Research and Application of the Uncertain NNSB-OPTICS Clustering Algorithm for Landslide Hazard Prediction: Postprint

Authors: Mao Yimin, Chen Huabin, Li Zhongli, Zhang Canlong

Date: 2018-05-20T00:00:00+00:00

Abstract

This research investigates the problems of ineffective characterization and processing of uncertain factors such as rainfall in landslide risk prediction, along with the issues of manual density threshold setting and high time complexity in the existing OPTICS-PLUS clustering algorithm. To improve the accuracy of landslide risk prediction, an uncertain NNSB-OPTICS clustering algorithm is proposed and applied to landslide prediction. First, the expansion strategy of the OPTICS-PLUS algorithm is optimized to avoid manual density threshold setting and improve algorithmic efficiency. Then, based on the distribution characteristics of rainfall data, an EC-type distance formula is proposed by integrating the EW-type distance formula and cloud model theory, effectively handling uncertain rainfall data. Finally, the uncertain NNSB-OPTICS clustering algorithm is applied to landslide risk prediction in Baota District, Yan' an City, establishing a landslide risk prediction model that achieves 89.7% prediction accuracy. Experimental results demonstrate that this method can effectively improve landslide risk prediction accuracy and is highly feasible.

Full Text

Preamble

Research and Application of Uncertain NNSB-OPTICS Clustering Algorithm in Landslide Hazard Prediction

Mao Yimin, Chen Huabin, Li Zhongli, Zhang Chanlong
(School of Information Engineering, Jiangxi University of Science & Technology, Ganzhou, Jiangxi 341000, China)

Abstract: Landslide hazard prediction faces challenges in effectively characterizing and processing uncertain factors such as rainfall, while existing OPTICS-

PLUS clustering algorithms require manual density threshold setting and suffer from high time complexity. To improve landslide hazard prediction accuracy, this paper proposes an uncertain NNSB-OPTICS clustering algorithm and applies it to landslide prediction. First, the expansion strategy of the OPTICS-PLUS algorithm is optimized to avoid manual density threshold setting and improve algorithm efficiency. Then, based on the distribution characteristics of rainfall data, an EC-type distance formula is proposed by integrating the EW-type distance formula and cloud model theory, enabling effective handling of uncertain rainfall data. Finally, the uncertain NNSB-OPTICS clustering algorithm is applied to landslide hazard prediction in Baota District, Yan' an City, establishing a landslide hazard prediction model that achieves 89.7% prediction accuracy. Experimental results demonstrate that this method can effectively improve landslide hazard prediction accuracy and exhibits high feasibility.

Keywords: landslide; hazard prediction; uncertain data; OPTICS algorithm

0 Introduction

Landslides are among the most widely distributed and frequently occurring geological disasters, posing serious threats to human survival and development. Landslide formation is influenced by multiple factors, including basic elements such as topography, lithology, and slope structure, as well as triggering factors with significant uncertainty like rainfall and human activities. Under the combined action of these factors, landslide occurrence is extremely complex and uncertain, making landslide hazard prediction highly challenging.

Clustering analysis is a key technique in data mining that can classify landslide data objects based on similarity among influencing factors without prior samples, extracting potentially useful information. Consequently, an increasing number of scholars have employed clustering algorithms for landslide prediction research. Reference [3] applied the fuzzy K-means algorithm to classify topographic features of road and non-road landslides, combined with GIS technology to establish a landslide hazard probability map for Clearwater National Forest in Idaho, USA, demonstrating that the method could accurately predict road-related landslides and provide important guidance for road planning. Reference [4] used the K-means clustering method to divide debris flow landslide susceptibility in the Wenchuan disaster area into five subclasses, determining hazard levels based on expert experience. The study showed that K-means clustering results were highly consistent with actual conditions and provided good classification performance. Reference [5] conducted landslide hazard prediction in Badong County, Hubei Province, selecting representative landslide influencing factors as evaluation indicators, using entropy weight and APH methods to obtain factor weights for application in K-means clustering to predict hazard levels for 86,216 prediction units. The results basically matched local landslide disaster conditions and could process multi-attribute data with practical value. Reference [6] studied the Three Gorges Reservoir' s Wanzhou District, selecting seven major disaster-causing factors as evaluation indicators, using landslide area ratio

