A novel ensemble decision tree-based CHi-squared Automatic Interaction Detection (CHAID) and multivariate logistic regression models in landslide susceptibility mapping

ABSTRACT

An ensemble algorithm of data mining decision tree (DT)-based CHi-squared Automatic Interaction Detection (CHAID) is widely used for prediction analysis in variety of applications. CHAID as a multivariate method has an automatic classification capacity to analyze large numbers of landslide conditioning factors. Moreover, it results two or more nodes for each independent variable, where every node contains numbers of presence or absence of landslides (dependent variable). Other DT methods such as Quick, Unbiased, Efficient Statistic Tree (QUEST) and Classification and Regression Trees (CRT) are not able to produce multi branches based tree. Thus, the main objective of this paper is to use CHAID method to perform the best classification fit for each conditioning factors, then, combined it with logistic regression (LR) to find the corresponding coefficients of best fitting function that assess the optimal terminal nodes. In the first step, a landslide inventory map with 296 landslide locations were extracted from various sources over the Pohang-Kyeong Joo catchment (South Korea). Then, the inventory was randomly split into two datasets, 70 % was used for training the models, and the remaining 30 % was used for validation purpose. Thirteen landslide conditioning factors were used for the susceptibility modeling. Then, CHAID was applied and revealed that some conditioning factors such as altitude, soil drain, soil texture and TWI, as terminal nodes and reflected the best classification fit. Then, a proposed ensemble technique was applied and the interpretations of the coefficients showed that the relationship between the decision tree branch nodes distance from drain, soil drain, and TWI, respectively, leads to better consequences assessment of landslides in the current study area. The validation results showed that both success and prediction rates, 75 and 79 %, respectively. This study proved the efficiency and reliability of ensemble DT and LR model in landslide susceptibility mapping.

Keyword: Landslide; Ensemble modeling; Decision tree; CHAID.LR; GIS; Remote sensing; South Korea