Regional Landslide Sensitivity Analysis Based on CPSO-LSSVM

Yanze Li[1(✉)], Zhenjian Yang[1,2], Yunjie Zhang[1], and Zhou Jin[1]

[1] School of Computer and Information Engineering, Tianjin Chengjian University, Tianjin 300384, China
1341485766@qq.com

[2] Tianjin Key Laboratory of Soft Soil Characteristics and Engineering Environment, Tianjin, China

Abstract. Landslide sensitivity analysis is of great significance for predicting landslide hazards. Taking the landslide in the hilly area of Sichuan Province as an example, through the interpretation of high spatial resolution remote sensing images and the analysis of the occurrence mechanism of landslides in the low hilly areas of Sichuan Province, eight landslide susceptibility evaluation factors were obtained. (elevation, slope, terrain relief, rivers, roads, geotechnical types, NDVI, fault structures). Then, using the neighborhood statistical analysis, ArcGIS technology and other methods to obtain training sample data and regional sample data. According to the characteristics of the landslide development, the Chao Particle Swarm Optimization (cpso) is used to optimize the parameters of the Least Square Support Vector Machine (lssvm), the cpso-lssvm landslide sensitivity prediction model was formed. The experimental results show that cpso-lssvm has obtained good prediction results in landslide sensitivity evaluation, and the prediction accuracy has increased to 70.5%.

Keywords: Landslide · CPSO · LSSVM · ArcGIS

1 Introduction

The Sichuan Province of China is a region with frequent geological disasters of landslides. There are 2,664 landslides caused by rainfall, and the disaster density is as high as 1/100 km^2, ranking first in the country. Moreover, the population density of the region is relatively high, and the level of regional economic development is relatively high, which makes the region one of the areas where the landslide geological disasters in China are concentrated and the damage is extremely high. Therefore, the scientific and rigorous analysis of landslide sensitivity in this area is of great practical significance for disaster mitigation and disaster risk management [1].

There are two reasons for the difficulty in accurately assessing the sensitivity of regional landslides [2]. One of them is the predisposing factor, which includes internal factors, fundamental factors such as lithology, stratigraphic structure, topography, etc.; external factors are direct Factors such as rainfall, vegetation, and human activities, and the uncertainty, high latitude, and nonlinearity of these factors make it difficult to identify the landslides and their connections. Secondly, compared with evaluating

S. Han et al. (Eds.): AICON 2019, LNICST 287, pp. 213–223, 2019.
https://doi.org/10.1007/978-3-030-22971-9_18

single landslides, the sensitivity of landslides in the evaluation area is not only the need to find the sensitivities of landslides in a certain area statistically, but also the participation of other disciplines. From a practical point of view, it is also more research significance. So far, no better way to evaluate the landslide susceptibility has been found, which is why many scholars are keen to study this topic.

Advances in human scientific research in geological exploration, GIS, and AI have catalyzed more scientific research results in evaluating landslide susceptibility, such as Artificial Neural Network models, grey system models, logistic regression prediction models, and support vector machines (SVM). Applied to the sensitivity evaluation of landslides. Among them, support vector machine (SVM) [3] has higher prediction accuracy and has been successfully applied in landslide sensitivity. SVM is a small sample learning method and can effectively solve the problem of low input space dimension. Least Square Support Vector Machine (LSSVM) [4, 5] is a derivative of support vector machine, which is suitable for solving small sample problems and solves the "big sample" required by BP neural network. The convergence speed is slow, and the shortcomings of SVM operation time are overcome. Based on this, this paper proposes to use the LSSVM model to evaluate the landslide sensitivity. However, due to the problem of premature ripening when LSSVM model parameters are optimized, the Chao Particle Swarm Optimization (CPSO) algorithm is used to optimize the LSSVM model. The CPSO has the characteristics of ergodicity and strong searching ability. Applying it to the optimization parameter process, when the LSSVM algorithm falls into the premature phenomenon, the chaotic sequence method is used to chase the LSSVM model, and the optimal solution is quickly found, which improves the accuracy and convergence speed of the solution and improves the landslide. The accuracy of prediction of sensitivity evaluation is of great significance [6, 7].

