^T\\ \boldsymbol{I}_n & \Phi+\boldsymbol{E}/\gamma\end{bmatrix}\begin{bmatrix}b\\ \alpha\end{bmatrix}=\begin{bmatrix}0\\ y\end{bmatrix} \quad (16)$$

$\boldsymbol{I}_n$ 为单位矩阵的转置矩阵，$\boldsymbol{E}$ 为 n 维单位矩阵，Φ 为核矩阵：

$$\boldsymbol{\Phi}_{ij}=y_iy_j\varphi(x_i)^T\varphi(x_j)=$$

$$y_iy_j\boldsymbol{K}(x_i,x_j),i,j=1,2,\cdots,n \quad (17)$$

解方程（16），可得出一组 α、b，最后得出 LSSVM 分类表达式为：

$$y(x)=\text{sign}[\sum_{i=1}^{n}\alpha_iy_i\boldsymbol{K}(x,x_i)+b] \quad (18)$$

LSSVM 的训练框架如图 1 所示，LSSVM 算法中的正则化系数和核函数参数需要进行寻优防止出现参数随机导致的精度不佳问题及过早收敛导致陷入局部最优的问题。

1.3 松鼠觅食算法

松鼠觅食算法（sparrow search algorithm）[16] 对于搜索空间中的一些复杂问题搜索能力及精度有明显优势，松鼠

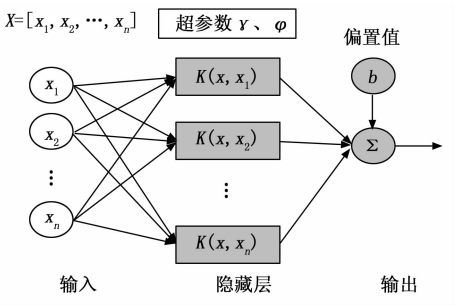

图 1 LSSVM 模型网络结构图

虽不会飞行，但可以通过滑翔的方式来躲避天敌捕食，SSA 算法就是模拟其这一行为的过程。松鼠的搜寻过程伴随其觅食的开始，寻找食物的方式通过其从不同的树木之间移动来获取，森林中不同区域的搜索通过松鼠位置的改变来实现。

假设松鼠的数量为 n，松鼠移动的位置通过矢量来确定，并在边界范围内随机初始化其位置。

$$\boldsymbol{FS}=\begin{bmatrix}FS_{1,1} & FS_{1,2} & \cdots & FS_{1,D}\\ FS_{2,1} & FS_{2,2} & \cdots & FS_{1,d}\\ \vdots & \vdots & \vdots & \vdots\\ FS_{n,1} & FS_{n,2} & \cdots & FS_{n,d}\end{bmatrix} \quad (19)$$

$FS_{n,d}$ 为第 n 只老鼠在第 d 维度上的值，松鼠在森林中的初始位置为：

$$FS_i=FS_L+U(0,1)\times(FS_U-FS_L) \quad (20)$$

FS_U 和 FS_L 为松鼠移动的上下界，U（0，1）为随机数[0，1]。

食物源的等级通过每一只松鼠位置的适应度表示，计算适应度值并进行升序分类，适应度最小的位置：最佳食物源①山核桃，接下来三只位置正常食物源②橡树，其他的位置无食物来源③普通树。

依据其天敌的出现概率 P_{dp} 松鼠更新移动的位置。

滑翔路径一：②→①

$$FS_{at}^{t+1}\begin{cases}FS_{at}^{t}+d_g\times G_c\times(FS_{ht}^{t}-FS_{at}^{t}), & R_1\geqslant P_{dp}\\ \text{Randomlocation}, & R_1<P_{dp}\end{cases} \quad (21)$$

滑翔路径二：③→②

$$FS_{nt}^{t+1}\begin{cases}FS_{nt}^{t}+d_g\times G_c\times(FS_{at}^{t}-FS_{nt}^{t}), & R_2\geqslant P_{dp}\\ \text{Randomlocation}, & R_2<P_{dp}\end{cases} \quad (22)$$

