

# Mining E-shopper's Purchase Rules by Using Maximal Frequent Patterns: An E-commerce Perspective

Md. Rezaul Karim, Jae-Hyun Jo, Byeong-Soo Jeong, and Ho-Jin Choi<sup>1</sup>

Department of Computer Engineering, Kyung Hee University, Korea

<sup>1</sup>Department of Computer Science, Korea Advanced Institute of Science and Technology (KAIST), Korea

E-mail: {whwo86, asif\_karim, jeong} @khu.ac.kr; hojinc@kaist.ac.kr

**Abstract**—Market basket analysis is very important to everyday's business decision, because it seeks to find relationships between purchased items. Undoubtedly, these techniques can extract customer's purchase rules by discovering what items they are buying frequently and together. Therefore, to raise the probability of purchasing the corporate manager of a shop can place the associated items at the neighboring shelf. For these reasons, the ability to predict e-shopper's purchase rules basing on data mining has become a competitive advantage for the company. On the other hand, mining maximal frequent patterns are also a key issue to the recent market analysis since; a maximal frequent pattern for a particular customer reveals the purchase rules. Moreover, if the dataset is sparse due to the presence of null transactions, the mining performance degrades drastically in existing approaches. In this paper, first we remove null transactions from the original dataset then we apply the bottom-up row enumeration tree approach to generate the maximal frequent patterns; later on the modified version of the sequence close level is used for counting the distance between a pair of items for mining the customer's purchase rules in an online transactional database. Experimental results show that our proposed approach is superior to previous approaches and can predict more accurate customer's purchase rules in reasonable time.

**Keywords**- Market basket analysis; e-shopper's purchase rules; sequence close level; maximal frequent patterns; transactional database.

## I. INTRODUCTION

Online shopping is the process whereby consumers directly buy goods or services from a seller in real-time, without an intermediary service, over the Internet. It is a form of electronic commerce. An online shop, e-shop, e-store, internet shop, web-shop, web store, online store, or virtual store evokes the physical analogy of purchasing products or services at a bricks-and-mortar retailer or in a shopping centre. The process is called Business-to-Consumer (B2C) online shopping [21]. When a business buys from another business it is called Business-to-Business (B2B) online shopping.

Data mining is crucial for extracting and identifying useful information from a large amount of data [2-5]. E-retailing companies operate purchase databases in a longitudinal way, such that all transactions are stored in chronological order. A record-of-transaction database typically contains the transaction date and the products bought in the course of a given transaction. Usually, each record also contains an e-shopper ID, particularly when the purchase was made using a credit card or a frequent-buyer card. Therefore, the purchasing sequence of an e-shopper in the database that has made repeated purchase

can easily be determined. This purchase sequence provides a description of the changes in an e-shopper's preferences over time [27]. Because, a purchase sequence can reveal the changes of e-shopper's preferences over time.

So far, a good number of efficient methods for extracting frequent patterns have been proposed [6-8]. The many algorithms are utilized to mine frequent patterns or frequent Patterns in time-based data. Frequent pattern mining has already become one of the widely applied Techniques in data mining. However, in the present domain of knowledge, there has been little study of mining the online customers' behavior based on purchase sequences. To predict e-shoppers' purchase behavior changes, an organization's management first identifies target e-shoppers who share similar preferences for products and looks for those products that target e-shoppers are most likely to purchase. The purchase transaction records of e-shopper are used to build e-shoppers' profile describing his or her likes and dislikes. A set of e-shoppers known as neighbors who have exhibited similar behavior in the past, can be found through calculating the correlations among e-shoppers [27].

In this paper, first we remove null transactions from the original dataset then apply the bottom-up row enumeration tree approach to generate the complete set of maximal frequent patterns. Later on, for mining customers purchase rules in an online transactional database we remove the limitations of existing approach of sequence close level.