2 Model Principle

LSSVM is an algorithm for improving SVM. It replaces the inequality constraint in SVM with equality constraint, and uses the loss function to change the quadratic programming problem of SVM into solving linear equation and set input n-dimensional vector. $\{(x_1, y_1), (x_2, y_2)\ldots(x_l, y_l)\} \subset R^n \times R$, Linear function set to

$$\mathrm{f(x)} = w^T - x + b \tag{1}$$

Optimization problem is

$$\min\left(\frac{1}{2}||w||^2 + \frac{1}{2}\gamma \sum\nolimits_{i=1}^{t} \xi_i^2\right) \tag{2}$$

Constrained to

$$y_i = w^T x_i + b + \xi_i,\ i = 1, 2, \ldots, l \tag{3}$$

Define the Lagrangian function as

$$\mathrm{L} = \frac{1}{2}\|w\|^2 + \frac{1}{2}\gamma \sum\nolimits_{i=1}^{l} \xi_i^2 - \sum\nolimits_{i=1}^{l} \alpha_i \left(w^T x_i + b + \xi_i - y_i\right) \tag{4}$$

Seeking partial derivatives for each parameter

$$\frac{\partial L}{\partial w} = 0 \rightarrow w = \sum\nolimits_{i=1}^{l} \alpha_i x_i \tag{5}$$

$$\frac{\partial L}{\partial \mathrm{b}} = 0 \rightarrow \sum\nolimits_{i=1}^{l} \partial_i = 0 \tag{6}$$

$$\frac{\partial L}{\partial \xi_i} = 0 \rightarrow \alpha_i = \gamma\, \xi_i,\ i = 1, 2, \ldots, l \tag{7}$$

$$\frac{\partial L}{\partial \alpha_i} = 0 \rightarrow w^T + b + \xi_i - y_i = 0,\ i = 1, 2, \ldots, l \tag{8}$$

Equations (5)-formula (8) is expressed as a matrix form

$$\begin{bmatrix} I & 0 & 0 & -x \\ 0 & 0 & 0 & -\mathbf{1} \\ 0 & 0 & \gamma I & -I \\ x^T & \mathbf{1} & I & 0 \end{bmatrix} \begin{bmatrix} w \\ b \\ \xi \\ \alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ y \end{bmatrix} \tag{9}$$

where I is a unit matrix; $x = [x_1 \ldots x_l]; y = [y_1 \ldots y_l]; \mathbf{1} = [1 \ldots 1]^T; \xi = [\xi_1 \ldots \xi_l]$; $\alpha = [\alpha_1 \ldots \alpha_l]]$;

The solution of the formula (9) is determined by the formula (10).

$$\begin{bmatrix} 0 & \mathbf{1}^T \\ \mathbf{1} & x^T x + \gamma^{-1} I \end{bmatrix} \begin{bmatrix} b \\ a \end{bmatrix} = \begin{bmatrix} 0 \\ y \end{bmatrix} \tag{10}$$

By choosing the best kernel function $\mathrm{K(x, y)}$, we can get linear regression to

$$\mathrm{f(x)} = \sum\nolimits_{i=1}^{l} \alpha_i K(x_i, x) + b \tag{11}$$

This paper selects the radial basis kernel function and the expression is

$$\mathrm{K}\left(x_i, x_j\right) = \exp\left(\frac{-||x_i - x_j||^2}{\sigma^2}\right) \tag{12}$$

where, σ is the perceived variable, and this paper needs to optimize (γ, σ) parameters.

2.1 CPSO-LSSVM Model

The PSO algorithm is an evolutionary algorithm that imitates the foraging process of the bird group. Compared with other algorithms, the PSO algorithm has the advantages of simple algorithm implementation and few parameters, but it is easy to fall into local extremum and cause premature convergence. In this paper, in order to make the group jump out of the local optimum quickly, the chaos idea and the PSO algorithm are combined, and the chaotic sequence is used to optimize the search of the PSO algorithm, which improves the accuracy and convergence speed of the PSO algorithm.

