

Applying Split and Merge Algorithm in Customer Relationship Management

G. SHIVAKUMAR (M.Tech)

Software Engineering
ATRI, Uppal Hyderabad, India

G. VARA LAKSHMI

Sr. Associate Professor
Department of CSE
ATRI, Uppal Hyderabad, India

ABSTRACT---Recently, customer relationship management (CRM) has become the core of growth of the company. Data mining, as a powerful data analysis tool, extracts critical information supporting the company to make better decisions by processing a large number of data in commercial databases. This paper introduced the basic concepts of data mining and CRM, and described the process how to use data mining for CRM. This paper presents SaM, a split and merge algorithm for frequent item set mining. Its core advantages are its extremely simple data structure and processing scheme, which not only make it very easy to implement. At last, the paper described the applications of several main data mining methods in CRM.

Keywords---Data mining, frequent item set mining, customer relationship management; classification, association rule.

1. INTRODUCTION

Frequent item set mining plays an important role in data mining and knowledge discovery techniques such as association rule mining, classification, clustering, graph mining, web mining etc[1]. Frequent item set mining is the following task: we are given a set of *items*, called the *item base*, and a database *T* of Transactions. for example, represent a product, and the item base may then represent the set of all products offered by a supermarket. The term *item set* refers to any subset of the item base *B*. Each transaction is an item set and may represent, in the supermarket a set of products that has been bought by a customer.

Recently, with the development of computer technology and IT (Information Technology), human-beings are entering into the information era, in which we are in dire need of some new tools and technologies to deal with such “mass” data, picking up useful information automatically from these data to solve the “data plenty, useful spare” problem “information explosion” takes to us. Therefore, how to dig up the valuable information in the data sea and apply this information into the practical work becomes the problem needs to be solved

urgently, which leads to the emersion of data mining technology .More and more enterprises start to focus on a new marketing management theory—customer relationship management (CRM), which places the retaining and upgrading of customer value as the core of enterprise organic development. Data Mining is just the very tool suitable to solve this problem.

2. DATA MINING

Generally, 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 costs, or both. Data mining software is one of a number of analytical tools for analyzing data. It allows users 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 of fields in large relational databases.

Data mining is primarily used today by companies with a strong consumer focus - retail, financial, communication, and marketing organizations. It enables these companies to determine relationships among "internal" factors such as price, product positioning, or staff skills, and "external" factors such as economic indicators, competition, and customer demographics. An effective Customer Relationship Management solution can ensure your marketing and sales teams have customer information that they can easily use to significantly grow the value of your customer base – typically[8].

Data mining theory and tool have already been adopted in every walk of life, and developed pretty good in some areas as engineering development, scientific research and commercial management in mobile, bank, telecom, insurance, retail, etc. CRM is becoming the focus of enterprises, and the main content of this thesis is just research on the application of data mining in customer relationship management.

3. FREQUENT ITEMSET MINING

Frequent item set mining is the term *item set* refers to any subset of the item base $B[2]$. Each transaction is an item set and may represent, in the supermarket setting, a set of products that has been bought by a customer. Note that the item base B is usually not given explicitly, but only implicitly as the union of all transactions. The *support* $sT(I)$ of an item set $I \subseteq B$ is the number of transactions in the database T it is contained in. Given a user-specified *minimum support* $smin \in \mathbb{N}$, an item set I is called *frequent* (in T) iff $sT(I) \geq smin$. The goal of frequent item set mining is to find all item sets $I \subseteq B$ that are frequent in the database T and thus, in the supermarket setting, to identify all sets of products that are frequently bought together.

3.1 Apriori algorithm

One of the most popular data mining approaches is to find frequent itemsets from a transaction dataset and derive association rules. Finding frequent itemsets (itemsets with frequency larger than or equal to a user specified minimum support) is not trivial because of its combinatorial explosion. Once frequent itemsets are obtained, it is straightforward to generate association rules with confidence larger than or equal to a user specified minimum confidence.

```

 $F_1 = \{ \text{Frequent itemsets of cardinality } 1 \};$ 
for( $k = 1; F_k \neq \phi; k++$ ) do begin
     $C_{k+1} = \text{apriori-gen}(F_k); // \text{New candidates}$ 
    for all transactions  $t \in \text{Database}$  do begin
         $C'_t = \text{subset}(C_{k+1}, t); // \text{Candidates contained in } t$ 
        for all candidate  $c \in C'_t$  do
             $c.count++$ ;
        end
         $F_{k+1} = \{ C \in C_{k+1} \mid c.count \geq \text{minimum support} \}$ 
    end
end

```

Frequent item set mining is a data analysis method that was originally developed for market basket analysis and that aims at finding regularities in the shopping behavior of the customers of supermarkets, mail-order companies and online shops.

Example of Transactional Database.

Fig 1. Example of a transaction database

Item id	Milk	Bread	Butter	Sugar	Ghee
T1	0	1	1	1	0
T2	1	1	0	0	1
T3	0	0	1	0	1
T4	1	0	1	1	0
T5	1	1	1	0	0
T6	1	1	0	1	1

From the above transactional database this frequent item set By consider by Using this formula:

Support: The rule holds with support *sup* in T (the transaction data set) if $\text{sup}\%$ of transactions contain $X \cup Y$.

$$\text{sup} = \text{Pr}(X \cup Y).$$

Confidence: The rule holds in T with confidence *conf* if $\text{conf}\%$ of transactions that contain X also contain Y .

$$\text{conf} = \text{Pr}(Y | X)$$

Items Id	Frequents Items
T1	Bread, Butter, sugar
T2	Milk, Bread, Ghee
T3	Butter, Ghee
T4	Milk, Butter, Sugar
T5	Milk, Bread, Butter
T6	Milk, Bread, Sugar, Ghee

