APPLICATION OF SWARM INTELLIGENCE FOR LANDSLIDE SUSCEPTIBILITY MODELING FROM GEOSPATIAL DATA FUSION

Jhe-Syuan Lai[1] and Fuan Tsai[2]
[1]Department of Civil Engineering, National Central University, Taiwan,
Email: jslai0726@gmail.com
[2]Center for Space and Remote Sensing Research, National Central University, Taiwan,
Email: ftsai@csrsr.ncu.edu.tw

KEY WORDS: Landslide Susceptibility, Ant Colony Optimization, Particle Swarm Optimization.

ABSTRACT: This study utilizes and explores the feasibility and application of swarm intelligence for landslide susceptibility modeling based on collected inventory of rainfall-induced shallow landslide events. Eleven geospatial factors are considered, including topographic, vegetative, environmental, geological and man-made information. Landslide inventory and factors are overlapped to obtain the training data for modeling (classification) and verification. Experimental results indicate that swarm intelligence algorithms can provide plausible results for landslide susceptibility modeling, comparing with conventional landslide detection and prediction methods.

1. INTRODUCTION

Landslide is one of the natural hazards causing human casualties, property damages and civil problems. Rainfall-induced shallow landslides (the landslide depth is smaller than 2 meter) are frequently triggered by typhoons from June to September every year in Taiwan. Therefore, modeling and predicting this type of landslides in order to prevent and mitigate disasters has become an important issue. Geospatial technologies and data can be used effectively to monitor and investigate long-term landslides in a large area (e.g. Peduzzi, 2010; Tsai & Chen, 2007). The focus of this study is to evaluate and model landslide susceptibility based on different typhoon events. Landslide susceptibility indicates the potential of landslide occurrence that ignores the triggered factor (rainfall in this study) to calculate the landslide likelihood instead of probability (Dai et al., 2002). A number of related works have proposed or applied different algorithms in order to construct robust landslide models using geospatial data fusion, e.g. statistical approach (Chang et al., 2014), data mining(Tsai et al., 2013), and artificial intelligence (Oh & Pradhan, 2011).

On the other hand, swarm intelligence is a novel concept to learn the collective behavior of biology for numerical analysis and computational optimization. It is primarily employed in the artificial intelligence domain. In particular, Ant Colony Optimization (ACO) and Particle Swarm Optimization (PSO) are famous and feasible ways of swarm intelligence. The strategies of ACO and PSO are that the former considers the interaction of ants using pheromone and the latter changes the velocity and position of particles. The similarity of both approaches is to perform iteration process to find the global or best solution. However, few studies have adopted ACO and PSO algorithms to model landslide events. This study utilizes and explores the feasibility and application of swarm intelligence for landslide susceptibility modeling based on four rainfall-induced shallow landslide events. Eleven geospatial factors are considered, including topographic, vegetative, environmental, geological and man-made information. Landslide inventory and factors are overlapped to obtain the training data for modeling (classification) and verification.

2. STUDY SITE AND DATA

The Shimen reservoir watershed is selected as the study site that covers a region of about 763.4 square kilometers in northern Taiwan. The elevation in the study site ranges between 250m to 3,500m measured from DEM (Digital Elevation Model). Forest is the primary land-cover, but there are few agricultural activities. Rainfall-induced shallow landslides commonly occurred because of typhoons in this region, causing various problems in water supply and resource management.

Eleven factors are considered in this study as listed in Table 1, including three original and eight derived data sets. The NDVI (Normalized Difference Vegetation Index) is an only multi-temporal factor, selected before typhoon events. In addition, this study normalized all NDVI images using PIFs (Pseudo Invariant Features) to reduce the different radiometric and atmospheric conditions. It is very convenient and suitable to analyze multi-temporal NDVIs (Du et al., 2002).

For identifying landslide samples, the landslide extents of four typhoons were digitized based on change detection results that were checked against auxiliary ground truth data and field investigations to generate a landslide inventory (Tsai and Chen, 2007). Consequently, this study transformed all landslide extents into pixels (10 x 10 m) in order to overlay with other raster data. However, these landslide pixels probably contain deposition area (non-landslide), thus may affect the fidelity of the model and cause inaccurate results. Therefore, this study removes landslide pixels with an empirical criterion to overcome this problem. In this study, landslide pixels identified on satellite images but whose slope is less than 10 degrees are considered as deposition instead of landslides. Table 2 displays four typhoon events and the numbers of landslide pixels occurred after each typhoon.