The position and velocity of the particles are updated using Eqs. (13) and (14).

$$v_j^{k+1} = wv_j^k + c_1 r_1 \left(p_{j-}^k x_j^k\right) + c_2 r_2 (p_g^k - x_j^k) \tag{13}$$

$$x_j^{k+1} = x_j^k + v_i^{k+1} \tag{14}$$

where, v_j is particle velocity; w is the weight; c_1 and c_2 are acceleration coefficients; r_1 and r_2 are random Numbers within the range of [0,1]. p_j is the individual optimal solution; p_g is the overall optimal solution; x_j^k is the decision variable.

Its iterative formula is

$$z_{i+1} = \mu z_i(1 - z_i), i = 1, 2, 3..., \mu \subset (2, 4] \tag{15}$$

at $\mu = 4$ and $0 \leq z_i \leq 1$, the system is completely chaotic.

The CPSO algorithm is mainly used to optimize the regularization parameter γ and kernel function σ of LSSVM. I'm going to use $x_j = (\gamma_j, \sigma_j)$ for the position of the JTH particle. The fitness function is

$$\text{fitness} = \sum_{i=1}^{N} |y_i - \widehat{y_l}|/\text{N} \tag{16}$$

where, N is the sample number; y_i is the actual sample value and $\widehat{y_l}$ is the predicted sample value.

The flow chart of cpso-lssvm is shown in Fig. 1.

3 Regional Landslide Data Acquisition and Feature Extraction Were Studied

3.1 Study the Geology of the Area

The overall terrain of Sichuan Province is high in the west and low in the east, with high and low disparity, complex landforms, rich soil types, plateaus and mountains in the west, more than 3,000 m above sea level, basins and hills in the east, and altitudes between 500 and 2000 m. The experimental observation area is 112,000 km^2 and is located in the hilly area of eastern Sichuan Province, accounting for 23% of the total area of Sichuan Province, as shown in Fig. 2.

