Applying Sequential Pattern Mining Technique to Build an Optimized Classification Model for Time-interval Sequences

Student：Chun-Ju Chien   Advisor：Dr. Chieh-Yuan Tsai

Institute of Industrial Engineering and Management
Yuan-Ze University

ABSTRACT

To survive in current competitive and fast-changed business environment, enterprises need to know their customers behavior in depth to provide the right services to customers at right time. However, the customer databases in enterprises are usually large and disordered which makes customer behavior analysis difficult. Sequential pattern mining and sequence classification are two popular data mining methods to explore customer behavior. The former can discover frequent occurring patterns, while the latter can assign a most probable class label to a given sequence based on the characteristics of the sequence. However, previous researches seldom discussed sequence classification problem related to time information. Without time information, two sequences with the same itemsets but different time-intervals will be classified as the same class, which is inappropriate in customer behavior analysis. For this reason, this research presents a time-interval sequence classification methodology to help decision makers make better business strategies to satisfy their various customers.

The proposed sequence classification methodology includes two main stages. The first stage is time-interval sequential pattern mining, which employs I-PrefixSpan algorithm to discover time-interval sequential patterns in the large database. The second stage is time-interval sequence classification method, which contains sequence similarity measure, time-interval sequence classification model establishment, and model optimization procedure. A simple case and NorthWind database which is a large scale database are employed to test the classification method. The experiment results indicate the proposed time-interval sequence classification method is feasible and efficient.

Keyword: Time-interval sequential pattern mining   Time-interval sequence classification   Sequence similarity measure   Particle swarm optimization