ABSTRACT

The key technology is the smart consumer advice Mining systems in information technology plays an increasingly an important role in the actions and decision making. Also common Mining is an important step research properties of association rules. It is one of the most important areas of research in data mining. Database to learn by frequent jobs in matters of doubt, the weighted they are so likely to be important, and the elements into often found great item sets users. But the introduction of weight makes frequent item sets may fall back to the enclosure is to be a matter longer. Hence frequently used items in the distance, not to search they are what are referred to the poor efficiency. Said weight trial frequency locked to its own gravity and item set sit is a serious amount of material introduced and tested. From these two properties WD-FIM. (own weight and the end of the trial Evaluation of the algorithm is set before us by frequent, repeated the space of the face of the frequent item sets, and improves the hurdle weighted with stones then efficiency. We proposed the alternative method of Pincer-search algorithm yields maximum frequent items, association rules, accuracy and minimum execution time.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Abbreviation	Full form
WD-FIM	Weight judgement Downward closure property based frequent Item sets Mining
ARM	Association rule mining
IWFPM	Interested Weighted Frequent Pattern Mining with Multiple supports
IWI	Infrequent Weighted Itemset
WARM	Weighted Association Rule Mining
FIM	Frequent Item set Mining
RDBMS	Relational Database Management System
UML	Unified Modeling Language

CHAPTER 1

INTRODUCTION

The key technology is the smart consumer advice Mining systems in information technology plays an increasingly an important role in the actions and decision making. The weight judgment downward closure property for weighted frequent item sets and the existence property of weighted frequent subsets are introduced and proved first. Based on these two properties, the WD-FIM (Weight judgment Downward closure property based Frequent Item sets Mining) algorithm is proposed to narrow the searching space of weighted frequent item sets and improve the time efficiency**.** Data mining is an emerging technique that addresses the problem of restructuring the data into the useful information. The rule generation methods are used in the area of data mining to determine the relationship between different items. The Association Rule Mining (ARM) is widely used for finding data patterns that reveals the combination of events occurring simultaneously based on the associations among a large set of data items. The Apriori algorithm (Agrawal and Srikant, 1994) is a popular algorithm for extracting high frequent item sets from a database using the predefined threshold measures such as minimum support and minimum confidence. In this paper, on the basis of the weight judgment downward closure property, the Pincer search algorithm is proposed to yields maximum frequent items, association rules, accuracy and minimum execution time.

The main contributions of this paper are listed as following.

The weight judgment downward closure property and the existence property of weighted frequent subsets for uncertain databases are introduced and proved. The weight judgment downward closure property can be used to narrow the searching space of weighted frequent item sets. The existence property of weighted frequent subsets can ensure all the weighted frequent item sets be discovered.

The Pincer-search algorithm is proposed to yields maximum frequent items, association rules, accuracy and minimum execution time.
A considerable amount of experiments are conducted on both real-life and synthetic datasets to evaluate the performance of the proposed Pincer-Search algorithm in terms of runtime, number of patterns and memory consumption.

Frequent item set mining is one of the best known and most popular data mining methods. Originally developed for market basket analysis, it is used nowadays for almost any task that requires discovering regularities between (nominal) variables. This paper provides an overview of the foundations of frequent item set mining, starting from a definition of the basic notions and the core task. It continues by discussing how the search space is structured to avoid redundant search, how it is pruned with the a priori property, and how the output is reduced by confining it to closed or maximal item sets or generators. In addition, it reviews some of the most important algorithmic techniques and data structures that were developed to make the search for frequent item sets as efficient as possible

EXISTING SYSTEM

A hyperlinked structure based algorithm called UH-mine to mine frequent patterns from uncertain data.
a tree-based mining algorithm called UF-growth which also constructs a tree structure to store the contents of the uncertain datasets, like its counterpart - the FP-growth algorithm for mining precise data.
In order to reduce the tree size at proposed the UFP-growth algorithm. To further reduce the tree size. An uncertain frequent pattern mining algorithm called CUF growth, which builds a new tree structure called CUF-tree.
AT-Mine is another tree-based efficient approach proposed to overcome the fatal problems of CUFP-Mine tree.

Disadvantages of existing system

In the WDFIM method to using to get low amount of mining item sets.

Low mining efficiency and coverage

Proposed system

We introduce an Pincer-search algorithm for efficient mining of the frequent item sets. The prefiltering method is applied to the input dataset to remove the item having low variance. Data discretization is performed and Pincer-search is applied for mining the frequent item sets.

The proposed Pincer-search algorithm yields maximum frequent items, association rules, accuracy and minimum execution time than the existing algorithms.