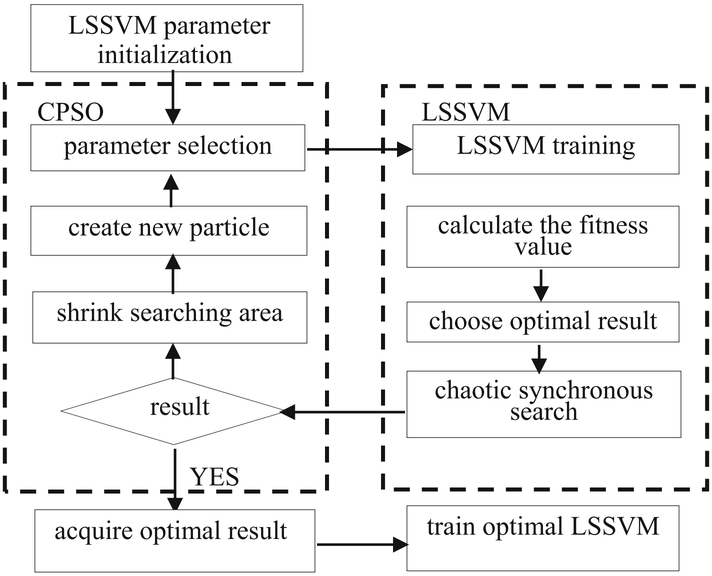


Fig. 1. CPSO-LSSVM flow chart

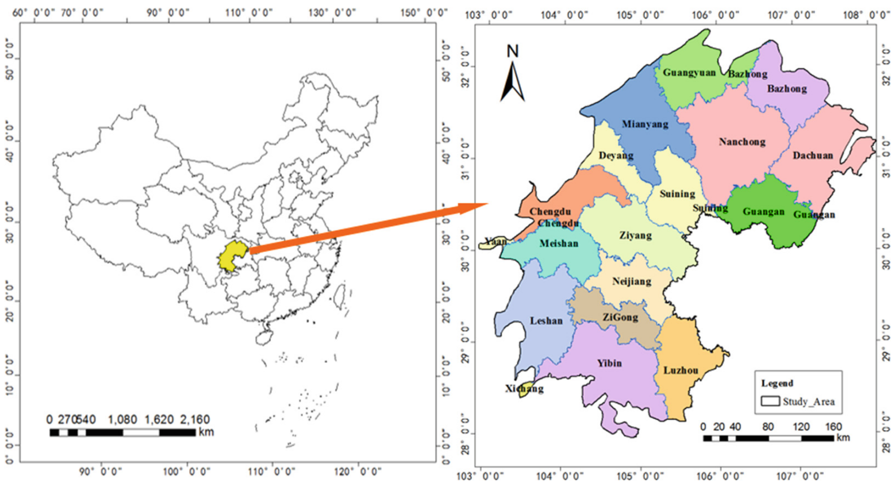


Fig. 2. Distribution of hilly areas in Sichuan Province

The administrative units included in the study area include: Chengdu, Zigong City, Zhangzhou City, Deyang City, Mianyang City, Guangyuan City, Suining City, Neijiang City, Leshan City, Nanchong City, Meishan City, Yibin City, Guang'an City, Da 18 cities (autonomous prefectures) including Chuan City, Ya'an City, Bazhong City, Ziyang City, and Liangshan Yi Autonomous Prefecture. From the perspective of world geography, Sichuan is located in the Asia-Europe earthquake zone. From the geographical division of China, it is located in the southwest seismic belt. The crustal movement is active, the earthquake is frequent, the mountain structure is unstable, and some areas are in the monsoon climate zone. The dry rainy season is distinct and the precipitation is concentrated. In the summer, there are many heavy rains, and there are many types of meteorological disasters, and the frequency of occurrence is high, so it is easy to cause landslide phenomenon.

3.2 Landslide Data Acquisition and Feature Extraction

3.2.1 Landslide Data Acquisition

The various influencing factors leading to landslides are called landslide susceptibility factors [8], including topography, geological lithology, artificial slopes (roads, reservoirs, etc.), natural slopes (rivers, lakes, etc.), vegetation cover, geological structures. Wait. Whether it can be selected as the landslide susceptibility factor has the following two criteria: First, the landslide susceptibility factor should be related to the landslide activity, that is, whether it is the root cause or direct cause of the landslide phenomenon; secondly, landslide sensitivity factor is obtainable and quantifiable. Under the analysis of the observation of Sichuan hilly area, firstly, considering the disaster-causing mechanism of landslides, landslides are mainly affected by topography, lithology, geological structure, slope cutting and underlying surface conditions. Second, consider to the availability of data, select elevation, slope, terrain relief, fault structure, geotechnical types, rivers, roads, NDVI. Wang et al. [9] analyzed the landslide impact factors of the region based on the variable dimension fractal theory and the deterministic coefficient probability model, and used the multi-layer perceptron model (CF-MLP) to carry out the landslide susceptibility in the region. Evaluation [10] verified the importance of the above eight landslide impact factors on the landslide impact in the region. The details of the eight landslide impact factors are shown in Table 1.

Table 1. Sensitivity index system construction table of low-mountain and hilly landslide in Sichuan Province

Number	The main class	The class	The data format	Access	Scale
1	Topography	Elevation	GeoTiff	DEM	30 m × 30 m
		Slope	GeoTiff	Extraction based on DEM	30 m × 30 m
		Relief	GeoTiff	Extraction based on DEM	30 m × 30 m
2	Cut slope	The river	Shape File	River network map	1:250,000
		The road	Shape File	Dart map	1:250,000
3	The lithology	Rock and soil types	Shape File	Geological map extraction	1:500,00
4	Vegetation coverage	NDVI	GeoTiff	MODIS offers products	30 m × 30 m
5	Geological structure	Fracture structure	Shape File	Geological map extraction	1:500,000

3.2.2 Landslide Data Preprocessing

The eight landslide impact factors mentioned in Sect. 3.2.1 of this paper constitute the conditional attributes of the LSSVM model. The state of the landslide (landslide state: 1; non-landslide state: 0) constitutes the decision attributes of the model. In order to achieve data standardization, different hierarchical quantization methods are used for different factors. For numerical factors, if the classification has a clear geographical significance, it is graded according to its geographical significance (including: slope, terrain relief); if its classification has no clear geologic significance, it is graded at equal intervals (Including: elevation, distance from the fault structure, distance from the river, distance from the road and vegetation cover). For type value factors, no classification is required, such as geotechnical types. The grading of each factor is shown in Table 2.