The rest of this paper is organized as follows: Section II analyzes previous works. Section III describes some basic concepts of sequential pattern and the modified version sequence close level  $CL(T^k)$  and analyzes relationships between items in a transaction. Section IV proposes a modified algorithm for extracting the maximal frequent pattern which will be used to mine the purchase rules. Section IV illustrates an example for extracting online customers' purchase rules with maximal frequent patterns and sequence close level. Section V shows experimental results and we conclude at section VI.

## II. RELATED WORKS

### A. Related Works

At the present time, there are usually are two principal approaches to be used to extract frequent patterns, methods based on association rule mining [6-10] and techniques based on using regular structures to denote navigation patterns, such as path, trip, tree, and graph structures. Several well known approaches for mining association rules [12-14] have been modified and applied in exploring frequent patterns for recent years. Furthermore, web mining is a primary application for

frequent pattern mining [15, 16]. Web mining extracts frequent patterns in analyzing users surfing behaviors. Additionally, a substantial amount of research has investigated approaches [17, 18] that can identify frequently occurring frequent patterns in extremely complex structures, such as Tree and graph. The researchers are developing for mining tree structures, whereas other researches focused on exploiting graph patterns. A tree structure, which is a special hierarchical structure, can be widely applied to mine in many domains, such as Customer relationship management, purchasing behavior analysis, and knowledge discovery [19, 20]. Frequently occurring frequent patterns, identified in a large database, comprise lists of subsequences.

Literature [28], addressed the problem of mining e-shopper's purchase rules based on k-trees sequential patterns by finding maximal frequent patterns. The idea goes as follows: first it constructs k-tree sequential patterns using Apriori based algorithm and then they mine the maximal frequent patterns and finally mine purchase rules based on the calculation of sequence close level for each maximal frequent patterns in every transaction. But this works has two problems: firstly, it has no proper bound due the incorrectness of the proposed sequence close level formula and secondly it leads a tedious problem due to multiple times scanning of the original database. For example, if the number of maximal frequent patterns increases; so increases the number of scan to the database scan even if the database could be held into the main memory.

#### B. Motivations and the Screening of Null Transactions

A null transaction is a transaction that does not contain any item-sets being examined. Typically, the number of null-transactions can outweigh the number of individual purchases because, for example, many people may buy neither milk nor coffee, if these itemsets are assumed to be two of the frequent itemsets. So, it is highly desirable to have a measure that has a value that is independent of the number of null-transactions. A measure is null-invariant if its value is free from the influence of null-transactions. Null-invariance is an important property for measuring association patterns in large transaction databases [9].

Although, lots of works have proposed and developed [17-20, 28], but, if the transactional database is sparse due to the presence of 'null transactions', mining performance degrades drastically. Unfortunately, above mentioned works did not have the null-invariance property. Since, large data sets typically have many null-transactions, it is important to consider the null-invariance property when selecting appropriate interestingness measures for pattern evaluation.

In this proposed approach an attempt has been made to eliminate the null transactions thereby, attempting to reduce the processing time for finding frequent  $k$ -itemsets. Finding null transactions and later eliminating them from future scheme of things is the initial part of this proposed framework. Consider for instance that, an electronic shop has 100 transactions of which, 40% are null transactions. FP-tree method of mining or any other related method in that case would scan all the 100 transactions while, our proposed approach attempts to reduce the transactions to 60% by considering just the valid 60 transactions after

screening the 40 null transactions. This saves a lot of precious computation time [9]. Besides, an attempt has been made to find null transactions by using vertical data layout format [27]. It is quite possible to find the null transactions by finding those transactions that don't appear against any frequent single-itemset with this representation.

Therefore, to avoid lots of sequential access to the original database, first we remove so called 'null transactions' then used the bottom-up row enumeration tree approach [29] to mine maximal frequent patterns.

If row/transaction IDs are saved along with the support of individual maximal frequent pattern, lots of sequential access through the database can be avoided substantially and thus faster the overall mining process. Finally, we remove the limitations of the concept of sequence close level for counting the distance between a pair of items for mining the customer's purchase rules.

