THE STUDY OF KPSO FOR GENERATING LANDSLIDE THEMATIC MAP

Shiuan Wan[1], Tsu chiang Lei[2], Zhuo Yi Lin[3]

[1]Associate Professor of Department of Information Management, Ling Tung University, Taichung, Taiwan, Tel: +886-4-24629863

Email:shiuan123@mail.ltu.edu.tw

[2]Associate Professor of Department of Urban Planning and Spatial Information, Feng Chia University, Taichung, Taiwan, Tel:+886-4-24511248

E-mail:tclei@fcu.edu.tw

[3]Graduate Student of Civil Engineering of NCHU, Taichung, Taiwan

Email: k74111en@yahoo.com.tw

KEY WORDS: Landslide, spatial information, clustering techniques

ABSTRACT: Spatial information of landslide interpretation is usually analyzed statistically based on topography, vegetation conditions and landforms. However, the uncertainty and the massive volume of data, the determination of landslide occurrence will result in many misjudgments. Hence, this study is decided to develop the clustering techniques to extract out some knowledge from landslide occurrence. More specifically, this study used spatial information technology (RS and DEM) to attain the vegetation cover and the landforms. Ancillary information is also adopted to attain the vegetation information. Similar results are obtained through KPSO and verifications are made.

1. INTRODUCTION

Particle swarm optimization (PSO) is a computational method that optimizes a problem by iteratively trying to improve a candidate solution with regard to a given measure of quality. PSO optimizes a problem by having a population of candidate solutions, here labeled particles, and moving these particles around in the search-space based on the particle's position and velocity, mathematically. The Swarm intelligence (SI) is the collective behavior of artificial intelligence. The connection of traditional k-means (Wan S., 2009) and particle swarm intelligence can be built as an evolution model for data analysis.

In this study, it is decide to develop a decision support system of landslide susceptibility map with spatial clustering analysis at the region of Wushe reservoir by remote sensing data. The entire study can be divided into three parts: The first part introduces the study area. The field backgrounds are generally introduced. The second part proposes the research method on the various cases of fuzzy c-means and discusses the difference between two different model's approaches. The third part presents the results and conclusions are drawn.

2. STUDY AREA AND MATERIAL

The study area is located at the downstream of Wushe reservoir, Nantou county, Taiwan. Figure 1 shows the information on study area. Figure 1 shows the site location and remote sensing image on the reservoir. This landslide area is 4.3 ha with an average slope of 30 degree. Typhoon generally brings heavy precipitations to trigger landslides in this area. The movement of landslide may cut off Wushe stream, and then destroy the Wushe hydraulic power plant. Consequently, the Taiwan Power Company (TPC) made many protection constructions (such as rock anchor) and monitoring equipments (such as the monitoring wells of ground-water levels) in the landslide. Figure 1 shows the picture of landslide around the reservoir.

