Build a PSO-SA Optimization Classifier for the Imbalance Sequence Data

Student：Wen-Jen Lee　　Advisor：Dr. Chieh-Yuan Tsai

Institute of Industrial Engineering and Management
Yuan-Ze University

ABSTRACT

Sequential pattern mining and sequence classification are two popular data mining methods used to explore the change of customer behavior. However, traditional methods have poor predictive ability to identify minority instances when dealing with the class imbalance problem. Actually, the imbalance class problem such as fraud detection, medical diagnosis, spam detection and fault monitoring/inspection exists everywhere in the real world. Therefore, this study develops an effective method to cope with the imbalance sequence classification problem. In this study, the sequences are divided into several sequence subsets according to the class label of sequences. Then, the AprioriAll algorithm is applied for each sequence subset and finds its sequential patterns. Next, the pairwise coupling method is used to combine every pair of sequence subsets and form a set of binary class datasets. For every binary class dataset, Force multi-class imbalance sequence (FMCIS) method is developed to build a classifier. Each classifier will generate two similarity values for a sequence first, then construct the units in fuzzy preference relations due to these two similarity values. The units will be composed by the fuzzy preference relations and computed a set of non-dominated values. Finally, the final class label of a sequence will be predicted due to the maximal non-dominated value. To increase the prediction accuracy of the proposed classifier, a hybrid PSO-SA algorithm is developed to adjust the weights of each pattern in each classifier and the weights of each class in fuzzy preference relations. The results show that the proposed classification model is useful for the sequence classification with imbalance data and especially in the low support value. But the applying optimized weighting in fuzzy preference relations does not perform well as expected.

Keyword: Sequential Pattern Mining、Class Imbalance Problem、Sequence Classification Problem、Particle Swarm Optimization