LEARNING TO PREDICT FOREST FIRES WITH DIFFERENT DATA MINING TECHNIQUES

Forest fires are the most prevalent cause of the global economy. There are already two systems operating in Slovenia that evaluates the possibility of fire threat in the natural environment: one operated by the Forest Institute of Slovenia and the other by Slovenian environment the agency, but spatial and temporal over-generalization, and in part, also outdated input of GIS data is the main problematic issues with these systems. This work intends to improve the existing models by including GIS, ALADIN (Aire Limitee Adaptation Dynamique Developpement International), MODIS (Moderate-resolution Imaging Spectroradiometer) data and the models for prediction of the stand height and canopy cover [12]; and extending their validity to the whole territory of Slovenia. The descriptive data is divided into 3 groups: Multitemporal MODIS, Meteorological ALADIN data, GIS Data. The data were analyzed with several different data mining algorithms for classification implemented in the WEKA data mining system [4]. They used: logistic regression, random forests, decision trees (J48), bagging and boosting ensemble methods. The widely used method for boosting is AdaBoost[9]. AdaBoost calls a given weak or base learning the algorithm repeatedly in a series of rounds. One of the main ideas of the algorithm is to maintain a distribution or set of weights over the training set. Several algorithms were used in the experiments, being logistic regression, random forests for boosting the trees. The results were analyzed using the Kappa statistics are used to evaluate the agreement between predicted and observed nominal values.

Reference:

Stojanova, Daniela & Panov, Pance & Kobler, Andrej & Džeroski, Sašo & Taškova, Katerina. (2006). Learning to predict forest fires with different data mining techniques.