Implementation of Myanmar Sculptures Gallery By using Frequent Pattern Mining

Hnin Hnin Aye,Khaing
University of Computer Studies, Yangon
decembersnow.14@gmail.com;khaingaugust@gmail.com

Abstract

In this paper, frequent pattern mining for Myanmar sculptures gallery is proposed. The proposed system is expects to study the FP-Growth theory, how the transaction database is preprocessed in a way that is common to basically all frequent item set mining algorithm. FP-Growth theory contains a transaction database. Transaction database contains transactions and items. In proposed system, transactions are transaction_Date and items are sculptures category. Sculptures category contains elephant, peacock, dragon, horse, ankone, lawkanaut, women, etc. This proposed system is running the FP-Growth theory, user is request date. And then building the FP-Tree and generating the frequent pattern. This frequent pattern is more seller things in our organization.

Keywords: frequent pattern mining, FP-Growth, FP-Tree

1. Introduction

Data mining represents a process developed to examine large amount of data routinely collected. The term also refers to a collection of tools used to perform the process. Data mining is used in most areas where data are collected marketing, health, communication, finance, etc. Data mining is a process of discovering advantages in data.

Mining frequent patterns in transaction database, time series database and many other kinds of database has been studied popularly in data mining research. So, we proposed the Myanmar sculptures gallery system searching the mining frequent patterns by using the FP-Growth theory in data mining. The FP-Growth method is compressing a large database into a compact. FP-Tree structure is a highly condense but compress for frequent pattern mining. FP-Tree based frequent pattern mining method becomes developing and efficient. No candidate generation. It only needs sub-database task [4].

In this paper, frequent pattern mining for Myanmar sculptures gallery system is proposed. The rest of this paper is organized as follows: Section 2 describes the motivation. Section 3 describes proposed system. Section 4 reviews related work. Section 5 reviews advantages of proposed system. Finally, concludes this paper in section 6.

2. Motivation

Data mining (sometimes called data or knowledge discovery) is the process of analyzing data from different perspectives and summarizing it into useful information. Information that can be used to increase revenue, cuts cost or both.

Data mining software is one of the numbers of analytical tools for analyzing data. It allows user to analyze data from many different dimensions or angles, categorize it and summarize the relationships identified. Technically, data mining is the process of finding correlations or patterns among dozens in large relational database [8].

Most of association rule algorithm use a candidate generation set and test approach. FP-Growth is such an algorithm that tried to avoid candidate generation in mining frequent pattern. In FP-Growth algorithm, a structure called frequent pattern tree (FP-Tree) is used to store the compresses and important into about frequent pattern [3].

We proposed the frequent pattern mining for Myanmar sculptures gallery system by using FP-Growth theory because association rule mining searches for interesting relationships among items in a given data set. Myanmar sculptures are more knowing and attraction the tourisms. We ambition is to become the Myanmar sculptures interest in many tourists. So, we proposed this system.

3. Proposed System

In this paper, we use frequent pattern mining system based on data mining aspect. Among of many data mining tasks, FP-Growth theory can be used for Myanmar sculptures gallery. Frequent pattern tree (FP-Tree) and algorithm is called FP-Growth algorithm that allows mining of the frequent item sets without candidate generation.

3.1 FP-Tree Definition

FP-Tree is a frequent pattern tree. Formally, FP-Tree is a tree structure defined below: 1.It consists of one root labeled as null a set of item prefix sub trees as the children of the root and a frequent item header table. 2. Each node in the item prefix sub trees has three fields: 2.1 Item name to register which item this node represents. 2.2 Count the number of transaction represented by the portion of the path reaching this node and node link that link to the next node in the FP-Tree .2.3 Carrying the same item name or null if there is one. Each entry in the frequent item header table two fields: 1.Item Name 2. Header of the node –link that points to the first node in the FP-Tree carrying the item name.

3.2 Mining the FP-Tree

Start at the last item in the table. Find all paths containing item. Follow the node link. Identify conditional patterns. In paths, pattern is required frequency. Build conditional FP-Tree C. Append item to all paths in C, generation frequent patterns. Mining C, items are appending recursively. Remove item from table and tree.

3.3 FP-Tree Construction Algorithm

Input: a transaction database and minimum support threshold
Output: Its frequent pattern tree, FP-Tree

(1). The FP-Tree is constructed in the following steps: (a). Scan DB once: Collect the set of frequent items F and their supports. Sort F in support descending order as L, the list of frequent items. (b). Create a root of an FP-Tree, T and label it as null. For each transaction Trans in DB do the following: Select and sort the frequent items in Trans according to the ordering to the order of L. Let the sorted frequent items in Trans be [p/P], where p is the first element and P is the remaining list. Call insert- tree ([p/P], T).

Insert-tree ([p/P], T)

{check if T has a child N where N.item-name = p.item-name then increment N count by 1.else Create a new node with count 1, its parent linked to nodes with some item-name via node-link structure. Call insert till P is non empty.}

(2). Mining of an FP-Tree is performed by calling FP-Growth (FP-Tree, null), which is implemented as follows:

Procedure FP-Growth (Tree, α)

(1) if Tree contains a single path P then

(2) for each combination (denoted as β) of the nodes in the path p.