### III. PROBLEM STATEMENT

In this section first we present some preliminary knowledge related to the problem of mining maximal frequent patterns. After that, we describe the process of mining customer's purchase rules by removing the limitations of existing approach of counting sequence close level based on the maximal frequent patterns found.

#### A. Mining Maximal Frequent Pattern

Let a set of distinct items  $I = \{i_1, i_2, \dots, i_n\}$  and  $n$  is the number of distinct items. Transaction  $t$  is an ordered list of items, denoted as  $t = [i_1, i_2, \dots, i_m]$  where  $i_1 < i_2 < \dots < i_m$  and  $m \leq n$ . A transactional database  $T = \{t_1, t_2, \dots, t_N\}$  is a set of  $N$  transactions and  $|N|$  is the number of total transactions. A set  $X \subseteq I$  is called a pattern. If  $X \subseteq t$ , it is said that  $X$  occurs in  $t$  or  $t$  contains  $X$ .  $Support(X)$  denotes the percentage of transactions that contain  $X$ . If  $Support(X) \geq min\_sup$  we say that  $X$  is a frequent pattern. If  $X$  is a frequent pattern and no superset of  $X$  is frequent, we say that  $X$  is a maximal frequent pattern.

TABLE I A TRANSACTIONAL DATABASE

TID	Itemset (Sequence of items)
10	A, B, C, F
20	C, D, E
30	A, C, E, D
40	A
50	D, E, G
60	B, D
70	B
80	A, E, C
90	A, C, D
100	B, E, D

So, given a transactional database, now the problem of mining maximal frequent patterns is to find the complete set of maximal frequent patterns whose support is no less than a user-given  $min\_sup$  threshold. For example, in table III, the occurrences of patterns "CD", "DE" and "CDE" are 3, 3 and 2 respectively. If the  $min\_sup$  is 2, all of these are frequent patterns. But, "CD" is not a maximal frequent pattern because its super pattern "CDE" is also a frequent pattern.

#### B. Sequence Close Level

In this subsection we present the sequence close level introduced at literatures [2, 4, 26, 28]. Items purchased by

online customers indicate a frequent pattern. Therefore, frequent pattern mining is widely used in analyzing e-shopper's behavior by mining frequent pattern [27, 28]. According to e-shopper's purchase items, a record for transactional data usually is depicted as tree-like sequence pattern [28]. In addition, most of the counting support approaches count only the frequency of occurring patterns and do not address the distance between a pair of items in a transactional database [28].

The length between two items in a transaction is called distance. Distance of the same pair of items in different sequences varies. For example, consider the filtered database presented in table III; in transaction 10; item C and D are neighbors, however C and D are far from each other in transaction 20. The distance between two items in a sequence assists in determining the closeness of their relationships. If two items are far from each other, their relationship is loose.

On the contrary, their relationship is close if two items are neighbors. Therefore, it is necessary for sequential pattern mining to calculate the distance between two items in a sequence in addition to counting the frequency of occurring pattern. Before describing the significance of distance between items, distance is first defined. If two items are neighbors, the distance between them is set to be 1. To estimate the significance of the distance for sequence, sequence close level is defined in [28] as follows:

$$CL(T^k) = \frac{\sum_{i=1}^{k-1} \left( \frac{1}{d_i} \right)}{k-1} \quad (1)$$

Here  $k$  is the length of  $k$ -maximal frequent pattern  $T^k$ ;  $d_i$  is the distance between two items, and  $d_i = p(t_{i+1}) - p(t_i)$  where  $p(t_i)$  is the position of  $i$ th item in transaction  $t_i$ . The value of  $CL(T^k)$  is between 0 and 1; but when  $CL(T^k)$  is equal to 1, all items in maximal frequent patterns in a transaction are close to each other and the sequence is very important. If the value is close to 0, the sequence has lower significance. When  $CL(T^k)$  is equal to 0, the sequence does not contain any useful information about the customer's purchase rules or purchase history.