Table 1. Landslide factors

Original data	Derived data	Resolution/scale
DEM	Elevation	40 x 40 m
	Slope	
	Aspect	
	Curvature	
SPOT Images	NDVI	10 x 10 m
River	Distance to river	1/5,000
Road	Distance to road	1/5,000
Fault	Distance to fault	1/50,000
Land use	-	1/5,000
Soil	-	1/25,000
Geology	-	1/50,000

Table 2. Typhoon events and numbers of landslide pixels

Typhoon event	Date	# of landslide pixels
Matsa	2005/8	1,480
Korsa	2007/10	296
Fung-wong	2008/7	296
Sinlaku	2008/9	654

3. PROCEDURE AND METHODS

There are four primary steps in this study, including (1) data pre-processing and integration, (2) data arrangement, (3) landslide susceptibility modeling and (4) accuracy assessment. In the data pre-processing and integration step, because the utilized analysis algorithms are record- (grid- or cell-) based, vector data need to be rasterized. In addition, all data were resampled to the same cell size (10 x 10 m) and subjected to PIFs normalization for the NDVI images. Consequently, some factors that can provide advanced information were derived from original data. Finally the pre-processed data were integrated for the subsequent analyses.

For the data arrangement, each typhoon-based non-landslide (non-occurrence) samples (pixels) were randomly selected and the number of pixels is the same as landslide. The attributes of integrated data were extracted according to different event-based landslide and non-landslide samples for inputting the classifiers. Both ACO and PSO classifiers were used to build the landslide susceptibility models. To assess the models, this study utilized K-fold cross validation to calculate Overall Accuracy (OA) derived from confusion matrix. In order to avoid insignificant results because of few samples in each fold (i.e. typhoon Korsa and Fung-wong), the 5-fold was chosen. This study also utilized the standard deviation of OAs to assess the models.

3.1 Ant Colony Optimization for Classification Task

The concept of ACO is to simulate the behavior of ants based on pheromone. Each Ant will select a path that contains strong pheromone. However, the strength of pheromone will decrease with the passing of time. According to previous phenomenon, the concentration area selected by ants is a short or suitable path. Many literatures have proved that ACO algorithm can be useful for choosing the optimal path. On the other hand, ant-miner (Parpinelli et al., 2002) was proposed to deal with classification task. The criteria of ant-miner is to construct many “If-Then” rules consisting of subsets of nominal factors or variables (or “terms”) using sequential covering approach (Witten & Frank, 2005). In other words, the classification rules are searched by ants in the terms of nominal (or discrete) factors. Each iteration of ant-miner calculation will choose some terms to construct a rule based on Eq. (1) in which a high value for a specific term means high probability to be selected. In addition, each iteration will update pheromone based on Eq. (4). Eq. (2), (3) and (5) are parts of Eq. (1) and (4), respectively. However, ant-miner cannot deal with numeric (or continuous) data. This study adopted an effective and classical way called multi-interval discretization (Fayyad & Irani, 1993) to discretize numeric factors.

There are five parameters which should be identified in the ant-miner platform, i.e. number of (1) ants and (2) iterations, (3) minimum numbers of cases per rule, (4) maximum number of uncovered cases in the training set and (5) number of rules used to test convergence of the ants. This study set parameter (3), (4) and (5) according to the default value (5, 10, 10, respectively) because these parameters were insignificant influence on classification results (Parpinelli et al., 2002). However, the number of iterations was tested in this study, including 1, 10, 25, 50, 100, 200, 300, 400, and 500, respectively. The number of ants was considered that 1, 5, 10, 100 and 500 times.