and classification area ratio curves to grade factors, then applying K-means clustering to classify susceptibility results and establishing a susceptibility zoning map based on the GIS platform, achieving satisfactory prediction accuracy. Reference [7] analyzed spatial distribution and deformation factors of landslides to extract potential centers, using cloud model theory to describe these centers and improve the K-means algorithm, performing cluster analysis based on data point membership degrees, and applying the improved algorithm to landslide prediction in the Three Gorges Reservoir area, proving the method could effectively predict regional landslide hazard. Reference [8] employed a two-stage analysis to extract main attributes and thresholds, calculated entropy values affecting landslide occurrence, and used particle swarm optimization to improve K-means, solving the problem of K-means easily falling into local optima. Drawing a sensitivity map for Miaoli Mountain Pool National Park in Taiwan, experiments demonstrated that the improved K-means algorithm achieved higher prediction accuracy.

While these clustering algorithms have achieved certain results in landslide prediction, they are far from satisfactory for two main reasons: (a) Rainfall uncertainty is a crucial factor in landslide occurrence, yet these algorithms focus on processing continuous and discrete numerical attributes and cannot effectively characterize uncertain rainfall data in landslide prediction; (b) Traditional clustering algorithms require pre-determining the number of clusters k and cluster centers, making them ineffective for handling non-uniformly distributed data and resulting in suboptimal clustering performance on landslide datasets. Due to these limitations, traditional clustering algorithms exhibit insufficient landslide hazard prediction accuracy. Therefore, there is a need to explore a new method that can effectively handle non-uniformly distributed data clustering while simultaneously addressing uncertain rainfall data to further improve landslide hazard prediction accuracy.

The OPTICS-PLUS clustering algorithm [9] is a density-based clustering algorithm that is more suitable for non-uniformly distributed landslide prediction data compared to partition-based K-means algorithms. It employs a one-time clustering result reorganization strategy during clustering, producing clearer reachability diagrams and achieving higher clustering accuracy. However, OPTICS-PLUS still has limitations in effectively characterizing rainfall, requires users to input density thresholds (introducing subjective influence on prediction results), suffers from high time complexity, and is unsuitable for large-scale landslide datasets. To address these issues, this paper improves upon OPTICS-PLUS by optimizing the algorithm' s expansion strategy and proposing a nearest neighbor search-based OPTICS algorithm (NNSB-OPTICS). Then, based on rainfall data distribution characteristics, an EC-type distance formula is proposed by combining the EW-type distance formula [10] and cloud model theory [11]. Integrating the EC-type distance formula into NNSB-OPTICS yields the uncertain NNSB-OPTICS algorithm, solving the problem of effectively characterizing uncertain rainfall data in landslide prediction. Finally, the uncertain NNSB-OPTICS algorithm is applied

to landslide hazard prediction in Baota District, Yan' an City, establishing a landslide hazard prediction model that demonstrates the algorithm' s feasibility and effectiveness.

1 Uncertain NNSB-OPTICS Clustering Algorithm

1.1 NNSB-OPTICS Clustering Algorithm Design

OPTICS-PLUS is an improvement over the OPTICS algorithm [12] that solves the problem of sparse points not being effectively clustered due to greedy search strategies, achieving higher clustering accuracy. However, OPTICS-PLUS still requires users to input density thresholds, making it difficult to avoid subjective and arbitrary influences on landslide prediction results, and suffers from high time complexity unsuitable for large-scale landslide datasets. Therefore, this paper proposes the NNSB-OPTICS clustering algorithm based on OPTICS-PLUS improvements. The algorithm first designs a global nearest neighbor pointer that always points to the point with the smallest nearest neighbor distance in the seed queue. After completing one expansion, the algorithm extracts the point pointed to by this pointer for the next iteration, eliminating the need for sorting operations and effectively improving time efficiency. Second, it introduces the concept of point average distance, obtaining each data object' s point average distance through iterative expansion to form a point average distance ordering that contains clustering structure information. Based on this ordering, the dataset can be divided into several clusters without requiring users to set thresholds, reducing human influence on landslide prediction results.