**Example 1:** For the filtered database presented in table 3, pattern  $\{A, C, D\}$  is a maximal frequent pattern occurs in TID 30 and 90 and it is also a 3-patterns since it contains two items, so sequence close level  $CL(T^k)$  in terms of TID 30 and 90 can be calculated by formula (1) as follows: The distances from A to C in TID 30 is  $d_1=1$  and from C to D is 2; so we have,

$$CL(T^2) = (1/1) + (1/2) / (3-1) = 0.75$$

On the other hand in terms of TID 90 distances from A to C is  $d_1=1$  and C to D  $d_2=1$  so we have,

$$CL(T^2) = (1/1) + (1/1) / (2-1) = 1$$

So, the maximal frequent pattern  $\{A, C, D\}$  does not contain equal importance in terms of sequence close level, that is a customer is more likely to purchase items A, C and D in TID 90 but an online customer first buy item A, C then another item E and item D in transaction 30.

But this measurement has one limitation. For example in table III, pattern  $\{A, C, E\}$  is a maximal frequent

pattern that occurred at TID 30 and 80. How do we calculate the sequence close level of  $\{A, C, E\}$  in terms of TID 80? Definitely the sequence close level for the pattern  $\{A, C, E\}$  for TID 30 and 80 should be equal, because A, C and E occurred consecutively. Therefore it implies that we need to remove this limitation. We proposed the modification of the distance calculation between items in a transaction. So, if using the occurrence of items in a maximal frequent pattern is consecutive then the distance can be calculated as follows:

$$d_i = |p(t_i) - p(t_{i+1})|$$

So, in both case we can use formula (1) for the calculation of sequence close level even if the occurrence of items in a maximal frequent pattern is consecutive.

**Example 2:** Pattern  $\{A, C, E\}$  is also a maximal frequent pattern occurs in TID 30 and 80 and it is also a 3-patterns since it contains three items, so the sequence close level  $CL(T^k)$  of TID 30 and 80 can be calculated by formula (1) as follows: The distances from A to C is  $d_1 = 1$  in TID 30 and from C to E is  $d_2 = 1$  respectively, so

$$CL(T^3) = (1/1) + (1/1) / (3-1) = 1$$

On the other hand in TID 80 distances from A to C is  $d_1 = 2$  and from C to E is  $d_2 = 1$  respectively, so

$$CL(T^3) = (1/1) + (1/1) / (3-1) = 1$$

Using similar way we can calculate the sequence close level for maximal frequent pattern  $\{C, D, E\}$  as well. So, maximal frequent pattern  $\{A, C, E\}$  and  $\{A, C, D\}$  does not contain equal importance in all the transactions in terms of sequence close level, for example in terms of TID 30 pattern  $\{A, C, E\}$  is more important than pattern  $\{A, C, D\}$ . Therefore, from the above example it is clear that the bigger value of  $CL(T^k)$  is the more desirably online customers purchase the items in transaction  $T^k$ .

#### IV. PROPOSED APPROACH

We use the bottom-up row enumeration tree approach to insert the frequent 1-itemset, 2-itemset and so on in bottom-up fashion. After that, complete set of maximal frequent patterns can be mine from this tree.

To avoid lots of sequential access to the original database, our approach goes as follows: (i) During first database scan we remove null transactions, after representing the dataset in vertical layout format from the database, since a null transaction has no contribution to an association rule, sequence close level or correlated patterns. (ii) In second scan to the filtered database, we insert frequent 1-itemsets into the bottom-up row enumeration tree; in addition to support at every node, we keep track the transaction ID sets; where, a frequent patterns occurs (iv) we mine frequent patterns and later on maximal frequent patterns are generated from that tree.

We apply pruning in every level of the tree, so it is guaranteed less number of redundant candidate patterns to be generated.

