Abstract

We are given a large database of customer transactions, where each transaction consists of transaction-id, the items bought in the transaction and the transaction time. The whole set of transaction is divided into a number of segments called durations (intervals) based on transaction time. And the dividing standard can be monthly, quarterly or yearly. We introduce the problem of mining strong association rules between consecutive durations using FP-tree and correlation coefficient, which is used to quantitatively describe the strength and sign of a relationship between two variables. This paper deals with the changes in the correlation between any two itemsets at the transition of the consecutive duration. Milestone is a change over point between durations. The transition may be positive or negative which are time points at which the pattern is either positively or negatively correlated. Also the method provides rare items, whose support is poor but are highly correlated.
- J. Han, J. Pei and Y. Yin. “Mining frequent patterns without candidate generation”. In ACM SIGMOD 2000.
- H. Xiong, S. Shekhal, P. N. Tan, V. Kumar, “Exploiting a support-based Upper Bound of Pearson’s Correlation Coefficient for Efficiently Identifying strongly correlated pairs.
- J. Zhang, J. Feigenbaum, “Finding highly correlated pairs efficiently with powerful pruning.”

Index Terms

Computer Science
Information Sciences
Keywords
Association Rule mining  support  Itemsets  Frequent Patterns  FP-Tree
Correlation
Correlation Coefficient