(3) generate pattern β U α with support = minimum support of nodes in β.

(4) else for each a_i , in the header of Tree {

(5) generate pattern $\beta = a_i$ U α with support = a_i . support ;

(6) Construct β, s conditional pattern base and then β's conditional FP-Tree Tree $_\beta$;

(7) If Tree $_\beta \neq \emptyset$ then

(8) Call FP-Growth (Tree $_\beta$, β); }

The FP-Growth algorithm for discovering frequent item sets without candidate generation.

3.4 Overview of the system

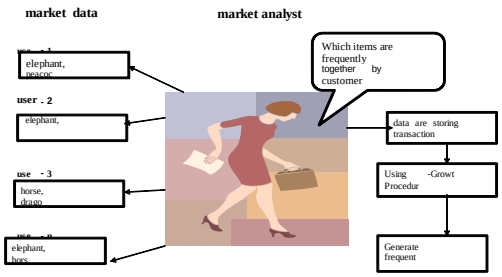

Figure1: Overview of the system

In figure 1, market data is a collection of users buying items. For example, user n is buying elephant, peacock, horse, etc. Market analysis is observed which items are frequently purchased together by my customer. And then data (such as elephant, peacock, horse, etc) are storing in transaction database and generating the frequent pattern by using FP-Growth theory.

3.5 System Implementation

In this system consists of two parts: admin and observer users. If username and password are corrected, admin can lock into the system. Admin can work running FP-Growth theory and insert new data. Inserting new data consists of transaction_ID, transaction_Date, customer name and sculptures category. Sculptures category contains elephant, peacock, dragon, horse, ankone, lawkanaut, women, etc. This data can save a customerTable contain a personnel database. In this way, user can insert next new data. Observer works running FP-Growth theory as shown in figure 2.

Before admin and observer users are running the FP-Growth theory, user is requesting the date. If user can request date, system can be detected in personnel database and discard date and saving a file. And then system is loading a save file and building the FP-tree and generate the FP-tree. Finally result is generating the mining frequent pattern as shown in figure 3. The proposed system can be served mining frequent pattern by using FP-Growth theory and insert new data. For example, customerTable contains transaction Date, transaction ID, customer name, sculptures category.

Table 1: example of customerTable

transaction Date	transaction ID	customer name	sculptures category
1/1/2008	1	Su Su	elepahnt,peacock
1/1/2008	2	Smith	elephant,ankone
1/2/2008	3	Marry	elephant,ankone
1/2/2008	4	James	ankone,women
1/3/2008	5	David	elaphant,dragon
1/4/2008	6	Helen	elepahnt,peacock
1/4/2008	7	Ki Ki	elepahnt,peacock
1/5/2008	8	Si Si	elephant,ankone
1/6/2008	9	Yuki	elaphant,dragon
1/7/2008	10	Marli	elephant,ankone

User is requesting date from 1/1/2008 to 1/3/2008. Data can be detected customerTable in personnel database and discard data and save a file. System can be loaded a save file and running the FP-Growth theory. Firstly, building FP-tree and generating the FP-tree. And then generates mining frequent pattern is {elephant, ankone}.

In proposed system, user input is request date. System can run the FP-Growth theory. System output is mining frequent pattern that are sculptures category. Result output can be changed in user request date.