Advantages of proposed system:

Prediction accuracy is improved by using the association rules.

Effective generation of the frequent item sets and association rules is ensured by maintaining the feasibility of the neural network.

Table 1.3.1.(a): A transaction database as running example

T-id	Items Bought	Frequent Items(ordered)
101	W,t,u,g,s,d,f	u,s,d,w
102	h,a,l,e,l,d	d,e,i
103	J,q,o,f,w	o,q
104	k,d,z,v,f	d,f,k
105	e,u,y,f,o,p	y,f,e

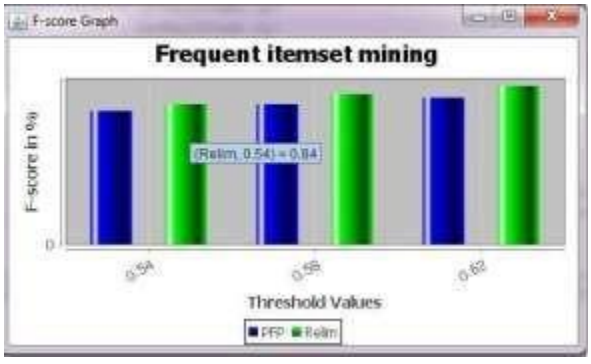

Fig.1.3.2 (b): F-score in % on Data Set

CHAPTER 2

LITERATURE SURVEY

INTRODUCTION

Literature survey is the most important step in software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then the next step is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need lot of external support. This support can be obtained from senior programmers, from book or from websites. Before building the system the above consideration are taken into account for developing the proposed system. The major part of the project development sector considers and fully survey all the required needs for developing the project. For every project Literature survey is the most important sector in software development process. Before developing the tools and the associated designing it is necessary to determine and survey the time factor, resource requirement, man power, economy, and company strength. Once these things are satisfied and fully surveyed, then the next step is to determine about the software specifications in the respective system such as what type of operating system the project would require, and what are all the necessary software are needed to proceed with the next step such as developing the tools, and the associated operations.

Title 1: IWFPM: Interested Weighted Frequent Pattern Mining with Multiple Supports
Author: Xuyang Wei1*, Zhongliang Li1, Tengfei Zhou1, Haoran Zhang1, Guocai Yang1

Association rules mining has been under great attention and considered as one of momentous area in data mining. Classical association rules mining approaches make implicit assumption that items'importance is the same and set a single support for all items. This paper presents an efficient approach for mining users' interest weighted frequent patterns from a transactional database. Our paradigm is to assign appropriate minimum support (minsup) and weight for each item, which reduces the number of unnecessary patterns. Furthermore, we also extend the support-confidence framework and define an interest measure to the mining algorithm for excavating users' interested patterns effectively. In the end, experiments on both synthetic and real world datasets show that the proposed algorithm can generate more interested patterns.

Title 2: Discovery of Infrequent Weighted Itemset with High Utility

Author: KalyaniTukaramBhandwalkar1 ,Mansi Bhonsle2

Knowledge discovery has been an interesting area of research due to its various applications. Traditionally, frequent pattern mining plays an important role. Generally, infrequent items within the dataset are ignored. Infrequent Itemset mining is a variation of frequent itemset mining where rarely occurring patterns are discovered. This paper tackles the issue of discovering rare and weighted item sets, i.e., the infrequent weighted itemset (IWI) mining problem. Also high utility item sets are discovered using UP Growth algorithm. Proposed system not only considers the frequency of the item sets but also considers the utility associated with the item sets.

Title 3: An Uncertainty-based Approach: Frequent Itemset Mining from Uncertain Data with Different Item Importance

Author: Gangin Lee, Unil Yun1 and HeungmoRyang

Since Itemset mining was proposed, various approaches have been devised, ranging from processing simpleitem-based databases to dealing with more complex databases including sequence, utility, or graph information. Especially, in contrast to the mining approaches that process such databases containing exact presence or absence information of items, uncertain pattern mining finds meaningful patterns from uncertain databases with items' existential probability information. However, traditional uncertain mining methods have a problem in that it cannot apply importance of each item obtained from the real world into the mining process. In this paper, to solve such a problem and perform uncertain itemset mining operations more efficiently, we propose a new uncertain itemset mining algorithm additionally considering importance of items such as weight constraints. In our algorithm, both items' existential probabilities and weight factors are considered; as a result, we can selectively obtain more meaningful item sets with high importance and existential probabilities. In addition, the algorithm can operate more quickly with less memory by efficiently reducing the number of calculations causing useless itemset generations. Experimental results in this paper show that the proposed algorithm is more efficient and scalable than state-of-the-art methods.