Table II VERTICAL LAYOUT FORMAT OF THE EXAMPLE DATABASE

Items	TID Sets
A	10, 30, 80, 90
B	60, 100
C	20, 30, 80, 90
D	20, 30, 50, 60, 90, 100
E	20, 30, 50, 80, 100
F	10
G	50

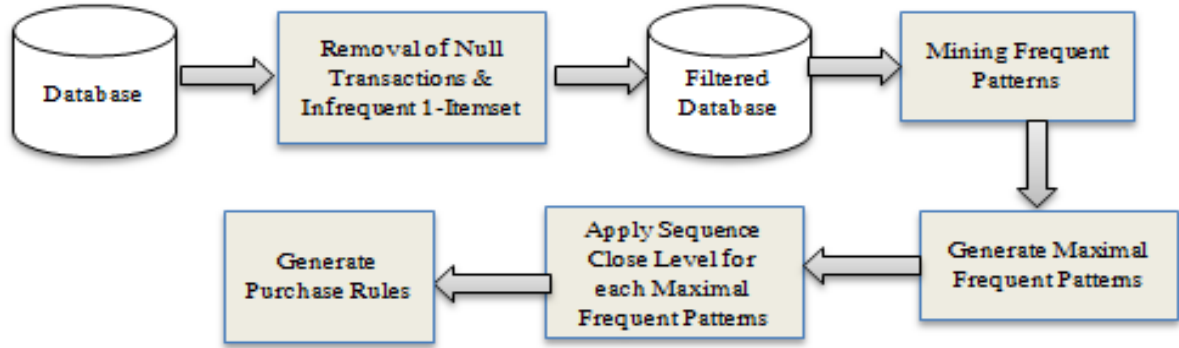


Figure 1. Workflow of our proposed approach for mining purchase rules.

#### Algorithm for Mining E-shopper's Purchase Rules

**Input:** i) A transactional database ii) minimum support threshold 'min\_sup' and (iii) minimum sequence close level 'min\_seq\_close\_lev'.

**Output:** i) Complete set of maximal frequent patterns which satisfy the minimum support threshold 'min\_sup' and ii) Complete set of purchase rules which satisfy the minimum sequence close level 'min\_seq\_close\_lev'.

**Step-1:** Scan the database and represent it using vertical layout format, then remove all the 'null transactions'.

**Step-2:** Scan the filtered database after removal of null transactions and constructs the complete pruned bottom-up row enumeration tree and find out complete set of frequent patterns from that tree. Suppose P is a candidate itemset in that tree:

1. If  $\text{Sup}(P) < \text{min\_sup}$   
Remove P from the tree. // Level 1 pruning
2. Else P is a frequent pattern.

**Step-3:** Check the maximality criterion and find out the complete set of maximal frequent patterns.

**Step-4:** Scan the transactional database for second time and calculate the sequence close level for the maximal frequent patterns found in step 2 for the TIDs contained by the maximal frequent patterns only. Suppose X is a maximal frequent k-pattern, then.

1. If  $(\text{CL}(T_x^k) > \text{min\_seq\_close\_lev})$   
X is an important purchase rule
2. Else //// Level 2 pruning  
X is not an important purchase rule

Figure 2. Algorithm mining E-shopper's purchase rules

#### V. AN ILLUSTRATIVE EXAMPLE

Online customers 'purchase rules can be discovered by mining maximal frequent pattern, since the maximal frequent patterns reveal online customers' purchase rules [26, 28]. Now, consider the vertical layout format in table II for the same database given in table I. Here, transaction 10 and 50 are supposed to be the 'null transactions'; also from table I transactions 40 and 70 are null transactions and are already not been considered for mining.