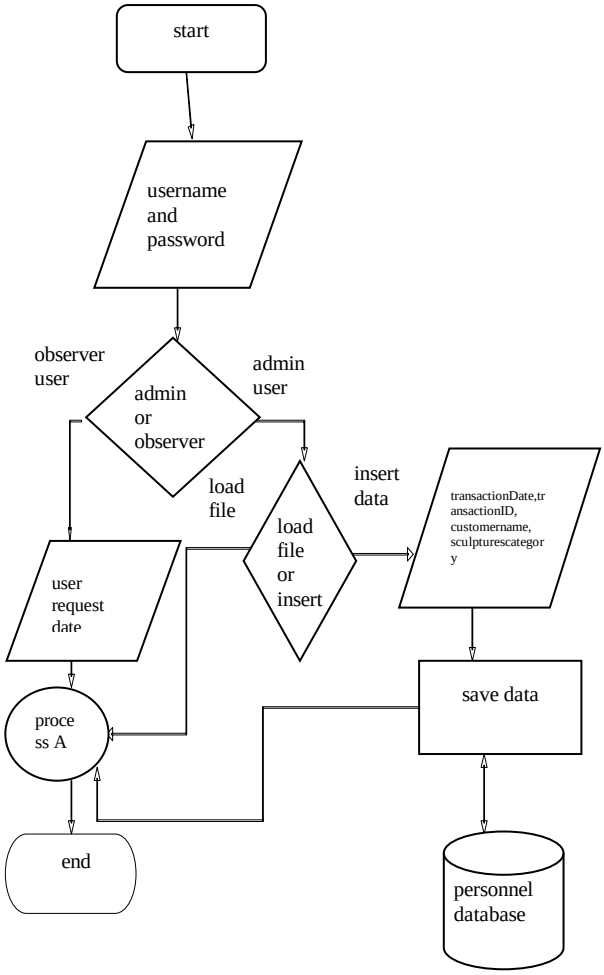

Figure 2: Process flow of proposed system

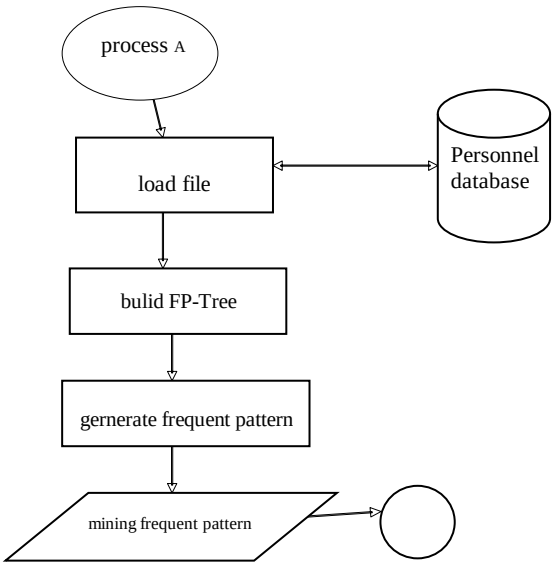

Figure 3: Process flow of proposed system for Process A

4. Related Work

Paper described for the development of an efficient frequent pattern mining system. In this study, we propose a frequent - pattern tree (FP - tree) structure, which is an extended prefix - tree structure for storing compressed, crucial information about frequent patterns, and develop an efficient FP - tree - based mining method, FP - growth, for mining the complete set of frequent patterns by pattern fragment growth. Efficiency of mining is achieved with three techniques: (1) a large database is compressed into a condensed, smaller data structure, FP - tree which avoids costly, repeated database scans, (2) our FP - tree - based mining adopts a pattern - fragment growth method to avoid the costly generation of a large number of candidate sets, and (3) a partitioning - based, divide - and - conquer method is used to decompose the mining task into a set of smaller tasks for mining confined patterns in conditional databases, which dramatically reduces the search space . Our performance study shows that the FP - growth method is efficient and scalable for mining both long and short frequent patterns, and is about an order of magnitude faster than the Apriori algorithm and also faster than some recently reported new frequent - pattern mining methods [4].

In this study, we propose a frequent - pattern tree (FP - tree) structure it uses compact data structure. Eliminate repeated database scan. Basic operation is counting and FP-Tree building. FP-Tree is much smaller than original FP-Tree. Conditional FP-Tree is smaller than pattern base. Mining process works on a set of usually much smaller pattern base and confidence FP-Tree. Divide and conquer and dramatic scale of shrinking [5].

5. Advantage of Proposed System

Myanmar sculptures are more knowing and attraction the tourisms. Users know popular things in Myanmar sculptures and Myanmar sculptures are interest in many tourists. Users are more studying the Myanmar sculptures distribution and knowing which items are frequently purchased together.

6. Conclusion

This paper proposed the frequent pattern mining for Myanmar sculptures gallery system. The proposed system is to provide Myanmar sculptures that are attracting the tourisms and users know popular things in Myanmar sculptures. For instance, if customers are buying an elephant, how likely they to also by ankone on the same trip to Myanmar sculptures gallery. Such information can lead to increased sales by helping retailers do selective marketing and plan their self space.

References

[1] A. I. Verkamo, H. Mannila, H. Toivonen M. Klemettinen and P. Ronkainen, “Finding interesting rules from large sets of discovered association rules ”,CIKM’94, 401-408, Gaitherbur, Maryland.

[2] A. Swami, R. Agrawal and T. Imielienski, “Mining Association Rules between Sets of Items in Large Databases”. Proc. Conf. on Management of Data, 207-216. ACM Press, New York, NY, USA 1993

[3] J. Han, J. Jewie and R. Mao, “Mining frequent pattern without candidate generation “, 2000

[4] J .Han, J. Pei, “Can we push more constraints into frequent pattern mining?” KDD’00. Boston, MA. Aug.2000

[5] J. Han, J. Pei and Y. Yin, “Mining frequent pattern without candidate generation”, proceeding ACM-SIGMOD International conference on Management of Data ACM Press, 2000, pp 1-12

[6] J. Han, J. Pei and Y. Yin, “Mining frequent pattern without candidate generation”, School of Computer Science Simon Fraser University, March 18th 2009

[7] J. Han, J. Pei, Y. Yin and R. Mao, “Data Mining and Knowledge Discovery”, 2004

[8] J. Han, M. Kamber, “Data Mining: Concepts and Techniques”, Academic Press, USA, 2001

[9] J. Han, M. Kamber, “Intelligent Database System Research Lab”, School of Computing Science Simon Fraser University, Canada http://www.cs.stu.ca