Sequent Data Mining Analysis for Rainfall-based Landslide Events with the Refinement of Landslide Samples and Feature Reduction

Jhe-Syuan Lai[1], Fuan Tsai[2]

[1]Department of Civil Engineering, National Central University, No. 300, Zhong-Da Rd., Jhongli, Taoyuan, Taiwan 32001, jslai0726@gmail.com

[2]Center for Space and Remote Sensing Research, National Central University, No. 300, Zhong-Da Rd., Jhongli, Taoyuan, Taiwan 32001, ftsai@csrsr.ncu.edu.tw

Abstract: Rainfall-based landslides have frequently induced by typhoons in Taiwan. Geoinformatics technologies and spatial data become mature and popular that can be used effectively to investigate and monitor natural disasters contiguously and to support decision-making for hazard prevention, mitigation and reconstruction. However, these huge data and various data types have pose a great challenge in processing and analysis.

Data mining (DM) provides a viable and effective solution to extract information or knowledge from vast and complex data set and can be used to analyze and construct landslide factor models. Bayesian network (BN) is the one of algorithms that explores the relationship between variables by conditional probability. It has been proved a powerful approach in the DM domain. This study adopts this operator to analyze landslide events. There were eleven landslide variables considered, including elevation, slope, aspect, curvature, Normalized Difference Vegetation Index (NDVI), fault, geology, soil, land use, river and road. For landslide identification, the candidate extents were extracted by image-based change detection. Subsequently, auxiliary data and field investigation were applied to reduce false alarms of landslides. Then analytic attributes were extracted from all landslide variables by landslide.

The objective of this study is to adopt the BN algorithm to predict sequent rainfall-based landslides of Shimen watershed reservoir in Taiwan since 2004 to 2008. Previous landslide events are used as training data to classify next event. Two subjects were further explored in this study. The first is landslide runout elimination from landslide samples, and the other is feature (variable) reduction. The former was performed in order to refine landslide samples for prediction improvement, because image-based interpretation cannot discriminate landslide and landslide runout area. The latter is to reduce redundancy of landslide events and variables. Experimental results demonstrate that the landslide runout elimination and feature reduction can improve the prediction accuracy and the computation efficiency while maintaining acceptable results.

Keywords: Landslide, Data Mining, Bayesian Network, Feature Reduction.