

The Effect of Feature Selection on Pattern Classification

Authors : Chih-Fong Tsai, Ya-Han Hu

Abstract : The aim of feature selection (or dimensionality reduction) is to filter out unrepresentative features (or variables) making the classifier perform better than the one without feature selection. Since there are many well-known feature selection algorithms, and different classifiers based on different selection results may perform differently, very few studies consider examining the effect of performing different feature selection algorithms on the classification performances by different classifiers over different types of datasets. In this paper, two widely used algorithms, which are the genetic algorithm (GA) and information gain (IG), are used to perform feature selection. On the other hand, three well-known classifiers are constructed, which are the CART decision tree (DT), multi-layer perceptron (MLP) neural network, and support vector machine (SVM). Based on 14 different types of datasets, the experimental results show that in most cases IG is a better feature selection algorithm than GA. In addition, the combinations of IG with DT and IG with SVM perform best and second best for small and large scale datasets.

Keywords : data mining, feature selection, pattern classification, dimensionality reduction

Conference Title : ICIST 2014 : International Conference on Intelligent Systems and Technologies

Conference Location : Tokyo, Japan

Conference Dates : May 29-30, 2014