Abstract

Associative (association rule) classification has been widely applied in various classification applications due to its interpretability. Conventionally, association rule mining approaches (e.g., Apriori or FP-Growth) generate if-then rules (i.e., frequent itemsets) for each class separately, which are then used in a majority voting classification scheme. However, generating a large number of rules is computationally time-consuming as data (feature) size gets larger. In addition, most of generated rules, involving many features, are complicated so that new data points may fail to satisfy these rules. Therefore, it is hard to make an accurate decision on classification. In this paper, we propose a new algorithm that induces simple rules (i.e. individual “if-then” rules) based on the frequency information in different classes using a nearest neighbor concept. These rules are directly applied to determine the classification of new data points. In addition, an improved discretization approach is designed and embedded in the proposed algorithm to obtain categorical features from continuous features. The computational complexity is better than existing classification algorithms such as k-nearest neighbors (KNN), support vector machine (SVM), logistical regression (LR), decision tree (DT), and lazy associative classification (LAC). The proposed algorithm has been extended to handle imbalance and multi-class classification problems. The experimental tests show higher accuracy compared to the existing associative classification and other well-known classification algorithms. Furthermore, a case study of brain aneurysm prediction is conducted using the proposed algorithm. A real public dataset is employed to find the risk factors of brain aneurysms. The most significant association rules are induced, which are used for rupture risk prediction and aneurysm detection. A user-friendly interface system is designed, called Brain Aneurysm Prediction System (BAPS), to support the diagnostic and treatment planning.