$$P_{ij}(t) = \frac{\tau_{ij}(t)\ \eta_{ij}(t)}{\sum_i^a \sum_j^{b_i} \tau_{ij}(t)\ \eta_{ij}(t)} \quad (1)$$

$$\eta_{ij} = \frac{\max\left(\sum_n \mathrm{freq}T_{ij}^1, \sum_n \mathrm{freq}T_{ij}^2, \ldots, \sum_n \mathrm{freq}T_{ij}^k\right)}{\sum_n T_{ij}} \quad (2)$$

$$\tau_{ij}(t=0) = \frac{1}{\sum_{i=1}^{A} b_i} \quad (3)$$

$$\tau_{ij}(t+1) = \tau_{ij}(t) + \tau_{ij}(t) \times Q \quad (4)$$

$$Q = \frac{TP}{TP+FN} \times \frac{TN}{TN+FP} \quad (5)$$

where τ and η are frequency and pheromone, respectively; ij and t indicate a index for term space and iteration, respectively; a is a particular factor; bi represents the set of a particular factor; Tij is a particular term; k describes the number of classes; n is the total number of Tij; A indicates the number of factors; Q is the quality measure of a rule that the formula can be simplified as sensitivity x specificity ; TP (True Positive), TN (True Negative), FN (False Negative) and FP (False Positive) are derived from confusion matrix.

3.2 Particle Swarm Optimization for Classification Task

PSO was proposed to solve optimization problems that simulate the local (or neighbor) and global relationships from social population. It has been explored for clustering task. A hybrid PSO/ACO algorithm was designed with the ACO concept and sequential covering approach to discover classification rules where it can be used for both nominal and numeric factors (Holden & Freitas, 2008). For a consistent comparison with ACO results, this study only considered a nominal case of the hybrid PSO/ACO algorithm using multi-interval discretization to discretize numeric factors before classification. The basic idea of the hybrid PSO/ACO classifier in a nominal case is to calculate the probability of each term for the same class and utilize the roulette selection choosing terms into a rule one by one the particles. Consequently, two quality measures, sensitivity x specificity (Eq. (5)) and precision (Eq. (6)) can be used to compare rules constructed by particles and decide the best one (global solution). Finally, neighbor particles searched by the Von-Neumann topology (Kennedy & Mendes, 2002) will update pheromone to increase probability of the suitable terms (local solution) for the next iteration.

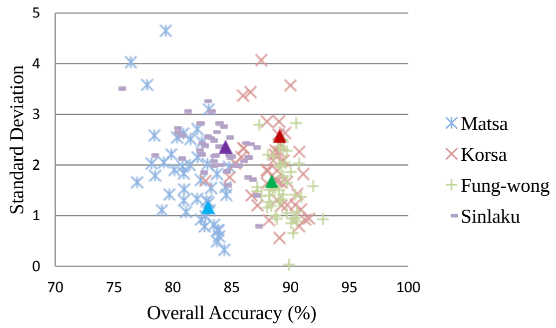
There are two parameters in the hybrid PSO/ACO algorithm, i.e. number of particles and iterations. The former was tested, including 1^2, 10^2, 20^2, 30^2, 40^2 and 50^2, respectively. The later contained 1, 10, 25, 50, 100, 200, 300, 400, and 500 times in this study.

$$Q = \frac{1+TP}{1+TP+FP} \quad (6)$$

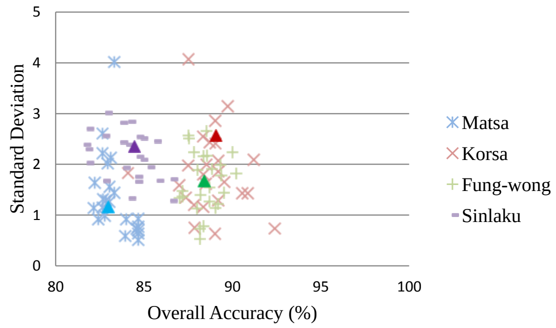
4. RESULTS

A two-phase experiment was tested to explore the acceptable and reasonable parameters for the classification task. This study tested the main parameters of the ACO and PSO classifiers, i.e. the numbers of iterations and ants (ACO) or particles (PSO). Firstly, different numbers of iterations are performed, and set the number of ants or particles as a default vale (the former is 5, the latter is 10^2) in order to observe the effect of iterations. Consequently, this study changed different numbers of ants or particles using a default of iterations (ACO is 100, PSO is 200).