Table 2. Classification of landslide impact factors in hilly areas of Sichuan Province

The main class	Factor	Grading\class	Area (%)	Factor	Classification	Area (%)
Topography	Slope	0–3°	20.57	Elevation	0–400 m	37.2
		3–5°	17.09		400–600 m	45.84
		5–15°	43.01		600–800 m	11.31
		15–25°	14.42		800–1000 m	3.41
		25–30°	2.55		1000–1200 m	1.18
		30–45°	2.13		>1200 m	1.06
		>45°	0.24			
The geological structure	Distance from fault	0–2 km	5.83	Distance from fault	10–12 km	8.84
		2–4 km	14.36		12–14 km	7.54
		4–6 km	13.68		14–16 km	6.26
		6–8 km	12.09		>16 km	5.04
		8–10 km	10.23			
Cut slope	Distance from river	0–2 km	10.14	Distance from road	0–2 km	13.6
		2–4 km	33.49		2–4 km	11.65
		4–6 km	22.16		4–6 km	10.43
		6–8 km	16.01		6–8 km	9.16
		8–10 km	11.17		8–10 km	8.38
		>10 km	7.02		10–12 km	7.48
		0–200 m	71.81		12–14 km	6.36
		200–600 m	25.96		>14 km	32.93
		>600 m	2.23			
Geological lithology	The lithology	1(loose deposits)	9.96	The lithology	6 (conglomerate)	37.7
		2 (mudstone)	16.78		7 (dolomite)	
		3 (limestone)	3.35		8 (carbonate rock)	
		4 (basalt)	0.84		9 (granite)	0.54
		5 (sandstone)	20.83		10 (shale)	
Vegetation coverage	NDVI	[–1, 1]				

When there is abnormal data or missing data when acquiring sample data, mainly due to meshing or boundary value, the ignore value method is used for such data, that is, the missing attribute values are ignored, the purpose is to keep Enter the integrity of the data. In addition, in order to reduce the computational difficulty, this paper normalizes and normalizes the attributes. The normalization formula is as follows:

$$x^{'} = (x - x_{min})/(x_{max} - x_{min}) \tag{17}$$

where, $\mathrm{x}(\mathrm{x} \in [x_{min}, x_{max}])$ is the real attribute value, x_{min} and x_{max} are the minimum and maximum values of this attribute respectively, and $x^{'} \in (0, -1)$ is the normalized attribute value.