滑翔路径三：③→①

$$FS_{nt}^{t+1}=\begin{cases}FS_{nt}^{t}+d_g\times G_c\times(FS_{ht}^{t}-FS_{nt}^{t}), & R_3\geqslant P_{dp}\\ \text{Randomlocation}, & R_1<P_{dp}\end{cases} \quad (23)$$

d_g 为随机滑翔距离，$R_1R_2R_3$ 为 [0，1] 范围内的随机数，FS_{at}^{t} 为松鼠在橡树上位置，FS_{ht}^{t} 为松鼠在山核树上位置，FS_{nt}^{t} 为松鼠在普通树上位置，G_c 为滑动系数。

季节的变化会影响松鼠的觅食活动，使用季节的变换来防止出现算法陷入局部最优。

$$S_c^t = \sqrt{\sum_{z=1}^{3}\sum_{k=1}^{d}(FS_{at,k}^{t,z} - FS_{ht,k})^2} \tag{24}$$

$$S_{\min} = \frac{10E-6}{365^{2.5t/t_m}} \tag{25}$$

t、t_m 分别为当前值和最大迭代值，$S_{\min}$为季节常数最小值，S_c^t 为季节常数，季节变换条件为 $S_c^t < S_{\min}$，满足此条件，普通松鼠位置随机改变。

$$FS_{nt,i}^{t+1} = FS_{i,L} + Levy(FS_{i,U} - FS_{i,L}) \tag{26}$$

$FS_{i,U}$和 $FS_{i,L}$为松鼠移动的上下界，Levy 为列维分布，有效地全局搜索，来找到距离当前地点最优的一个新地点。SSA 算法具体步骤如图 2 所示。

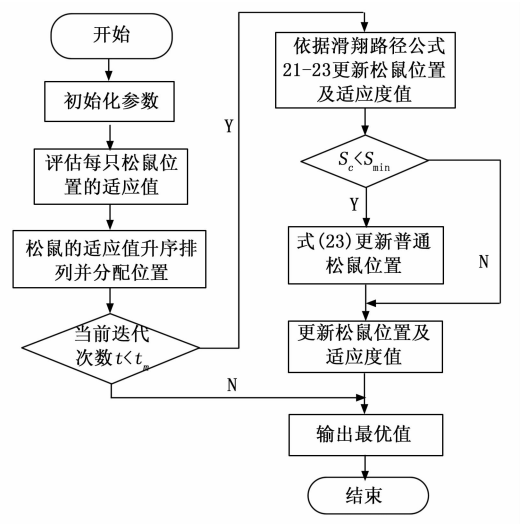

图 2　SSA 算法流程

2　基于 KPCA-SSA-LSSVM 山区泥石流灾害预测模型

基于 KPCA-SSA-LSSVM 的山区泥石流灾害发生预测流程如图 3 所示，具体实现步骤如下：

1）首先对监测的原始数据预处理，并使用 KPCA 核主成分分析法筛选出覆盖率满足需求的 6 个影响因子。

2）数据集合理划分，确定训练集及测试集百分比。

3）初始化寻优参数及 LSSVM 参数。

4）根据各影响因子建立 LSSVM 预测模型，并在训练集训练最佳适应度函数。

5）将不同预测模型在测试集上分析对比，筛选得出最佳模型及预测结果。

3　仿真验证和结果分析

3.1　研究区概况及数据来源

陕西省商洛市山阳县的中村镇，因其地处秦岭山下，山脉沟壑众多，属于中、低山地形，山体土石量多达 180 多万方，占地高达 80%以上，位于地势差异较大的峡谷地区，地形地质复杂，山体石量多，更易引发灾害。同时也属于长江流域汉江水系，地区水源较多，河流较多、尤其在夏秋季降雨量也较多，年平均降水量达到 671～865 毫米，如