For a given dataset, the following definitions are provided:

Definition 1 (Nearest Neighbor Distance). Let there be two sets and , where and . The nearest neighbor distance of is:

Definition 2 (Point Average Distance). Let , the point average density of is:

where is the distance between and in .

The NNSB-OPTICS clustering algorithm generates a point average distance ordering queue through nearest distance iterative expansion. By analyzing steep rising and falling regions in this queue, dense and sparse regions can be effectively distinguished, dividing the dataset into several clusters without manual density threshold setting. During iterative expansion, to avoid multiple sorting operations and repeated similarity calculations that reduce time efficiency, NNSB-OPTICS makes two improvements on OPTICS-PLUS: (a) It defines a GPNP (global point to nearest point) pointer, as shown in [Figure 1: see original paper], which records the point with the smallest nearest neighbor distance in each iteration and points to it, allowing direct extraction of the GPNP-pointed point for expansion in the next iteration; (b) NNSB-OPTICS adds an SD (Sum of Distance) field for each object to record the total distance between that object and already-expanded objects, enabling direct reading of the total distance

from the SD field when calculating a point' s average distance.

NNSB-OPTICS optimizes the OPTICS algorithm' s expansion strategy. During iterative expansion, it does not compare already-expanded data objects, reducing one comparison object per iteration. Therefore, the algorithm' s time complexity is , equal to OPTICS' s time complexity [13]. However, when determining expansion direction, OPTICS requires sorting the ordered seed queue with time complexity (where is seed queue length) [14], while NNSB-OPTICS only needs to search according to the GPNP pointer and extract the pointed object for expansion with time complexity . Thus, NNSB-OPTICS achieves higher actual efficiency than OPTICS.

1.2 Uncertain Data Processing

Landslide occurrence is influenced by multiple factors, among which rainfall is an important triggering factor. However, rainfall values constitute uncertain data where only approximate ranges can be determined, and precise numerical values cannot be described [14]. The proposed NNSB-OPTICS clustering algorithm applies to continuous and discrete attribute data but cannot effectively characterize and process uncertain rainfall data. Therefore, this paper introduces the EW-type distance formula and cloud model theory, combining them based on rainfall data distribution characteristics to obtain a new uncertain data distance formula.

Definition 3 (Uncertain Data) [15]. Let there exist a mapping such that , then is called uncertain data, where is the value range of , and is the probability density function of .

In uncertain data distance measurement, the EW-type distance is the most widely used method. For given uncertain data and , the EW-type distance between and is:

where and are the expectations of and , and and are the interval widths of and .

The EW-type distance assumes uncertain data follows a uniform distribution within the value interval, comprehensively measuring distance through expectation and interval width. However, the uncertain rainfall data studied in this paper approximately follows a normal distribution within its value interval [16], making the EW-type distance unsuitable for rainfall characterization. According to Yan' an Meteorological Bureau data, adjacent regions exhibit similar rainfall amounts. Based on this property, the backward cloud algorithm [17] can be used to obtain the normal cloud model digital characteristics corresponding to rainfall , providing qualitative description. Extract neighboring uncertain data value intervals as -level cuts of normal fuzzy numbers for fuzzification to obtain fuzzy membership function curves. According to the expansion principle, expand the fuzzy membership function curves into cloud expectation curves. From these curves, cloud model digital characteristics are obtained, where ex-

pectation and hyper-entropy are calculated as:

where is the mean of the cloud expectation curve equation, is the variance of the cloud expectation curve, and is the mean of . The cloud model expectation and hyper-entropy are introduced into the EW-type distance, using hyper-entropy instead of interval width to describe the fuzziness of uncertain data, yielding the EC-type distance formula:

For discrete and continuous attributes whose values are not fuzzy, directly set their expectation equal to the attribute value and hyper-entropy equal to 0, then use the EC-type distance formula for distance measurement. Thus, the EC-type distance formula applies to datasets containing discrete, continuous, and uncertain attributes.