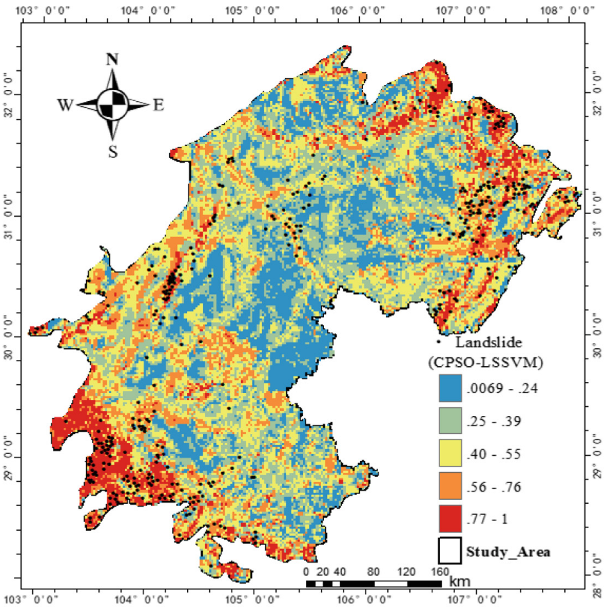
4 Simulation Research

In order to verify the validity of the algorithm model in this paper, the training sample data are arranged, the training set samples are taken from the odd items of the sample data, and the test set samples are taken from the even items of the data. Set the total number of particle swarm to 100, because the fitness converges rapidly with the number of iterations, so set the number of iterations to 100, the learning factor $c_1 = 1$, $c_2 = 5.9$, and the inertia weight $\mathrm{W} = 0.5$. The optimal parameters obtained in the end $\gamma = 25$, $\sigma = 2.76$. In order to compare the map-making effect of ArcGis [11] predicted by cpso-lssvm model, this paper also uses PSO-LSSVM for experiments. The results are shown in Fig. 3. In this paper, the research area is divided into five grades by natural breakpoint method, which are extremely low sensitivity, low sensitivity, medium sensitivity, high sensitivity and extremely high sensitivity. That is, the more sensitive the part of the landslide is, the closer it is to red. In the CPSO-LSSVM (Fig. 3A), most of the areas where the landslide points fall are highly sensitive areas, and the sensitive areas of each level are more accurately divided. In the PSO-LSSVM (Fig. 3B), some areas with dense landslide sites are not highly sensitive, which is inconsistent with the characteristics of the landslide development in the study area. Therefore, CPSO-LSSVM is more in line with the actual landslide distribution law than PSO-LSSVM.

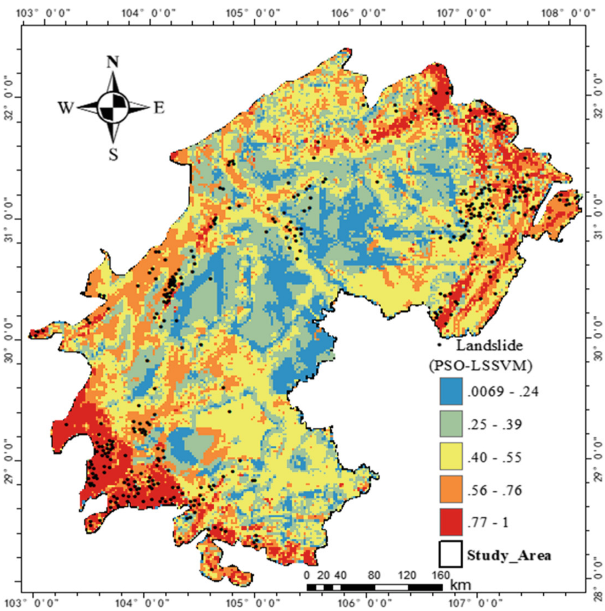
And using the ROC curve [12], the CPSO-LSSVM model PSO-LSSVM model was compared and analyzed. The ROC curve is a comprehensive indicator of the sensitivity (Sensitivity) and specificity (1-Specificity) continuous variables. Sensitivity refers to the proportion of all positive categories that are predicted to be positive, and specificity refers to the proportion of all negative classes that are predicted to be positive. Then, the ROC curve is plotted with the sensitivity as the ordinate and the specificity as the abscissa. The ROC is calculated as follows:

$$\text{Sensitivity} = \frac{TP}{TP + FN} \tag{18}$$

$$1 - \text{Specificity} = \frac{FP}{FP + TN} \tag{19}$$



(A) CPSO-LSSVM



(B) PSO-LSSVM

Fig. 3. Landslide sensitivity areas obtained by different prediction models (Color figure online)

where TP refers to positive samples with positive predictions, FP refers to positive samples with negative predictions, TN refers to negative samples with positive predictions, and FN refers to negative samples with negative predictions.

Figure 4 shows the ROC curves formed by two prediction models. As shown, $CPSO - LSSVM(AUC) = 0.716$, $PSO - LSSVM(AUC) = 0.708$. It can be seen that the prediction accuracy of CPSO-LSSVM is better than that of PSO-LSSVM.