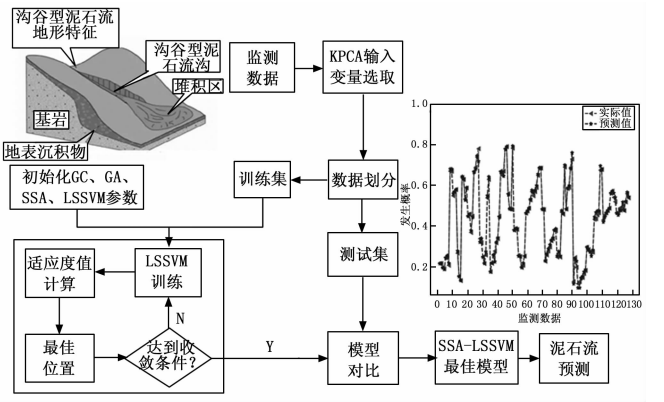

图 3　KPCA-SSA-LSSVM 算法流程

果连续降雨量大容易导致土质疏松[17]，从而增加了地质灾害的安全隐患点。

参照《T/CAGHP 006-2018 泥石流灾害防治工程勘察规范》、《滑坡崩塌泥石流灾害调查规范（1：50 000）（DZ/T0261-2014）》[18]结合山阳县实地监测区域监测泥石流活动数据，对泥石流发育机制及成灾特征分析，本研究区域按照水源和物源成因划分为崩塌型泥石流，其中固体物质主要由滑坡崩塌等重力侵蚀提供[19-20]。去掉规范量级评分表中 5 分以下影响因子，最后结合监测区实际泥石流数据得出 11 个影响因子，分别为沟岸山坡坡度、降雨量（24 h、1 h、10 min 最大降雨量）、土壤含水率、沟床平均坡度、岩性影响、流域相对高差、河沟堵塞程度、河沟纵坡、产沙区沟槽横断面、松散物平均厚度、流域面积、泥砂沿程补给长度比、孔隙水压力、沿沟松散物量、区域构造影响、流域植被覆盖率。因各降雨量 24 h、1 h、10 min 最大降雨量对泥石流的发生有极大的影响，所以选取暴雨强度 R 作为泥石流灾害模型的影响因子。暴雨强度 R 计算如式（27），各参数选取如表 1 所示。

$$R = K(H_{24}/H_{24(D)} + H_1/H_{1(FD)} + H_{1/6}/H_{1/6(D)}) \tag{27}$$

表 1　泥石流降雨量因子参数

参数	K	$(H_{24}/H_{24(D)})$/mm	$(H_1/H_{1(D)})$/mm	$(H_{1/6}/H_{1/6(D)})$/mm
描述	前期雨量修正系数 $K=1$ 无前期降雨 $K>1$ 有前期降雨 取值 $K=1.15$	H_{24}：24 h 最大降雨量 $H_{24(d)}$：24 h 临界值 取值 $H_{24(D)}=30$ mm	H_1：1 h 最大降雨量 $H_{1(D)}$：1 h 临界值 取值 $H_{1(D)}=15$ mm	$H_{1/6}$：10 min 最大降雨量 $H_{1/6(D)}$：10 min 临界值 取值 $H_{1/6(D)}=6$ mm

3.2　数据预处理

监测数据由于环境的影响会出现一些如缺失、离群或维度不统一的数据，这些数据对于模型的建立有极大的消极影响，会产生跳跃，且无法与其他数据统一，因此需要对监测数据进行预处理。

1）异常值处理：监测数据中存在一部分偏离传感器本身范围的值或偏离观测值较大的值，不处理会影响数据预测准确性，距离达到 5 倍或者与均值的距离≥3 倍标准差的数据称之为离群点。

2）缺失值的处理：监测数据通过泥石流灾害区域多传感器实时传输，传输过程中经常会出现遗漏或者个别离群点情况，会出现失真损失有效信息，导致属性值缺失不准确。按照属性因素方法进行统计得出缺失率，本文划分两种类别数据的缺失值，如表 2 所示。

表 2 数据缺失值

类型	类别型	数值型
$q\geqslant 90\%$	缺失值属性剔除	缺失值属性剔除
$40\%\leqslant q<90\%$	缺失值属性作为一种新的类别	相邻属性加权值填充
$20\%\leqslant q<40\%$	多重插补	均值填充
$q\leqslant 20\%$	同类均值插补	众数填充

3）数据归一化：监测数据种类较多样且数量较多，多传感器数据量纲不同有较大的差异，使用原始数据直接建模对于预测的准确性有极大的影响，所以需要对数据进行归一化处理，归一化处理公式如（28）：