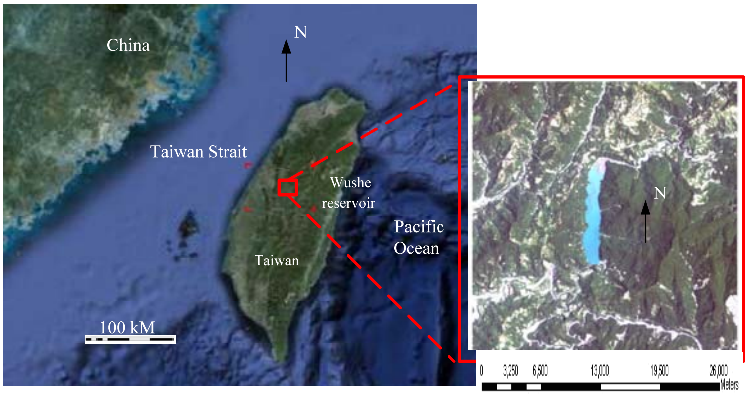

Figure 1 Information on Study Area

2.1 Investigated Scenario and Background

In general, the analyses of landslide occurrences are usually relied on inventory map. Unfortunately, this area is short of historical data for landslide, thus, the inventory map is not exist. In this research, the study area is the Wushe reservoir and the data collected belongs to the period after Chi-Chi earthquake(1999). The Wushe reservoir (area: 125000 ha.) is situated 20-40 km from the Chelung Pu fault, which transmitted tremendous energy to the central part of Taiwan. Utilizing DEMs, SPOT-image data, the researchers investigated maps and GIS spatial attribute data (morphology, geology, landslide, slope, and soil type etc.). Those data were preprocessed using the spatial information system to build a database of landslide events. During 2006/3/1 to 2008/10/1, two violent typhoons invaded the landslide area and induced a great amount of landslide. The cumulative precipitation depth of the Sinlaku typhoon is 976mm and the Jangmi typhoon is 451mm. The strong earthquake and the heavy rainfall are the trigger to the occurrence of the landslide area.

2.2 Ancillary Information

In general, the vegetation indicators are employed as a tool to model the spatially varying phenomenon of natural (Pradhan and Lee 2010). We incorporate eight indicators as (1) Normalized Difference Vegetation Index ; *NDVI* (2) Cropping management factor index; *CMFI* (3) Band-Ratio; *BR* (4) SquareBand-Ratio; *SQBR* (5) Vegetation Index; *VI* (6) Soil Adjusted Vegetation Index; *SAVI* (7) Modified Soil Adjusted Vegetation Index; *MSAVI* (8) Average Brightness Index; *ABI*

There are 4 basic spectrums: B, G, R, and IR. Based on these fundamental spectrums, 8 different indicators of vegetations are generated. The 30 study samples are listed in Figure 2. Then, twelve different image data are created as one single image by image fusion. We performed a Data Mining approach to dig out the landslide zone from those image data. More specifically, the Fuzzy C-means and KPSO are used as a parallel classifier to identify the landslide area, respectively.

3. RESEARCH METHOD

KPSO-clustering algorithm is to detect the cluster centers of a given data set as an optimal case or not (Chen and Ye, 2003). It combines the algorithms of K-means and Particle Swamp Optimization. An initial population of solutions called particles is randomly generated, and each string is a sequence of real numbers representing the *K* cluster centers. After the encoding of the string of the particles, the execution of KPSO is as follow:

Step 1) Initialize position vector *X* and associated velocity *V* of all particles in the population randomly.

$$V_{id} = V_{id} + C_1 * rand() * (P_{id} - X_{id}) + C_2 * rand() * (P_{gd} - X_{id}) \quad (1)$$

$$X_{id} = X_{id} + V_{id} \quad (2)$$

Step 2) Evaluate the fitness function for each particle. We use metric function which proposes by [2] to measure the similarity dissimilarity between the various elements of a data set. For each data point x_i, we assign point x_i, $i \in \{1,2, \ldots, N\}$ to cluster C_j, $j \in \{1,2, \ldots, K\}$ if

$$1\text{-}\exp(-\beta\|x_i - z_j\|^2) = \min_p \{1\text{-}\exp(-\beta\|x_i - z_p\|^2)\},\ p=1,2,\ldots,K \quad (3)$$

where $$\beta = \left(\frac{\sum_{i=1}^{N}\|x_i - \bar{x}\|^2}{N}\right)^{-1},\quad \bar{x} = \frac{\sum_{i=1}^{N} x_i}{N} \quad (4)$$

In our research, the selected fitness function of the AKPSO is given by:

$$\text{fitness} = \frac{k_0}{J_0 + \sum_{j=1}^{K}\sum_{x_i \in C_j}\left\{1 - \exp(-\beta\|x_i - z_j\|^2)\right\}} \quad (5)$$

where the k_o is a positive constant, and J_o is a small valued constant.

Step 3) Compare every particle's fitness value with previous particle's best solution (*pbest*). If current solution is better than previous value (*pbest*), then update *pbest* with current solution.

Step 4) Compare fitness evaluation with the population's overall previous best. If current value is better than the *gbest* (the global version of the best value), then reset *gbest* to the current particle's value and position.

Step 5) Use the one step of K-means algorithm to replace the result of the *gbest*. The cluster centers encoded in the *gbest* are replaced by the mean points of the respective clusters [11]:

$$z_j^* = \frac{1}{N_j} \sum_{x_i \in C_j} x_i \text{ , } j=1,2,\ldots,K \qquad (9)$$

where N_j is the number of points belonging to cluster C_j. The effect of the K-means algorithm is to direct the best solution towards the area of the training data. The drawback of the hybridization is that the running time considerably grows as the number of K-means step increases. For better convergence and lower computing time purpose, the Step 5 work in the initial five iterations (or less) is enough.