1.3 Uncertain NNSB-OPTICS Clustering Algorithm Design

Using the EC-type distance formula as the similarity calculation metric applied to the NNSB-OPTICS clustering algorithm yields the uncertain NNSB-OPTICS clustering algorithm. The algorithm flow is as follows:

Algorithm: Uncertain NNSB-OPTICS
Input: Dataset
Output: Result queue

1. Calculate the expectation and hyper-entropy of each attribute in dataset

2. Initialize result queue as empty

3. While dataset is not empty:
   3.1 Extract point from dataset via GPNP pointer
   3.2 If dataset is empty, algorithm ends
   3.3 For each point in dataset:
   3.3.1 Calculate distance using EC-type distance formula and update
   3.3.2 Update
   3.4 Update and add to result queue

4. Return result queue

2 Experiments and Analysis

2.1 Experimental Environment

All experiments were conducted on a Windows 7 operating system with an Intel(R) Core(TM) i5-4210U 2.80GHz CPU and 8GB RAM. Landslide experimental data were extracted using ARCGIS 10.3 software, with Oracle 12c as the database platform. Algorithms were tested in Python language on the PyCharm 5.03 platform.

2.2 Simulation Experiments

To verify the clustering effectiveness of the proposed algorithm, NNSB-OPTICS was compared with OPTICS [12], OPTICS-PLUS [9], and EOPTICS [18] on four UCI datasets. Dataset characteristics are shown in . Experiments focused on clustering accuracy, result stability, and time efficiency.

This paper adopts the Micro-precision standard to measure clustering accuracy using data classification information, calculated as:

where represents the number of correctly clustered sample points, is the total number of dataset samples, and is the number of clusters. The value ranges in [0,1], with values closer to 1 indicating higher clustering accuracy. Before experiments, multiple tests were conducted to obtain the optimal neighborhood radius for OPTICS-PLUS on each dataset. The experimental neighborhood radius value set was defined as , and the MinPts value set was . Combinations of neighborhood radius and core points yielded 9 parameter groups, with each group running 10 times for a total of 90 experiments. To better analyze algorithm accuracy and stability, the best result and worst result were recorded, and the mean of multiple experiments was calculated using:

where is the number of experimental repetitions (in this study). Clustering accuracy and single-run times for each algorithm on UCI datasets are shown in and .

The experimental results on UCI datasets indicate that NNSB-OPTICS achieves higher average accuracy than the other three algorithms on all four datasets. The difference between best and worst results is significantly smaller for NNSB-OPTICS, demonstrating higher clustering accuracy and better result stability. This is attributed to two factors: first, NNSB-OPTICS avoids manual density threshold setting, reducing human influence on clustering results; second, after expansion completion, NNSB-OPTICS employs a result reorganization strategy to assign boundary points to the nearest dense region, improving clustering accuracy. However, optimal results show that on the Iris and Wine datasets, NNSB-OPTICS performs slightly worse than OPTICS-PLUS because Iris class 1 and class 2 samples, as well as Wine class 3 samples, exhibit data overlap with minimal density distribution changes in overlapping regions, causing the point average distance ordering produced by NNSB-OPTICS to fluctuate smoothly and resulting in poorer cluster identification—an aspect requiring further improvement. In terms of time efficiency (), NNSB-OPTICS demonstrates the shortest single-run time among all algorithms, showing clear advantages in time efficiency.

2.3 Case Application

To verify the feasibility of the uncertain NNSB-OPTICS algorithm in landslide hazard prediction and whether the proposed uncertain data processing method can effectively characterize rainfall, Baota District in Yan' an City was selected

as the study area for validation. Baota District is located in the central part of the Loess Plateau in northern Shaanxi, with complex geological conditions, frequent human activities, and high landslide frequency due to rainfall and other influencing factors, posing significant threats to human safety and property. Rainfall uncertainty is difficult to effectively characterize in actual landslide hazard prediction. Based on the proposed uncertain data processing method, rainfall was processed and combined with landslide-related theoretical foundations to apply the uncertain NNSB-OPTICS clustering algorithm to landslide hazard prediction research in Baota District, verifying its feasibility.