Since, null transactions containing just 1-itemsets are not significant since these itemsets do not have contribution while mining correlated patterns or association rule mining, hence, have been removed prior to mining. Also, itemsets F and G do not satisfy the minimum support count of 2, and is hence, avoided for mining. For the ease of the reader we just showed the corresponding filtered transactional database in table III. From table III, we can see that the support of the frequent 1-itemset A, B, C, D and E are 3, 2, 4, 5 and 4 respectively. Now we construct the tree with bottom-up row enumeration approach. Fig-3 shows the complete pruned row-enumeration tree with support and TID sets.

TABLE III THE FILTERED DATABASE

TID	Itemset
20	C, D, E
30	A, C, E, D
60	B, D
80	A, E, C
90	A, C, D
100	B, E, D

TABLE IV FREQUENT PATTERNS AND THEIR SUPPORT FROM THE TREE

Frequent Pattern	Support	Frequent Pattern	Support
AC	3	CE	3
AD	2	DE	3
AE	2	ACD	2
BD	2	ACE	2
CD	3	CDE	2

From the tree presented by figure 3, we have generated frequent patterns with support and TID sets and shown in table IV. Later on four maximal frequent patterns are generated from table IV as follows: BD, ACD, ACE and CDE and each with support 2. From the row enumeration tree we find TID set where maximal frequent patterns are occurred. For example, pattern BD occurs at TIDs 60, 100; pattern ACE occurs at TIDs 30, 80; pattern ACD occurs at TIDs 30, 90 and pattern CDE occurs at TIDs 20, 30 respectively. Therefore, to understand the importance of the maximal frequent patterns to transaction records, calculation of  $CL(T^k)$  is made for the maximal frequent patterns found.

Now suppose the minimum sequence close level is 0.60. The results of  $CL(T^k)$  indicate that for maximal frequent patterns (shown in table V), the bigger value of  $CL(T^k)$  is, the closer all items in the sequence are each other, and the more desirably online customers purchase the items according to k-trees frequent pattern [28]. For example, purchase rules or purchase sequence like {A, C, E}, {A, C, D} in terms of TID 30; sequence or rules {B, D} in terms of TID 60 are very important purchase rules. On the other hand {A, C, D} in terms of 90; {A, C, E} in terms of TID 80; and {C, D, E} in terms of TID 30 are also important but with less significant rules compared to others. But purchase rules like {B, D} is not an important rules or purchase sequence.

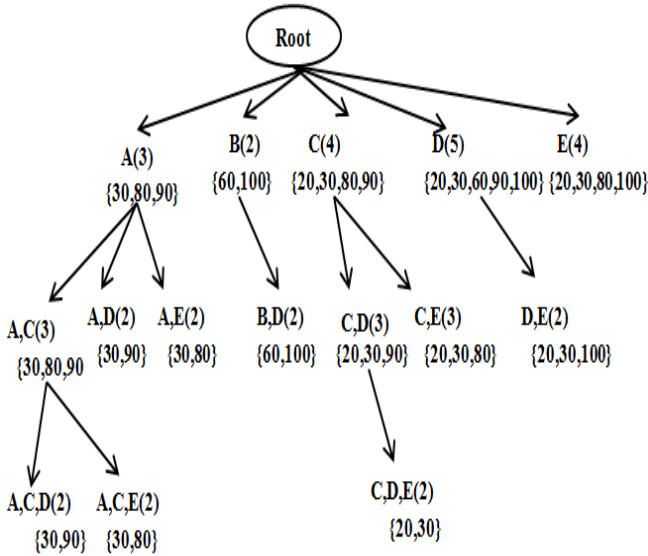


Figure 3. Complete row enumeration tree (pruned).

