Contents

Chapter 1

Our Proposed Approaches

In this section we will discuss about our proposed approach for mining frequent pattern over large uncertain stream data. Stream Data has a special property that it comes and flows away. For this reason we will lose data after data stream has flown away. To resolve this we will proposed a window based approach where we will keep the most recent information in a tree structure as the most recent data is most valuable. Later we will show how the window will slide, remove old data and insert new data in the window. As, for uncertain data stream each same item in different transaction has different existential probability, it becomes very hard to merge (share) this node in the tree. This uncertainty property of item makes the tree huge. We have proposed a new U^{cap} value for each item that helps to share a single node when constructing the tree which we named as *US-tree*. We will show that our proposed tree *US-tree* will be very compact and very efficient for later mining. Later will describe an approach for mining the *US-tree* named *USFP-growth* which is *FP-growth* like approach. Later we will propose a method for filtering false positive from found most probable frequent patterns.

1.1 Analysis Uncertain Stream Data Properties

1.1.1 Stream Property

1.1.2 Uncertainty Property

For expected support calculation in an uncertain database we get the following equation

$$ExpSup = \sum_{i=0}^{UDB} [\prod_{x \in I} p(x, t_i)] \quad (1.1)$$

where,

- ***I*** is itemset,
- ***p(x, ti)*** is existential probability value for any item ***x*** in transaction t_i
- ***UDB*** is an uncertain database.

1.2 Preliminaries

1.3 Mining Frequent Patterns from Uncertain Databases

1.3.1 US-tree Construction

1.3.2 Sliding Window

1.3.3 Mining US-tree : FPUS-growth

1.3.4 False positive reduction

1.3.5 Algorithm

1.4 Summary