2.3.1 Data Sources and Preprocessing This study conducted landslide hazard prediction research based on the Baota District geological disaster detailed survey project. Experimental data sources were as follows: Using ARCGIS software, Baota District was divided into grid cells of 5m×5m size, resulting in 5,672,922 grid units. Each grid unit was treated as a point and imported into a 1:5000 digital elevation map to derive slope type, gradient, height, and aspect thematic maps, from which corresponding data were extracted. Rock-soil structure data were obtained from 1:10000 geological maps. Vegetation coverage values were derived from Spot5 remote sensing data using ERDAS remote sensing image processing software to calculate normalized difference values from Spot5 near-infrared band B3 and visible red band B2. Rainfall values included 24-hour rainfall for 7 days before landslide occurrence and future 7-day 24-hour rainfall from meteorological rainfall maps.

The original dataset contained millions of samples with numerous attributes, including many missing, duplicate, and erroneous values. To improve experimental accuracy, data preprocessing was required. First, dimensionality reduction was performed based on landslide theory and the special geological environment characteristics of the Loess Plateau, retaining seven attribute items as clustering features: slope gradient, height, aspect, vegetation, slope type, rock-soil structure, and rainfall, with landslide hazard level as the decision attribute, deleting other less influential attributes. Then data cleaning removed records containing missing, duplicate, or erroneous values. After preprocessing, 5,647,382 valid records were obtained, with attribute characteristics shown in .

2.3.2 Uncertain NNSB-OPTICS Clustering Model Construction First, cloud model mean and hyper-entropy for each object in the dataset were calculated using equations (1) and (2). The result queue was initialized as empty. Starting from any object in the dataset, distances between objects were calculated using the proposed EC-type distance formula. After each iteration, expansion proceeded according to the GPNP pointer direction until all objects in the dataset were added to the result queue, obtaining a point average distance ordering queue containing clustering structure information. Finally, the Gmdlent Clustering method [19] was used to identify steep rising and falling regions in the point average distance ordering for cluster recognition, outputting clustering results to obtain clusters.

After clustering, all evaluation units in the landslide sample dataset were divided into clusters. According to clustering algorithm properties, objects within the same cluster have high similarity, while objects between different clusters have high dissimilarity—meaning evaluation units within the same cluster have similar topographic, geomorphic, and climatic environmental characteristics. Reference [20] demonstrates that similar landslide development characteristics imply similar landslide occurrence trends. Based on this theory, using 293 landslide observation points with rainfall information obtained from field surveys, combined with direct search and expert evaluation methods [21], hazard levels for each cluster can be quickly determined. Through direct search, each cluster is examined: if it contains one confirmed hazard level, the cluster' s hazard level equals that level; if it contains two or more confirmed hazard levels with different counts, the majority principle determines the cluster's hazard level; if counts are equal or no confirmed hazard levels exist, landslide disaster experts determine the hazard level based on prior experience and familiarity with regional geological environmental conditions combined with survey results, thereby classifying the hazard levels of remaining evaluation units in the study area.

2.3.3 Evaluation Criteria An error matrix was established through statistical analysis of actual survey data and prediction results. In the error matrix, columns represent actual observed values while rows represent predicted values from the model. For example, represents the number of samples with low-risk predictions and low-risk observations, represents samples with low-risk predictions but medium-risk observations, and represents samples with low-risk predictions but high-risk observations.

The Kappa coefficient [22] is a relatively simple yet highly accurate evaluation method. The Kappa coefficient based on the error matrix can statistically reflect the superiority of classification results. The calculation formula is:

where is the overall accuracy of the prediction model, representing the probability that predicted values match observed values in the dataset; is the total of row records; is the total of column records; is the total sample size; and is the number of classification types. In this experiment, . The Kappa coefficient ranges in [0,1], with higher values indicating better prediction performance. When predicted values completely match observed values for all samples, the Kappa coefficient equals 1.