$$R=\frac{R-R_{\min}}{R_{\max}-R_{\min}} \tag{28}$$

式（28）中，R 为某因素归一化处理后的数据，$R_{\min}$ 和 $R_{\max}$ 表示某因素数据中的最小值及最大值。

3.3 数据影响因子筛选

由于样本影响因子彼此之间存在相关性，为避免相关性对预测结果的影响，本文通过 KPCA 核主成分分析法选取成灾因子，各主成分的特征值及贡献率如图 4 所示，实验表明前 6 个主成分的累计的贡献率已经达到 95.48%，覆盖的信息超过了 90%，覆盖率达到要求，所以文中选取前 6 个影响因子作为泥石流灾害模型训练的输入数据。并依据《T/CAGHP 006-2018 泥石流灾害防治工程勘察规范》及泥石流相关资料分析，得出影响因子与泥石流发生量化等级关系如表 3 所示。以陕西省山阳县重点地灾监测区的历史数据作为研究样本，数据使用山阳县 2019 年 4 月到 2021 年 4 月的 10 个监测点的数据作为数据集，经过数据预处理及成灾因子选取后数据集总共筛选出 1 300 组数据，分别分为 80%测试集和 20%的两个验证集。

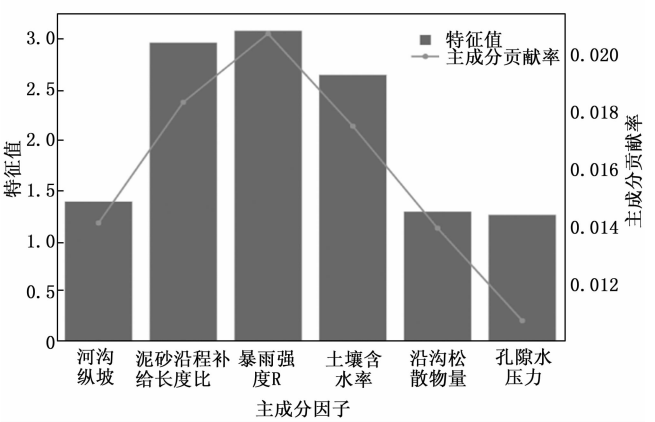

图 4 影响因子特征值及主成分贡献率

表 3 泥石流影响因子量化表

序号	影响因子	量化等级	
1	河沟纵坡	A:>12°	B:6～12°
		C:3～6°	D:<3°
2	泥砂沿程补给长度比	A:>60%	B:30%～60%
		C:10%～30%	D:<10%
3	暴雨强度 R	A:>10	B:4.2～10
		C:3.1～4.2	D:<3.1
4	土壤含水率	A:>29%	B:19%～29%
		C:10%～19%	D:<10%
5	沿沟松散物量	A:>10	B:5～10
		C:1～5	D:<1
6	孔隙水压力	A:>60%	B:10%～30%
		C:30%～60%	D:<10%

注：A. 泥石流发生严重；B. 泥石流中等发生；C. 泥石流轻微发生；D. 泥石流不发生。

3.4 超参数寻优

LSSVM 建模过程中调优参数为正则化系数和核函数 φ 参数，文中选取 SSA 寻优算法与遗传算法（GA，genetic algorithm）及网格搜索（GC，gridsearchCV）在相同 1 040 组训练集对 LSSVM 模型的正则化系数 γ 和核函数 φ 参数进行寻优。种群的规模设置为 90，最大迭代次数设置为 200，每个优化算法分别进行 60 次独立实验，并分别画出最优适应度函数值与迭代次数曲线图进行比对，结果如图 5 所示，适应度函数值随着迭代次数的增加而逐渐减小，最终搜寻到最优参数后收敛。GC 在第 16 次迭代大幅下降。跳出了局部最优状态，GA 整个迭代过程收敛速度较慢，但也逐渐趋向最优，SSA 优化效果最好，明显引导种群向最优位置处，说明使用 SSA 算法寻优，对松鼠移动的位置不断调整可以跳出局部最优值，且收敛速度快，而且早熟现象明显，能够取得更小的适应度。最终选取正则化系数 $\gamma=0.274$ 和核函数 $\varphi=7.642$。

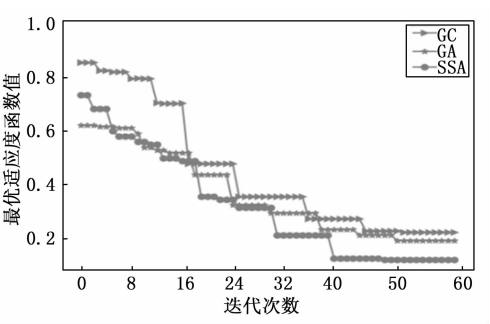

图 5 寻优适应度曲线对比图

3.5 仿真验证及结果分析

为验证模型的预测精度，引入模型评价指标 AUC 值，

其为 *ROC* 曲线结合坐标轴围成的面积值，范围一般介于[0.5～1] 之间，预测的真实性取决与 *AUC* 值接近 1 的程度，靠近 1 真实性高反之则反。*MAE* 预测值真实误差，*RMSE* 预测值和真实值间偏离程度，*MSE* 真实值与预测值差异，越接近零预测精度越高，如式（29）所示：

