

MINING TEMPORAL RARE UTILITY ITEMSETS IN LARGE DATABASES USING RELATIVE UTILITY THRESHOLDS

CHUN-JUNG CHU¹, VINCENT S. TSENG² AND TYNE LIANG¹

¹Department of Computer Science
National Chiao Tung University
No.1001, Dasyue RD., HsinChu City 300, Taiwan
{cjchu; tliang}@cis.nctu.edu.tw

²Department Computer Science and Information
National Cheng Kung University
No.1, Dasyue RD., Tainan City 701, Taiwan
tsengsm@mail.ncku.edu.tw

Received July 2007; revised November 2007

ABSTRACT. *Utility itemsets are considered to be the different values of individual items such as utilities, and utility mining aims at identifying the itemsets with highest utilities. The temporal significant rare utility itemsets are those itemsets which appear infrequently in the current time window of large databases but are highly associated with specific data. In this paper, we propose two novel algorithms, namely TP-RUI (Two-Phase Rare Utility Itemsets) -Mine and TRUI (Temporal Rare Utility Itemsets) -Mine, for mining temporal rare utility itemsets from temporal databases. To the best of our knowledge, this is the first work on mining temporal rare utility itemsets from temporal databases. The novel contribution of TRUI-Mine is particularly that it can effectively identify the temporal rare utility itemsets by generating fewer temporal high transaction-weighted utilization 2-itemsets in temporal databases. In this way, the process under all time windows of temporal databases can be achieved effectively with limited memory space, less candidate itemsets and CPU I/O time. The experimental results show that TRUI-Mine can discover the temporal rare utility itemsets with higher performance and less candidate itemsets compared to the other algorithm TP-RUI-Mine that is also proposed in this paper by us under various experimental conditions.*

Keywords: Utility mining, Temporal significant rare utility itemsets, Temporal databases, Association rules

1. Introduction. The mining of association rules for discerning the relationship between data items in large databases is a well studied technique in the data mining field with representative methods like *Apriori* [1,2]. The problem with mining association rules can be distilled into two steps. The first step involves finding all frequent itemsets (or say large itemsets) in databases. Once the frequent itemsets are found, generating association rules is straightforward and can be accomplished in linear time.

An important research issue that extends from the mining of association rules is the discovery of temporal association patterns in temporal databases due to the wide variety of applications on various domains. Temporal data mining can be defined as the activity of looking for interesting correlations or patterns in large sets of temporal data accumulated for other purposes [6]. For a database with a specified transaction window size, we may use an algorithm like *Apriori* to obtain frequent itemsets from the database. For time-variant temporal databases, there is a strong demand to develop an efficient and effective method to mine various temporal patterns [11,19]. However, most methods designed for