Fig 2: Frequent Item set

3.2 Split and Merge Algorithm

In this paragraph we describe the basic form of our SaM (split and merge) algorithm [4] for generating association rules. Preprocessing is very similar to many other frequent item set mining algorithms. The steps are illustrated in Figure 3 for a simple example transaction database:

step 1 shows the transaction database in its original form. In step 2 the item frequencies are determined in order to discard infrequent items. With a minimum support of 3, items f and g are infrequent and thus eliminated. In step 3 the (frequent) items in each transaction are sorted according to their frequency, because processing the items in the order of increasing frequency usually leads to the shortest execution times. In step 4 the transactions are sorted lexicographically into descending order, with an item with higher frequency preceding an item with lower frequency. In step 5 the basic data structure is built by combining equal transactions and setting up an array, in which each element consists of two fields: an occurrence counter and a pointer to the sorted transaction. The conditional database for frequent item sets *not* containing this item is obtained with a simple *merge step* (right part of Figure 4). The new array and the rest of the original array are combined with a procedure that is almost identical to one phase of the well-known merge sort algorithm.

Fig 3: 1.Original form 2. item frequencies 3. transactions with sorted items 4. Lexicographically sorted transactions 5. used data structure

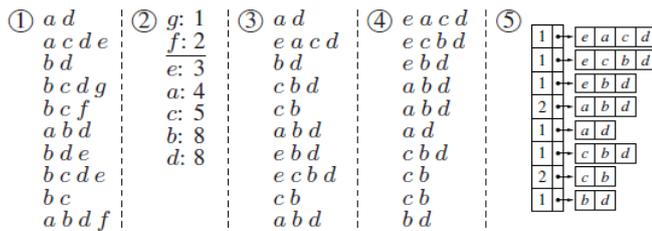
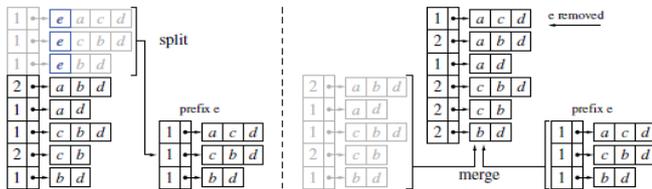


Fig 4: split (left) and merge (right).



4. OBJECTIVE

CRM, the technology [1], along with human resources of the company, enables the company to analyze the behavior of customers and their value. The main areas of focus are as the name suggests: customer, relationship, and the management of relationship and the main objectives to implement CRM in the business strategy. Understand customer needs better, Maintain long-term customer relationships

5. EXPERIMENTAL RESULTS

To evaluate the performance of our proposed Split and Merge algorithm these results were obtained.

Association Rules:

- rule 0: MILK ==> BUTTER
- rule 1: BUTTER ==> SUGAR
- rule 2: SUGAR ==> RICE
- rule 3: SUGAR ==> MILK
- rule 4: MILK ==> SUGAR
- rule 5: GHEE ==> SUGAR
- rule 9: MILK ==> BREAD
- rule 10: MILK ==> GHEE

6. CONCLUSION

In modern society, customer becomes the most important asset of enterprise. Efficient customer relationship management is the necessary method to improve enterprise competition advantage. By applying split and merge algorithm can get better result. At present, the local researches on CRM mainly focus on CRM system structure research, CRM value evaluation and calculation [1], CRM management modes and application of data mining and mathematical statistics methods in CRM and other aspect.

7 ACKNOWLEDGMENT

I express my earnest and deep sense of gratitude to Mrs. D.Sujatha, Head of the Department, Computer Science & Engineering without Her valuable suggestions and guidance, preparation of this Paper wouldn't have been possible. It was a great pleasure and honor to work under Her and I hope many more fruitful associations in the years to come. Her kindness and affection will remain with me forever. I also owe the gratitude towards T. Naga lakshmi Lecturer, Coordinator for me. And I want to thank my Internal guide G. Vara lakshmi Sr. Associate Professor and other faculty members of ATRI college, for their noteworthy suggestions and encouragement given to me in completing the paper. I would also like to thanks the members of PRC for their help, through provoking discussions invigorating suggestions extended to me with immense care, zeal throughout the work.

8. REFERENCES

- [1] Yu-feng Jia and Ying-xin Hu, "Research on Application of Data Mining in Customer Relationship Management", Journal of Shijiazhuang Railway Institute(Social Science), Vol.3, Jun.2009, pp. 31-34.
- [2] Bin Cao and Yue-xin Wu, "Data Mining Technology Application in the Customer Relationship Management System", <http://www.paper.edu.cn>, 2008.
- [3] G.I. Webb and S. Zhang. *k-Optimal-Rule-Discovery*. *Data Mining and Knowledge Discovery* 10(1):39–79. Springer, Amsterdam, Netherlands 2005
- [4] C. Borgelt. SaM: Simple Algorithms for Frequent Item Set Mining. IFSA/EUSFLAT 2009 conference- 2009.
- [5] Y. Cheng, U. Fayyad, and P.S. Bradley. Efficient Discovery of Error-Tolerant Frequent Itemsets in High Dimensions. Proc. 7th Int. Conf. on Knowledge Discovery and Data Mining (KDD'01, San Francisco, CA), 194–203. ACM Press, New York, NY, USA 2001
- [6] J. Han, and M. Kamber, 2000. *Data Mining Concepts and Techniques*. Morgan Kanufmann.
- [7] R. Agrawal, T. Imieliński, and A. Swami. 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.
- [8] Bin Cao and Yue-xin Wu, "Data Mining Technology Application in the Customer Relationship Management System", <http://www.paper.edu.cn>, 2008.