2.3.4 Landslide Prediction Accuracy Analysis and Comparison To verify whether the proposed uncertain data processing method can effectively handle rainfall data and improve landslide prediction accuracy, both NNSB-OPTICS and uncertain NNSB-OPTICS clustering algorithms were used to establish landslide prediction models for Baota District, Yan' an City. For uncertain rainfall data, the uncertain NNSB-OPTICS landslide prediction model uses equations (1) and (2) to obtain rainfall cloud model digital characteristics, then uses equation (3) for similarity calculation. The NNSB-OPTICS cluster-

ing model uses the quantitative method [23] traditionally employed in landslide prediction for rainfall discretization, dividing rainfall into: light rain (below 20 mm), moderate rain (20-44.9 mm), heavy rain (45-59.9 mm), storm (60-79.9 mm), heavy storm (80-99.9 mm), and extreme storm (above 100 mm), with numerical values substituted and Euclidean distance used for similarity calculation.

Baota District has 428 geological disaster observation points, including 293 landslide observation points. During data preprocessing, all landslide observation points were rasterized into 1,367 evaluation units, comprising 311 low-risk, 729 medium-risk, and 327 high-risk units. Statistical analysis of prediction results from both models yielded error matrices shown in .

** Error Matrix of Landslide Hazard Prediction for Two Models**

Using and equations (4) and (5), the overall accuracy and Kappa coefficient for the NNSB-OPTICS landslide prediction model were calculated as:

The overall accuracy and Kappa coefficient for the uncertain NNSB-OPTICS model were:

Results show both landslide prediction models achieve overall accuracy above 80%, meeting credibility requirements [24], demonstrating that both NNSB-OPTICS and uncertain NNSB-OPTICS clustering algorithms are feasible for landslide prediction. Comparing the two models, the uncertain NNSB-OPTICS model' s overall accuracy is 4.3 percentage points higher than NNSB-OPTICS, with a higher Kappa coefficient, proving that under the same landslide sample dataset, the uncertain NNSB-OPTICS model' s predictions better match actual landslide conditions. This is because the uncertain NNSB-OPTICS model employs the EC-type distance formula for rainfall characterization, fully considering rainfall distribution characteristics and compensating for information loss inherent in traditional landslide prediction methods that directly discretize rainfall, thereby improving landslide hazard prediction accuracy to some extent.

3 Conclusion

Traditional clustering algorithms cannot effectively characterize uncertain rainfall data and exhibit poor clustering performance on non-uniformly distributed data in landslide hazard prediction. To address this, this paper first introduced the density-based OPTICS-PLUS algorithm and proposed the NNSB-OPTICS clustering algorithm to overcome its limitations of requiring manual density threshold setting and high time complexity. Then, considering rainfall data distribution characteristics, the EC-type distance formula was proposed by combining the EW-type distance formula and cloud model theory. Integrating the EC-type distance formula into NNSB-OPTICS yielded the uncertain NNSB-OPTICS clustering algorithm, solving the problem of effectively characterizing and processing rainfall data. Case study validation in landslide hazard prediction demonstrated that the proposed uncertain data processing method can

more effectively characterize rainfall and improve landslide prediction accuracy.

References

[1] Huang Runqiu. Large-scale landslides and their occurrence mechanisms in China since the 20th century [J]. Chinese Journal of Rock Mechanics and Engineering, 2007, 26(3): 434-454.

[2] Zhou Tao, Lu Huiling. Research progress on clustering algorithms in data mining [J]. Computer Engineering and Applications, 2012, 48(12): 100-111.

[3] Gorsevski P V, Gessler P E, Jankowski P. A fuzzy K-means classification and a bayesian approach for spatial prediction of landslide hazard [M]// Handbook of Applied Spatial Analysis. Berlin: Springer, 2010: 653-684.

[4] Hu Kaiheng, Cui Peng, Han Yongshun, et al. Debris flow landslide susceptibility evaluation in Wenchuan disaster area based on clustering and maximum likelihood method [J]. Science of Soil and Water Conservation, 2012, 10(1): 12-18.

[5] Gui Lei, Yin Kunlong, Wang Jiajia. Research on landslide hazard zoning based on cluster analysis [J]. Hydrogeology & Engineering Geology, 2013, 40(1): 100-105.

