INTEGRATING RULE-BASED ALGORITHMS WITH FUZZY RULE INDUCTION ON REGIONAL LANDSLIDE SUSCEPTIBILITY MODELING

Jhe-Syuan Lai[1] and Fuan Tsai[2]
National Central University, No. 300, Zhong-Da Rd., Jhongli, Taoyuan, Taiwan,
Email: [1]jslai0726@gmail.com; [2]ftsai@csrsr.ncu.edu.tw

KEY WORDS: Landslide Susceptibility, Decision Tree, Particle Swarm Optimization, Fuzzy Rule Induction.

ABSTRACT: This study integrates Decision Tree (DT) and Particle Swarm Optimization (PSO) algorithms with Fuzzy Rule Induction (FRI) operator respectively (called DT-FRI and PSO-FRI) to assess landslide susceptibility according to existing rainfall-induced and shallow landslide events. The constructed landslide susceptibility models are applied to classify and verify occurrence samples. In this study, two strategies are applied for the model verification, i.e. space- and time-robustness. The former is to separate samples into training and check data based on a single event. The latter is to predict (classify) later landslide events with a landslide susceptibility model which is constructed from earlier events. Eleven geospatial factors are considered, including topographic, vegetative, environmental, geological and man-made information. The landslide inventory and factors are overlapped to obtain the training and check data for modeling and verification. Experimental results show that applying the conventional DT algorithm can reach high modeling accuracy respectively based on the space-robustness strategy but both have poor performance to predict (classify) consequent events (time-robustness). After integrating with FRI, the prediction (classification) results are significantly improved, especially using PSO-FRI models.

1. INTRODUCTION

Landslide is one of the natural and geological hazards that can cause serious property losses and human casualties. Rainfall-induced and shallow landslides are frequently triggered by typhoons and other extreme weather events from June to September every year in Taiwan. Therefore, modeling and predicting landslides in order to prevent and mitigate disasters has become an important issue. Modeling landslide susceptibility is a fundamental and essential task in the landslide risk assessment and management framework (Dai et al., 2002). In this study, landslide susceptibility represents the likelihood of landslide occurrence in an area with given local terrain attributes (Brabb, 1984) and the triggering factor (i.e. rainfall) is not taken into account because it may change in a short period (Dai et al., 2002). A number of related works for evaluating landslide susceptibility have been proposed and can be classified into three main categories(Clerici et al., 2006; Wan, 2009; Yilmaz, 2010), i.e. deterministic, heuristic and statistical approaches. On the other hand, remotely sensed images, LiDAR point clouds and GIS datasets can be used effectively to monitor and investigate long-term landslides in a regional scale (e.g. Peduzzi, 2010; Tsai and Chen, 2007). As geospatial technologies and data advance, data mining based methods have been intensively developed to build landslide susceptibility models, such as decision tree (e.g. Nefeslioglu et al., 2010; Tsai et al., 2013), particle swarm optimization (e.g Lai and Tsai, 2014; Wan, 2013) and fuzzy rule induction (e.g. Muthu et al., 2008; Gorsevski and Jankowski, 2010).

Among them, Decision Tree (DT) is a non-parametric method that does not require classification assumption. However, it might be inadequate in dealing with complicated cases. Particle Swarm Optimization (PSO) 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. The PSO strategy is to change the velocity and position of particles that performs iteration process to find the global or best solution. However, few studies have adopted the PSO algorithm to model landslide events. Fuzzy Rule Induction (FRI) algorithm has been proposed to solve non-linear and uncertainty cases, but the classification rules (or expert database) are usually constructed manually.