TABLE V  
CALCULATION OF  $CL(T^k)$  FOR MAXIMAL FREQUENT PATTERNS

FP	$CL(T^k)$	TID	Itemset
{B, D}	1	60	B, D
	0.5	100	B, E, D
{A, C, D}	1	30	A, C, E, D
	0.75	90	A, C, D
{A, C, E}	1	30	A, C, E, D
	0.75	80	A, E, C
{C, D, E}	1	20	C, D, E
	0.75	30	A, C, E, D

## VI. EXPERIMENTAL RESULTS

### A. Programming Environment

All programs we have written in Microsoft Visual C++ 6.0 running on Windows XP. And the Hardware configuration is as follows: Processor-Intel Core 2 Duo 2.4GHz, Main memory-4GB, Hard disk space- 500GB.

### B. Description of the Datasets

We apply our algorithm on mushroom and connect-4 datasets. Datasets 'mushroom' and 'connect-4' were downloaded from <http://archive.ics.uci.edu/ml/website>. The mushroom datasets contain 8124 transactions and the average length is 23; on the other hand connect-4 dataset contains 135,115 transactions with of lengths 8 to 35. Size of mushroom and connect-4 datasets is 5 MB and 20 MB respectively.

### C. Performance Analysis

We set `min_seq_close_lev` to 0.65. We compared our result with the algorithm proposed [28] *et al.* for this purpose we modified the algorithm of maximal itemset mining [22, 29] according to our algorithm and in graphs we named our approach as 'Proposed Approach' and the approach proposed at [28], as 'Latest Approach'. In the first experiment we compared the running time of our approach and the latest approach. Fig 4(a) shows the running time on mushroom and fig 4(b) shows the running time on connect-4 datasets. Connect-4 dataset takes more time because of more transactions and variable lengths of the transactions in the datasets.

In the second experiment we compared the memory usage of our approach to the latest approach. We can see that our approach consumes very low memory compared to the latest approach.

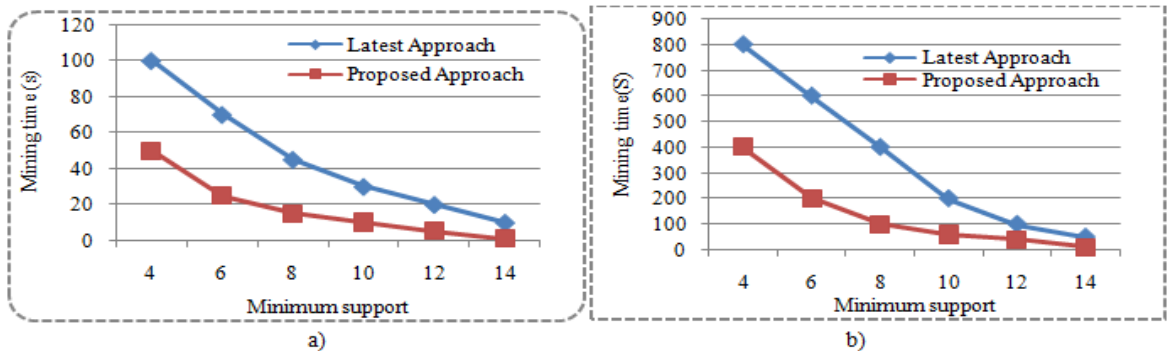


Figure 4. Mining time with change of minimum support on a) Mushroom b) Connect-4 dataset