$$\begin{cases} RMSE = \sqrt{MSE}\sqrt{\sum_{i=1}^{I}(y_i - y_i')^2/I} \\ MAE = |\ y_i - y_i'\ |\ /I \\ AUC = 1 - FP/(FN + TN) - FN/(TP + FP)/2 \end{cases} \tag{29}$$

其中：I 样本数量，y_i' 为模型预测输出，y_i 为真实输出。

为验证本研究优化模型的准确性，将经过数据预处理及降维后的训练数据作为泥石流预测模型构建的输入数据，总共 1 040 组训练集构建泥石流预测模型，并通过 10%验证集 1 验证各模型的准确性。实验采用 LSSVM 作为泥石流灾害预测模型，并用 SSA 算法超参数寻优。使用同一个验证集验证未优化的 LSSVM 模型及其他寻优算法对 LSSVM 预测效果比对。利用预测结果计算模型的 *MAE*、*MSE* 和 *RMSE*，值越接近于零精度越高，可以看出 SSA-LSSVM 的 *MAE*、*MSE* 和 *RMSE* 最小且接近于零，对比评估指标结果如图 6 所示，传统的 LSSVM 相对误差较大，最大相对误差达到 1.72%，而 SSA-LSSVM 最大误差达到 0.19%，误差是最低的，进一步说明了该模型预测的精度较高。

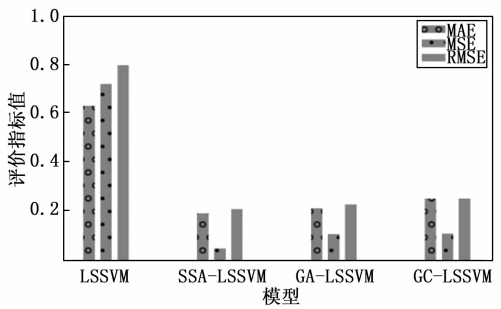

图 6　模型预测评估指标

为进一步验证模型的稳健性，选取 10%验证集 2，将其打乱的 130 个监测数据作为模型预测概率及预测等级误差的评价，图 7 为各模型寻优 LSSVM 模型后的实际发生概率与预测发生概率比对图。SSA 寻优后实际值与预测值基本吻合，拟合情况较好，极限的几个数据 27、38、89 及 111 发生概率存在一些差异，但是其对应的风险预报等级与实际数据风险等级结果吻合，不影响预报的等级，多个算法模型在同一预测集上的预测等级结果如图 8 所示，按照泥石流发生等级准确率降序排列：SSA 达到 100%，GA-LSSVM 达到 92.3%，GC-LSSVM 达到 90%。实验说明引入