This study tries to integrate DT and PSO classifiers with FRI algorithms respectively (called DT-FRI and PSO-FRI) for landslide susceptibility modeling based on existing rainfall-induced and shallow landslide events. The constructed landslide susceptibility models are applied to classify and verify occurrence samples. In this study, two strategies are investigated for the model verification, i.e. space- and time-robustness. The former is to separate samples into training and check data based on a single event. The latter is to predict (classify) later landslide events with a landslide susceptibility model which is constructed from earlier events. Eleven geospatial factors are considered, including topographic, vegetative, environmental, geological and man-made information. Finally, the landslide inventory and factors are overlapped to obtain the training and check data for modeling and verification.

2. STUDY SITE AND DATA

The Shimen reservoir watershed in Taiwan is selected as the study site as shown in Figure 1. It covers about 763.4 square kilometers. The elevation in the watershed measured from DEM (Digital Elevation Model) data ranges between 250 to 3,500 meters above sea level. The major land-cover is forest, but there are sparse agricultural activities. Landslides are commonly induced by heavy rainfall in this region and the debris flows are flushed into the reservoir, causing various problems in water supply and resource management.

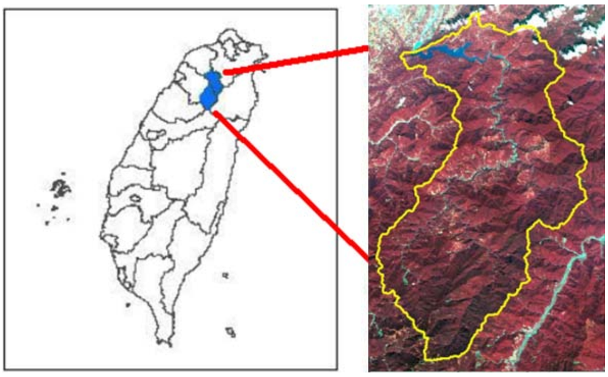

Figure 1. Study site

Eleven factors are considered in this study as listed in Table 1. The NDVI (Normalized Difference Vegetation Index) is the 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 analysis 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, vector-based factors and DEM into the 10 m X 10 m cell size in order to overlay with the NDVI image. 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 deposition 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 (Deng et al., 2016). Table 2 displays four typhoon events and the numbers of landslide pixels occurred after each typhoon.

Table 1. Used landslide causative factors

Original data	Useddata (Raster format)	Original resolution/scale
DEM	Elevation	40 x 40 m
	Slope	
	Aspect	
	Curvature	
SPOT Images	NDVI	10 x 10 m
River map	Distance toriver	1/5,000
Road map	Distance to road	1/5,000
Fault map	Distance to fault	1/50,000
Land-use map	Landuse	1/5,000
Soil map	Soil	1/25,000
Geology map	Geology	1/50,000

Table 2. Typhoon events and number of landslide pixels

Typhoon event	Date	No. of landslide pixels
Aere	2004/8	23,166
Matsa	2005/8	1,480
Sepat	2007/8	227
Wipha	2007/9	218

3. PROCEDURE AND METHODOLOGY

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. Subsequently, 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, for each typhoon, 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 spatial- and event-based landslide and non-landslide samples in order to input the classifiers for space- and time-robustness verifications respectively. This study selects 2/3 samples to build the model and the remainder are used for verification in the space-robustness tasks. For the time-robustness verification, the later events are predicted (classified) by the earlier event model. In our case, Typhoon Aere, that has the largest number of landslide samples, is treated as the training data to construct landslide susceptibility models for classifying Typhoon Matsa, Sepat and Wipha events, respectively. To assess the models, this study calculates Overall Accuracy (OA) and Kappa coefficient, and compares the landslide sample's susceptibility for the space- and time-robustness verification, respectively.

For constructing the DT-FRI and PSO-FRI models, landslide factors are discretized (Fayyad and Irani, 1993) firstly because the combination of decision tree and fuzzy rule induction is available only for discrete data. Consequently, the fuzzy memberships of landslide factors should be assigned. This study characterizes the fuzzy relationships as the triangular shape, and then the number of membership functions and discrete subsets are equal in each landslide factor. Figure 2 illustrates an example of fuzzy memberships. Finally, the representative rules conducted by the decision tree and particle swarm optimization (Holden and Freitas, 2008) algorithms under the space-robustness verification are selected and embedded in the fuzzy rule induction algorithm based on a simple and popular fuzzy inference, Mamdani's fuzzy inference, to perform the time-robustness verification further.

