

# A Compound Method for Solving Sequence

## Classification Problems

Student : Chih-Jung Chen

Advisor : Dr. Chi-Yang Tsai

Institute of Industrial Engineering and Management  
Yuan-Ze University

### ABSTRACT

Recently, considerable attention has focused on compound sequence classification methods which integrate multiple data mining techniques. Among these methods, sequential pattern mining (SPM) based sequence classifiers are considered to be efficient for solving complex sequence classification problems. Although previous studies have demonstrated the strength of SPM-based sequence classification methods, the challenges of pattern redundancy, inappropriate sequence similarity measures, and hard-to-classify sequences remain unsolved. This research proposes an efficient two-stage SPM-based sequence classification method to address these problems. In the first stage, during the sequential pattern mining process, redundant sequential patterns are identified if the pattern is a sub-sequence of other sequential patterns. A list of compact sequential patterns is generated excluding redundant patterns and used as representative features for the second stage. In the second stage, a sequence similarity measurement is used to evaluate partial similarity between sequences and patterns. Finally, a particle swarm optimization-AdaBoost (PSO-AB) sequence classifier is developed to improve sequence classification accuracy. In the PSO-AB sequence classifier, the PSO algorithm is used to optimize the weights in the individual sequence classifier, while the AdaBoost strategy is used to adaptively change the distribution of patterns that are hard to classify. In addition to apply in market database and protein database, the proposed method is improved to solve mobile sequence classification problem. Mobile sequence includes three information as the moving patterns (Location), purchase patterns (Items), visiting time of customers. The proposed method is improved to develop a Location-Item-Time (LIT) closed sequential pattern mining based sequence classification method to apply in mobile commerce database. The Location-Item-Time (LIT) closed sequential pattern mining method is proposed for efficiently discovering the user's behavior LIT closed patterns and can eliminate redundant LIT patterns. The extracted LIT closed sequential patterns, considered as representative features, are then used to build the

classification model.

Keyword: Sequence classification 、 ClosedMining algorithm 、 Particle swarm optimization (PSO) algorithm 、 Adaptive boosting (AdaBoost)