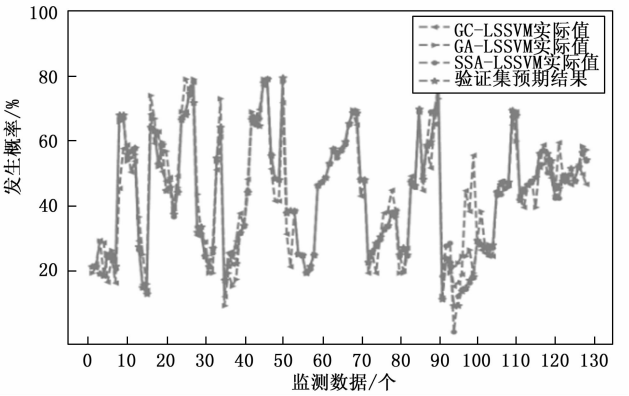

图 7　模型预测预测比对图

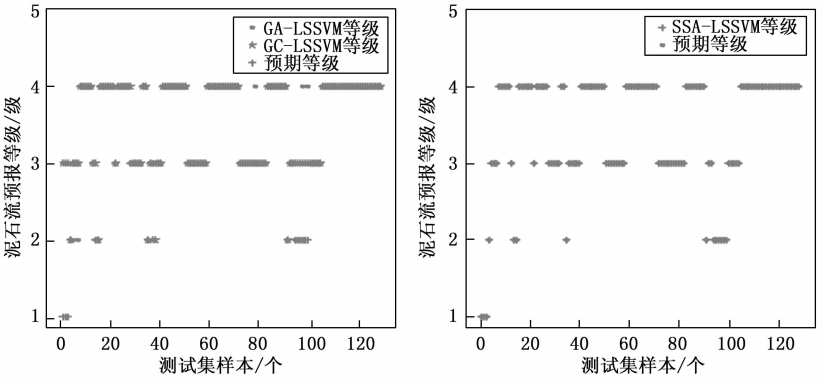

图 8　发生概率等级预测图

SSA 对 LSSVM 参数寻优，泥石流发生的概率及等级预测准确率皆有明显的提升。使用寻优后的 SSA-LSSVM 模型对研究区域泥石流进行预测，从发生的概率及预测的等级两方面都证明该模型具有较高的可行性。

此外通过 *AUC* 公式计算得出 20%验证集中各模型的 *ROC* 曲线如图 9 所示，*ROC* 曲线中横坐标为假阳性率（FPR/1-Specificity 特异度），纵坐标为真阳性率（TPR/Sensitivity），可以根据 *ROC* 曲线的面积下的 *AUC* 值看出各个预测模型对应评价指标的好坏，*AUC* 值越高说明模型精

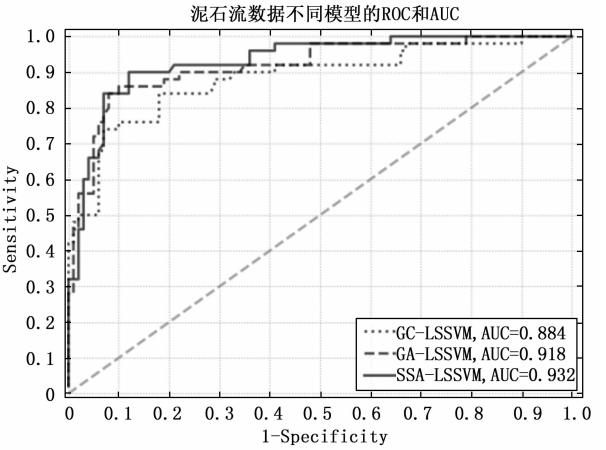

图 9　模型 *ROC* 曲线

度越佳，各模型 *AUC* 值均高于 0.88，但是 SSA-LSSVM 模型指标更加，无论从测试时间、*AUC* 值及 *ROC* 曲线均明显优于 GA 和 GC 寻优模型，模型 *AUC* 均值为 0.932，预测效果较其他模型理想。各模型的对比指标参数如表 4 所示，SSA-LSSVM 模型相比其它模型平均测试事件最短且平均 *AUC* 值最高且接近于 1。多组实验数据结果证明 SSA-LSSVM 模型具有较好的预测效果，在泥石流灾害预测中有较好的预测能力。

表 4 模型参数及结果比对

模型及训练测试样本	模型:LSSVM 训练样本:1 040 组 测试样本:260		
组寻优算法	GC	GA	SSA
模型参数	γ=0.286 φ=6.985	γ=0.295 φ=7.026	γ=0.274 φ=7.642
平均测试时间	0.025 4	0.021 2	0.014 3
平均 *AUC* 值	0.884	0.918	0.932

4 结束语

本文以普适的山阳县中村镇区域泥石流为例，结合泥石流全域的地形地貌成灾机理，提出 KPCA-SSA-LSSVM 泥石流预测模型，在研究区实践应用效果良好，说明模型具有一定的可行性和有效性。因此，可以得出以下结论：

1）参照《T/CAGHP 006-2018 泥石流灾害防治工程勘察规范》、《滑坡崩塌泥石流灾害调查规范（1：50 000）(DZ/T0261-2014)》并结合山阳县实地监测区域，监测泥石流活动数据，对泥石流发育机制及成灾特征分析，选出 11 个成灾因子，并使用 KPCA 主成分分析法依据因子的贡献率筛选出 6 个重要的成灾因子；