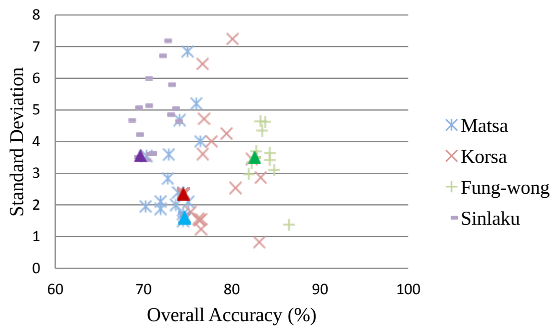
Figure 1(a), (c) and (e) show the results from different numbers of iterations for ACO and PSO based on four typhoon events. The same parameter setting repeated five runs; because the initial rule of ACO and PSO is random-based selection that will cause some variations. Every sample in these figures indicates one of the processes using different parameters or runs; and the triangles present the means of repeating runs derived from the default value. Figure 1(b), (d) and (f) display similar results but changing different numbers of ants or particles. It is obvious that using the default parameters of ACO or PSO can not only provide an acceptable and reasonable result but also avoid heavy work to explore the optimal parameters in our cases.



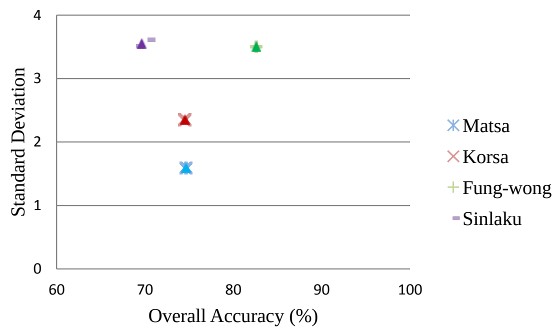
(a) Different numbers of iterations for ACO



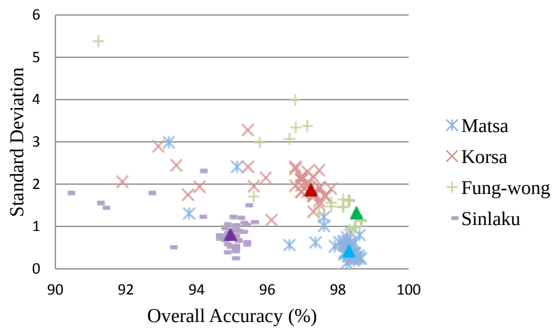
(b) Different numbers of ants for ACO



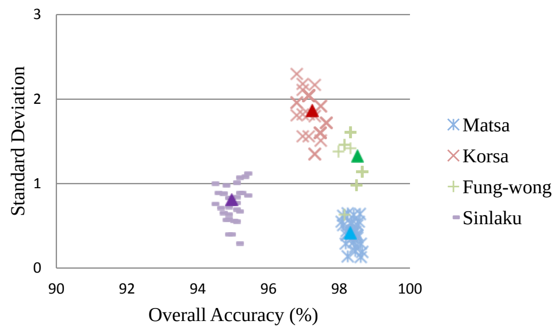
(c) Different numbers of iterations for PSO using sensitivity x specificity measure



(d) Different numbers of particles for PSO using sensitivity x specificity measure



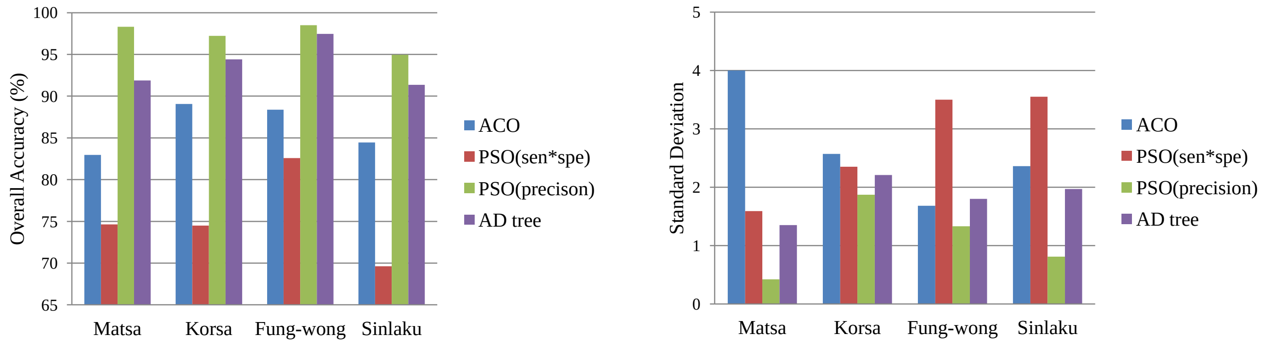
(e) Different numbers of iterations for PSO using precision measure