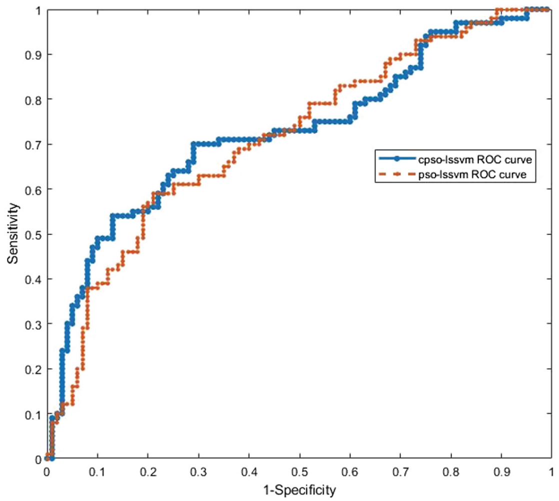


Fig. 4. ROC curves of relative ratios of different models

5 Conclusion

In view of the problem that the landslide sensitivity is difficult to accurately predict, the LSSVM is used to predict the landslide in Sichuan Province, and the hyperparameter of the model is optimized by the particle swarm optimization algorithm. Aiming at the problem of insufficient optimization ability of typical particle swarm optimization algorithm, an improved particle swarm optimization algorithm is proposed. The cpso-lssvm landslide sensitivity evaluation model was constructed, and the defect of LSSVM easily falling into the local minimum problem was overcome. In order to verify the effect of the model, the landslide point density renderings were drawn using ArcGis, and the CPSO-LSSVM model was compared with the PSO-LSSVM model in combination with the ROC curve. Finally, both evaluation methods show that CPSO-LSSVM is more suitable for landslide susceptibility evaluation, and the prediction effect is better than PSO-LSSVM, which provides a new way of thinking in the direction of landslide sensitivity prediction.

Acknowledgments. This research was supported by Tianjin Key Laboratory of Soft Soil Characteristics and Engineering Environment.

References

1. Huang, F., Yin, K., Jiang, S.: Evaluation of landslide vulnerability based on clustering analysis and support vector machine. J. Rock Mech. Eng. **37**(1), 156–167 (2008)
2. Wang, Z., Hu, Z., Zhao, W.: Determining the best statistical unit of topographic relief by using cumulative and analytical algorithm. Surv. Mapp. Sci. **39**(6), 59–64 (2014)
3. Niu, R., Peng, L., Ye, R.: Landslide susceptibility assessment based on rough sets and support vector machine. J. Jilin Univ.: Earth Sci. **42**(2), 430–439 (2012)
4. Suykens, J., Vandevalle, J.: Least squares support vector machine classifiers. Neural Process. Lett. **9**(3), 293–300 (1999)
5. Li, X., Xu, J.: Landslide deformation prediction based on wavelet analysis and LSSVM. Geodesy Geodyn. **29**(04), 127–130 (2009)
6. Liu, Q., Sun, J., Chen, X., Liu, J., Zhang, Q.: Application of CPSO-LSSVM in autoregressive clock difference forecasting. J. Jilin Univ. (Eng. Ed.) **44**(03), 807–811 (2014)
7. Liu, M.: Network intrusion detection based on CPSO-LSSVM. Comput. Eng. **39**(11), 131–135 (2013)
8. Nie, J., Lian, J., Hu, Z.: Analysis of landslide spatial characteristics in wenchuan earthquake-stricken area. Geogr. Res. **33**(2), 214–224 (2014)
9. Wang, Z., Hu, Z., Zhao, W.: Sensitivity analysis of environmental factors for rainfall-induced landslides based on probability model of deterministic coefficients-taking the low mountain and hilly areas of Sichuan Province as an example. Disaster Sci. **29**(2), 109–115 (2014)
10. Wang, Z., Hu, Z., Zhao, W.: Regional landslide sensitivity assessment based on multilayer perceptron model district as an example. J. Disaster Prev. Mitig. Eng. **35**(5), 691–698 (2015)
11. Saadatkhah, N., Kassim, A., Lee, L.M.: Qualitative and quantitative landslide susceptibility assessments in Hulu Kelang area, Malaysia. Electron. J. Geotech. Eng. **19**, 545–563 (2014)
12. Feng, F., Wu, X., Niu, R.: Landslide sensitivity evaluation of BP neural network based on particle swarm optimization. Sci. Surv. Mapp. **42**(10), 170–175 (2017)