

## CLUSTERING PERIODIC FREQUENT PATTERNS USING FUZZY STATISTICAL PARAMETERS

FOKRUL ALOM MAZARBHUIYA<sup>1</sup> AND MUHAMMAD ABULAISH<sup>2</sup>

<sup>1</sup>College of Computer Science  
King Khalid University  
P.O. Box 3236, Abha, Saudi Arabia  
fokrul.2005@yahoo.com

<sup>2</sup>Center of Excellence in Information Assurance  
King Saud University  
P.O. Box 92144, Riyadh 11653, Saudi Arabia  
mAbulaish@ksu.edu.sa

Received December 2010; revised April 2011

**ABSTRACT.** *Frequent pattern mining from super-market transaction datasets is a well-stated data mining problem and consequently there are number of approaches including association rule mining to deal with this problem. However, super-market transaction datasets are generally temporal in the sense that, when a transaction happens in a super-market, the time of the transaction is also recorded. A number of techniques have been proposed to find frequent itemsets from temporal datasets in which each itemset is associated with a list of time intervals in which it is frequent. Considering the time of transactions as calendar dates, there may exist various types of periodic patterns viz. yearly, quarterly, monthly, daily, etc. And, if the time intervals associated with a periodic itemset are kept in a compact manner then it turns out to be a fuzzy time interval for which the set superimposition method can be used. In this paper, we propose an agglomerative hierarchical clustering algorithm to find clusters among the periodic itemsets. Since the fuzzy number is invariant with respect to shifting, we define similarity measure using the variance of fuzzy intervals associated with frequent itemsets. The efficacy of the proposed method is established through experimentation on real datasets.*

**Keywords:** Data mining, Clustering, Temporal patterns, Locally frequent itemset, Set superimposition, Fuzzy time-interval

1. **Introduction.** Clustering is one of the well-known data mining problems which follows unsupervised learning approach and it is very useful for the discovery of data distribution and patterns in the datasets with unknown class-labels. The goal of the clustering process is to discover both the dense and sparse regions in a dataset. There are two main approaches to clustering: *hierarchical clustering* and *partitioning clustering*. In hierarchical clustering, the dataset is divided into a sequence of partitions, in which each partition is nested into the next partition in the sequence. The hierarchical clustering creates a *hierarchy* of clusters from small to big or big to small and consequently it is termed as *agglomerative* or *divisive* clustering techniques respectively. Clustering of numerical data has been studied in the past [8]. However, in real life, we come across datasets that contain different types of data such as binary, categorical, spatial, ordinal, temporal or mixture of these. During the last few years, a number of new and interesting algorithms for clustering categorical and spatial data have also been proposed [5, 15, 16, 18, 20].

Association rule mining is another data mining problem which focuses on deriving associations among data. The association rule mining problem was formulated by Agrawal