Step 6) Change the velocities and positions with Eq. (11) and Eq. (12).

Step 7) Repeat Step 2 to Step 6 until the predefined number of iterations is completed.

4. DISCUSSION AND RESULTS

There are 2600 observed samples which include 1427 non-occurrences and 1173 occurrences. As aforementioned, only 10 occurrence and 20 non-occurrence samples are selected to generate the image-based landslide rules.

Figure 2 presents the outcomes of K-PSO model. The KPSO can search the optimal cluster center among the raw-band and 8 ancillary information. It can be observed that the KPSO can approach approximate 85% of accuracy. Comparing Figure 2 and Figure 2, the major parts of the landslide are successfully detected. This is the optimal cluster-center for clustering techniques. The classification accuracy is shown. It has to notice that the performance of iterative clustering algorithms which converges to numerous local minima depends highly on initial cluster centers. KPSO can resolve the local minima problem and then produce the best cluster outcomes. Moreover, the results from KPSO can be taken as a consideration on the validation model of Fuzzy-C means. Hence, the results of both models are very similar.

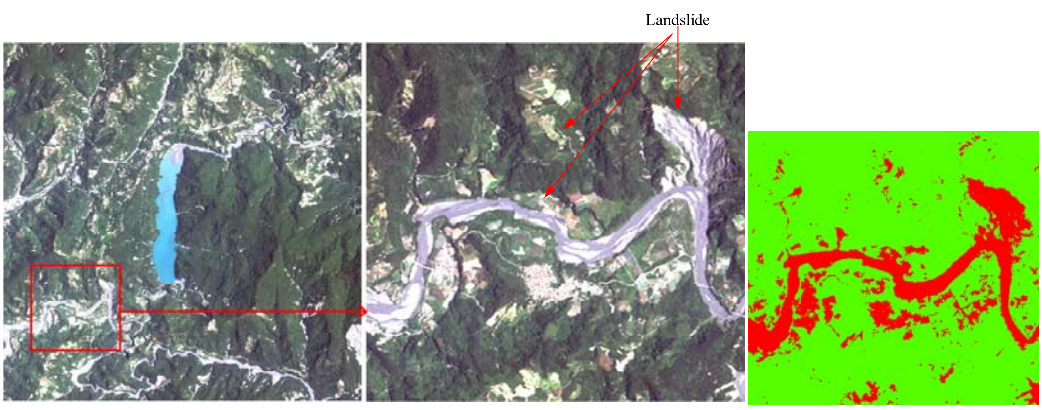

Figure 2 Image classification for landslide area

5. CONCLUSION

Monitoring and assessment of landslide hazard is an important task for decision making and policy planning in the landslide area. This study proposed two useful classifiers coupled with 4 original bands+ 8 ancillary information to derive landslide spectral feature. The accuracy of two classifiers is about 85%. Fuzzy Clustering can be considered the most important unsupervised learning problem; so, as every other problem of this kind, it deals with finding a structure in a collection of unlabeled data. Accordingly, the unlabeled data of can save a huge amount of time on investigating the in-situ landslide occurrence.

References

Chen, C.-Y. and Ye, F., K-means Algorithm Based on Particle Swarm Optimization, 2003 International Conference on Informatics, Cybernetics, and Systems, I-Shou University, Taiwan, R.O.C. pp. 1470-1475.

Lin W.T. 2008, Earthquake-induced landslide hazard monitoring and assessment using SOM and PROMETHEE techniques: A case study at the Chiufenershan area in Central Taiwan, International Journal of Geographical Information Science, 22:9,995 -1012.

Pradhan B., Lee S. 2010, Landslide susceptibility assessment and factor effect analysis: back propagation artificial neural networks and their comparison with frequency ratio and bivariate logistic regression modelling, Environmental Modelling & Software,25(6), 747-759.

Wan S., 2009, A spatial decision support system for extracting the core factors and thresholds for landslide susceptibility map, Engineering Geology 108, 237–251.