[6] Zhang Jun, Yin Kunlong, Wang Jiajia, et al. Study on landslide susceptibility evaluation in Wanzhou District of Three Gorges Reservoir [J]. Chinese Journal of Rock Mechanics and Engineering, 2016, 35(2): 284-296.

[7] Wang X, Niu R. Landslide prediction in three gorges based on cloud model and data field [C]// Proc of International Symposium on Computer Network and Multimedia Technology. 2009: 1-4.

[8] Wan S. Entropy-based particle swarm optimization with clustering analysis on landslide susceptibility mapping [J]. Environmental Earth Sciences, 2013, 68(5): 1349-1366.

[9] Zeng Yiling, Xu Hongbo, Bai Shuo. Improved OPTICS algorithm and its application in text clustering [J]. Journal of Chinese Information Processing, 2008, 22(1): 51-55.

[10] Bao Yu' e, Peng Xiaoqin, Zhao Bo. Interval number distance based on expectation and width and its completeness [J]. Fuzzy Systems and Mathematics, 2013, 27(6): 133-139.

[11] Li Deyi, Liu Changyu. On the universality of normal cloud model [J]. Engineering Science, 2004, 6(8): 28-34.

[12] Ankerst M, Breunig M M, Kriegel H P, et al. OPTICS: ordering points to identify the clustering structure [J]. ACM SIGMOD Record, 1999, 28(2): 49-60.

[13] Zhang Q, Wang X, Wang X. An OPTICS clustering-based anomalous data filtering algorithm for condition monitoring of power equipment [M]// Data

Analytics for Renewable Energy Integration. [S.l.]: Springer International Publishing, 2015: 917-922.

[14] Hou Rongtao, Lu Yu, Wang Qin, et al. Application of OPTICS algorithm in lightning nowcasting [J]. Journal of Computer Applications, 2014, 34(1): 297-301.

[15] Mao Yimin, Peng Zhe, Chen Zhigang, et al. Application of uncertain decision tree classification algorithm in landslide hazard prediction [J]. Application Research of Computers, 2014, 31(12): 3646-3650.

[16] Liu Weiming, Gao Xiaodong, Mao Yimin, et al. Research and application of uncertain genetic neural network in landslide hazard prediction [J]. Computer Engineering, 2017, 43(2): 308-316.

[17] Yu Shaowei, Shi Zhongke. A new backward cloud algorithm based on normal distribution interval numbers [J]. Systems Engineering—Theory & Practice, 2011, 31(10): 2021-2026.

[18] Alzaalan M E, Aldahdooh R T, Ashour W. EOPTICS: enhancement ordering points to identify the clustering structure [J]. International Journal of Computer Applications, 2012, 40(17): 975-8887.

[19] Brecheisen S, Kriegel H P, Kroger P, et al. Density-based data analysis and similarity search [M]// Multimedia Data Mining and Knowledge Discovery. Berlin: Springer, 2007: 94-115.

[20] Kwang Y, Jong H. Landslide susceptibility mapping in Injae, Korea, using a decision tree [J]. Engineering Geology, 2010, 116(3): 274-283.

[21] Guzzetti F, Carrara A, Cardinali M, Reichenbach P. Landslide hazard evaluation: a review of current techniques and their application in a multi-scale study, Central Italy [J]. Geomorphology, 1999, 31(1-4): 181-216.

[22] Xu Wenning, Wang Pengxin, Han Ping, et al. Application of Kappa coefficient in accuracy evaluation of drought prediction models—taking Guanzhong Plain drought prediction as an example [J]. Journal of Natural Disasters, 2011, 20(6): 45-53.

[23] Gao Huaxi, Yin Kunlong. Correlation analysis between rainfall and landslide disasters and discussion on early warning threshold [J]. Rock and Soil Mechanics, 2007, 28(5): 1056-1060.

[24] Sabokbar H F, Roodposhti M S, Tazik E. Landslide susceptibility mapping using geographically-weighted principal component analysis [J]. Geomorphology, 2014, 226(1): 15-24.

Note: Figure translations are in progress. See original paper for figures.