2）LSSVM 建模过程中调优参数为正则化系数和核函数参数，选取 SSA 寻优算法与遗传算法（GA，genetic algorithm）及网格搜索（GC，GridSearchCV）在相同 1 040 组训练集对 LSSVM 模型参数进行寻优，解决参数随机导致的精度不佳问题及陷入局部最优问题。

3）将寻优后的 SSA-LSSVM 预测结果与 GA、GC 参数寻优模型预测结果比对，从 *AUC* 值、*MAE*、*MSE*、*RMSE* 评价指标都验证了 SSA-LSSVM 预测的精度。

4）使用寻优后的 SSA-LSSVM 模型对研究区域泥石流进行预测，从发生的概率及预测的等级两方面都证明该模型具有较高的可行性。

参考文献：

[1] 何 伟．芦山县洞峡子沟泥石流工程地质特征及成因机制分析[J]．地质灾害与环境保护，2019，30（4）：3－8.

[2] 李 宁，唐 川，史青云，等．九寨沟震区“6·21”泥石流成因与致灾机制研究[J]．工程地质学报，2022，30（3）：740－750.

[3] 罗超鹏，常 鸣，武彬彬，等．基于 FLOW-3D 的泥石流龙头运动过程模拟研究[J]．中国地质灾害与防治学报，2022，33（6）：53－62.

[4] 张永军，李 松．基于“甘肃武都火烧沟”公式的泥石流横向分布计算[J]．甘肃地质，2021，30（2）：50－53.

[5] 罗小惠，裴向军，郭 斌．长白山天池地区泥石流激发雨型与临界雨量[J]．科学技术与工程，2020，20（18）：7183－7191.

[6] 徐根祺，曹 宁，李 璐，等．基于改进粒子群优化 ν 支持向量机的泥石流灾害预测模型[J]．国外电子测量技术，2022，41（9）：73－81.

[7] 翟淑花，冒 建，南 赟，等．基于遗传规划的泥石流多因子融合预测方法[J]．中国地质灾害与防治学报，2020，31（6）：111－115.

[8] 刘育林，周爱红，袁 颖．基于 GRA-PCA-BP 神经网络模型的单沟泥石流危险性预测[J]．河北地质大学学报，2022，45（4）：44－51.

[9] 李丽敏，程少康，温宗周，等．基于改进 KPCA 与混合核函数 LSSVR 的泥石流预测[J]．信息与控制，2019，48（5）：536－544.

[10] 李丽敏，张 俊，温宗周，等．基于布谷鸟优化轻量梯度提升机的泥石流预测[J]．科学技术与工程，2021，21（30）：13177－13184.

[11] 王英杰，丁明涛，张明慧．灰色 GM（1，1）模型在汶川县泥石流活动趋势预测中的应用[J]．地质灾害与环境保护，2020，31（1）：23－29.

[12] 张 研，吴康丽，邓雪沁，等．基于相关向量机的蒋家沟泥石流平均流速预测模型[J]．自然灾害学报，2019，28（6）：146－153.

[13] 李 璐．基于机器学习的滑坡地质灾害预报模型研究[D]．西安：西安工程大学，2019.

[14] 金 龙，曾德智，孟可雨，等．基于 GWO-LSSVM 算法的海底管道腐蚀预测模型研究[J]．石油与天然气化工，2022，51（2）：70－76.

[15] 肖亚宁，孙 雪，张亚鹏，等．基于 SOA-LSSVM 的 SLS 成形工艺参数优化研究[J]．机床与液压，2022，50（6）：36－42.

[16] 商立群，李洪波，侯亚东，等．基于 VMD-ISSA-KELM 的短期光伏发电功率预测[J]．电力系统保护与控制，2022，50（21）：138－148.

[17] 徐根祺，李丽敏，温宗周，等．基于宽度学习模型的泥石流灾害预报[J]．山地学报，2019，37（6）：868－878.

[18] 石振明，吴 彬，郑鸿超，等．泥石流防治措施与冲击力研究进展[J]．地球科学，2022，47（12）：4339－4349.

[19] HUANG Y，ZHANG B. Challenges and perspectives in designing engineering structures against debris-flow disaster[J]. European Journal of Environmental and CivilEngineering，2022，26（10）：4476－4497.

[20] ZHANG X Z，TANG C X，LI N，et al. Investigation of the 2019 Wenchuan County debris flow disaster suggests nonuniform spatial and temporal post-seismic debris flow evolution patterns[J]. Landslides，2022，19（8）：1935－1956.