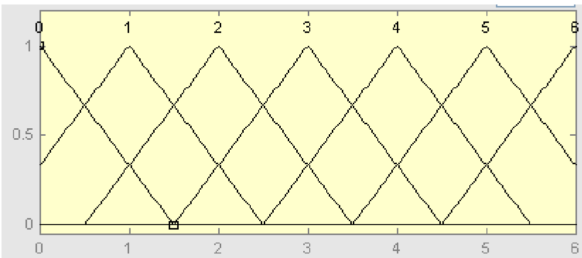

Figure 2. An example of fuzzy memberships in a specific factor (X: discrete subset, Y: fuzzy degree)

4. RESULTS

A two-phase verification is employed to examine the fidelity of the constructed models. First of all, this study performs the conventional decision tree algorithm to classify occurrence and non-occurrence samples based on the space-robustness verification event by event. The classification results are shown in Table 3. It is obvious that decision tree can separate landslide and non-landslide samples well to obtain high overall accuracies and Kappa coefficients. Therefore, the effectiveness of developing a space-robustness based algorithm for further improvement will be limited. Instead of confining in the space-robustness, this study switches to an approach focused on time-robustness.

Table 3. Evaluations of space-robustness verification using decision tree classifier

Typhoon Event	Aere	Matsa	Sepat	Wipha
OA (%)	92.85	97.91	92.86	91.22
Kappa	0.857	0.9583	0.8574	0.8241

To compare DT, PSO, DT-FRI and PSO-FRI performances, overall accuracy and landslide samples' susceptibility are used. Table 4 lists the overall accuracies considering different typhoon events and classifiers. Although overall accuracies are relatively lower than space-robustness verifications, these results show that the PSO-FRI outperforms others and can be further discussed in different perspectives. From a statistical point of view, a low overall accuracy means occurring errors in experiments. In other words, there exist significant disagreement between the training and check data. It can be connected to an assumption of data-driven approaches that the past landslide conditions will occur in the future but the collected data may be unsatisfied to predict all consequent events. Thus there are two scenarios that may cause classification errors. Firstly, the classified landslide samples reveal that similar conditions occurred in the past, but these locations in the check data are stable at present. On the contrary, the landslide areas are classified as a non-landslide class because the models do not have similar occurrence situations. These are reasons why the overall accuracies are lower than the space-robustness verification. Therefore, the crisp (hard) evaluators, such as overall accuracy and Kappa coefficient, may be inappropriate to reflect the advantage of landslide susceptibility analysis in the time-robustness verification. Instead, this study utilizes the landslide sample's susceptibility to assess the results.

Table 5 shows that decision tree classifies the landslide samples into the extreme susceptibility categories (i.e. [1, 0.75] and (0.25, 0)) and the hybrid methods (i.e. DT-FRI and PSO-FRI) estimate them into the middle susceptibility, (0.75, 0.5] especially. It is clear that DT-FRI and PSO-FRI classifiers provide more reasonable landslide susceptibility results than decision tree. Furthermore, the number of conducted rules is also shown in Table 6. The results reveal that PSO-FRI is more effective than other algorithms.