(f) Different numbers of particles for PSO using precision measure

Figure 1. Comparisons between different parameter settings (the triangles present the mean of repeating five runs derived from the default value)

According to above experiment, this study adopted the default parameters of ACO and PSO to model the landslide susceptibility for different events. On the other hand, a classical algorithm- AD tree (Active Directory tree) was applied to compare with the swarm intelligence approach. Figure 2 illustrates that PSO using the precision measure outperform ACO and PSO of sensitivity x specificity measure due to high overall accuracies and low variations each event. The results of both ACO and PSO indicate the means of repeating five runs. Furthermore, PSO using the precision measure can provide the competitive results comparing with AD tree. In summary, swarm intelligence has a potential and feasible to build landslide susceptibility model.



(a) Overall Accuracy (b) Standard Deviation

Figure 2. Comparisons between different classifiers

5. CONCLUSION

This study integrates eleven geospatial parameters and explores the feasible of swarm intelligence for landslide susceptibility modeling based on rainfall-induced shallow landslide events in the Shimen reservoir watershed in northern Taiwan. Experimental results indicate that applying the default value of ACO and PSO algorithm to build the model is an acceptable and reasonable way in the tested cases. In addition, the precision measure of PSO provides a better accuracy and stable result than the sensitivity x specificity measure of ACO and PSO. Comparing with a conventional algorithm, PSO produced a more plausible result. It may be of interest for future research to decrease the volume of data using feature and instance reduction for an effective modeling.

REFERENCES

Chang, K.-T., Chiang, S.H., Chen, Y.-Y., Mondini, A.C., 2014. Modeling the spatial occurrence of shallow landslides triggered by typhoons. Geomorphology, 208, pp. 137-148.

Dai, F.C., Lee, C.F., Ngai, Y.Y., 2002. Landslide risk assessment and management: an overview. Geomorphology, 64, pp. 65-87.

Du, Y., Teillet, P.M., Cihlar, J., 2002. Radiometric normalization of multitemporal high-resolution satellite images with quality control for land cover change detection. Remote Sensing of Environment, 82, pp. 123-134.

Fayyad, U.M., Irani, K.B., 1993. Multi-interval discretization of continuous valued attributes for classification learning. In: 13th International Joint Conference on Artificial Intelligence, pp. 1022-1027.

Holden, N., Freitas, A.A., 2008. A hybrid PSO/ACO algorithm for discovering classification rules in data mining. Journal of Artificial Evolution and Applications, 2008(2).

Kennedy, J., Mendes, R., 2002. Population structure and particle swarm performance. In: Proc. of the IEEE Conference on Evolutionary Computation, May, 2002, Hawaii, USA.

Oh, H.-J., Pradhan, B., 2011. Application of a neuro-fuzzy model to landslide-susceptibility mapping for shallow landslides in a tropical hilly area. Computer & Geosciences, 37, pp. 1264-1276.

Parpinelli, R.S., Lopes, H.S., Freitas, A.A., 2002. Data mining with an ant colony optimization algorithm. IEEE Transaction on Evolutionary Computing, 6(4), pp. 321-332.

Peduzzi, P., 2010. Landslide and vegetation cover in the 2005 Northern Pakistan earthquake: a GIS and statistical quantitative approach. Natural Hazards and Earth System Sciences, 10, pp. 623-640.

Tsai, F., Chen, L.C., 2007. Long-term landcover monitoring and disaster assessment in the Shiman reservoir watershed using satellite images. In: Proc. of 13th CeRES International Symposium on Remote Sensing, 29-30, October, 2007, Chiba, Japan.

Tsai, F., Lai, J.-S., Chen, W.W., Lin, T.-H., 2013. Analysis of topographic and vegetative factors with data mining for landslide verification. Ecological Engineering, 61, pp. 669-677.

Witten, I.H., Frank, E., 2005. Data mining: practical machine learning tools and techniques, 2nd edition. Morgan Kaufmann, San Francisco, USA.