Table 4. Evaluations of time-robustness verification using DT, PSO, DT-FRI and PSO-FRI classifiers

OA (%)	Typhoon Matsa	Typhoon Sepat	Typhoon Wipha
DT	69.59	52.64	72.71
PSO	67.33	50.44	70.64
DT-FRI	69.39	56.39	63.76
PSO-FRI	79.29	57.71	79.82

Table 5. Landslide sample's susceptibility of time-robustness verification using DT, DT-FRI and PSO-FRI classifiers

Check event	Classifier	Landslide susceptibility interval (%)			
		[1, 0.75]	(0.75, 0.5]	(0.5, 0.25]	(0.25, 0]
Matsa	DT	38.51	6.69	7.57	47.23
	DT-FRI	0	89.19	10.81	0
	PSO-FRI	0	90.07	9.93	0
Sepat	DT	10.57	0.88	6.17	82.38
	DT-FRI	0	65.2	34.8	0
	PSO-FRI	0	53.3	46.7	0
Wipha	DT	45.87	5.96	6.42	41.74
	DT-FRI	0	85.32	14.68	0
	PSO-FRI	0	90.37	9.63	0

Table 6. Number of rules of DT, PSO, DT-FRI and PSO-FRI classifiers under the time-robustness

DT	PSO	DT-FRI	PSO-FRI
2,886	432	48	45

5. CONCLUSION

This paper presents a procedure to construct landslide susceptibility models based on rainfall-induced and shallow landslide events. Eleven geospatial factors are used, including topographic, vegetative, environmental, geological and man-made information. The landslide inventory and factors are overlapped to produce the training data for modeling and verification. There are two verifications in the study, i.e. space- and time-robustness. This study integrates both decision tree and PSO with fuzzy rule induction (i.e. DT-FRI and PSO-FRI) to classify samples and verify the prediction (classification) by the time-robustness method. The proposed model is also compared with the decision tree and PSO classifiers. The results indicate that the decision tree classifier can reach high classification accuracy under the space-robustness strategy but it and PSO have poor performance to predict (classify) subsequent events. Consequently, the landslide sample's susceptibility is applied to evaluate the results of time-robustness verification. The results show that PSO-FRI is more reasonable and effective than other algorithms in the study cases.

BIBLIOGRAPHY

Brabb, E. E., 1984. Innovative approaches to landslide hazard mapping. In Proc. of 4th International Symposium on Landslides, 16-21 September, Toronto, Canada, pp. 307-324.

Clerici, A., Perego, S., Tellini, C., Vescovi, P., 2006. A GIS-based automated procedure for landslide susceptibility mapping by the condition analysis method: the Baganza valley case study (Italian Northern Apennines). Environmental Geology, 50, pp. 941-961.

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

Deng, Y. C., Tsai, F., Hwang, J. H., 2016. Landslide characteristics in the area of Xiaolin Village during Morakot typhoon. Arabian Journal of Geosciences, 9(5), 332, 16 pages.

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.

Gorsevski, P. V., Jankowski, P., 2010. An optimized solution of multi-criteria evaluation analysis of landslide susceptibility using fuzzy sets and Kalman filter. Computer and Geosciences, 36, pp. 1005-1020.

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, 11 pages.

Lai, J. -S., Tsai, F., 2014. Application of swarm intelligence for landslide susceptibility modeling from geospatial data fusion. In: Proc. of 35th Asian Conference on Remote Sensing, 27-31 October 2014, Nay Pyi Taw, Myanmar.

Muthu, K., Petrou, M., Tarantino, C., Blonda, P., 2008. Landslide possibility mapping using fuzzy approaches. IEEE Transactions on Geoscience and Remote Sensing, 46(4), pp. 1253-1265.

Nefeslioglu, H. A., Sezer, E., Gokceoglu, C., Bozkir, A. S., Duman, T. Y., 2010. Assessment of landslide susceptibility by decision trees in the metropolitan area of Istanbul, Turkey. Mathematical Problems in Engineering, 2010, 15 pages.

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.

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

Wan, S., 2013. Entropy-based particle swarm optimization with clustering analysis on landslide susceptibility mapping. Environmental Earth Sciences, 68, pp. 1349-1366.

Yilmaz, I., 2010. Comparison of landslide susceptibility mapping methodologies for Koyulhisar, Turkey: conditional probability, logistic regression, artificial networks, and support vector machine. Environmental Earth Sciences, 61, pp. 821-836.