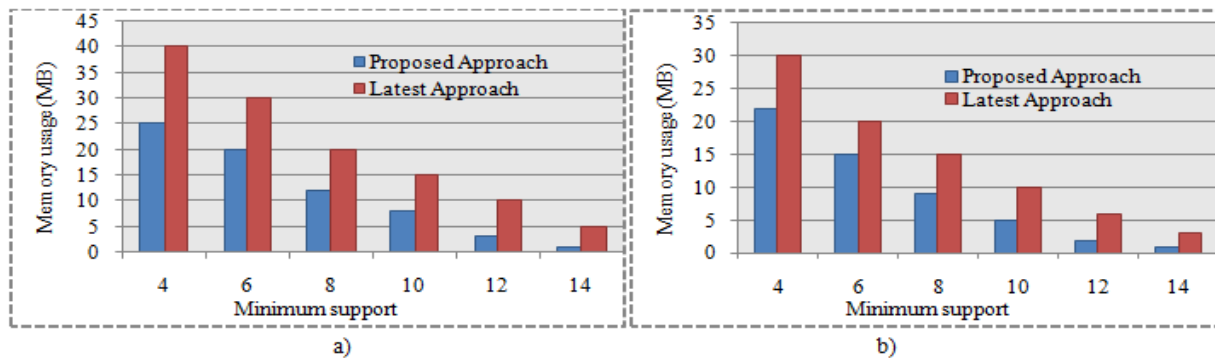


Figure 5. Memory usage with change of minimum support on a) Mushroom b) Connect-4 dataset

## VII. CONCLUSION

In this paper, we have proposed a complete solution of predicting e-shopper's purchase rules by using maximal frequent patterns and sequence close level. Experimental results indicate our proposed approach can give more accurate measurement in discovering e-shoppers purchase rules and useful for e-marketers to make marketing decisions.

On the other hand, since, businesses often attempt to adopt online shopping techniques without understanding them and/or without a sound business model, producing web stores that support the organizations' culture and brand name without satisfying consumer's expectations. Therefore, in future we intend to extend this work by considering different parameters and factors related to e-business models.

## ACKNOWLEDGEMENT

This work was supported by a grant from the NIPA (National IT Industry Promotion Agency, Korea) in 2012 (Global IT Talents Program) and partly supported by the National Research Foundation (NRF) grant (No. 2011-0018264) of Ministry of Education, Science and Technology (MEST) of Korea.

## REFERENCES

- [1] A. Gupta, B. Su, and Z. Walter: "Risk profile and consumer shopping behavior in electronic and traditional channels", *Decision Support System*, 200, Vol. 38, No. 3, PP: 347-367.
- [2] S. Liao, Y. Chen: "Mining customer knowledge for electronic catalog marketing", *Expert Systems with publications*, no.27, pp.521-532, 2004.
- [3] M. Chen, and J. Han: "Data mining: an overview from a database perspective", *IEEE Transactions on Knowledge and Data Engineering*, vol.8, 2006.
- [4] P. Giudici, and G. Passerone: "Data mining of association structures to model consumer behavior", *Computational Statistics and Data Analysis*, no.38, pp.533-541, 2002.
- [5] S. Changchien: "Mining association rules procedures to support online recommendation by customers and products fragmentation", *Expert Systems with Applications*, vol.20, no.4, pp.325-335, 2001.
- [6] F. Massegli, P. Poncelet, and M. Teisseire: "Incremental mining of frequent patterns in large databases", *Data & Knowledge Engineering*, vol.46, no.1, pp.97-121, 2003.
- [7] H. Cheng, X. Yan, J. Ha: "incremental mining of frequent patterns in large database", In *Proceedings of International Conference on Knowledge Discovery and Data Mining, KDD*, 2004.
- [8] J. Pei, J. Han, B. Mortazavi-Asl, J. Wang, H. Pinto: "Mining frequent patterns by pattern-growth: The PrefixSpan approach", *IEEE Transactions on Knowledge and Data Engineering*, vol.16, 2004.
- [9] H. Song, and J. Kim: "Mining the change of customer behavior in an Internet shopping mall", *Expert System with Applications*, Vol.21, no. 3, pp. 157-168, 2001.
- [10] R. Agrawal, and R. Srikant: "Fast algorithms for mining association rules", In *Proceedings of the International Conference on Very Large Data Bases*, pp, 1995.
- [11] J. Han, and M. Kamber, *Data Mining: Concepts and Techniques*, 2<sup>nd</sup> Edition: Morgan Kaufmann, 2006.
- [12] Gaborone, and G. Coenen: "Algorithms for computing association rules using a partial-support tree", *Knowledge-Based Systems*, vol. 13, 2004.
- [13] R. Agrawal, Imielinski, and T. A. Swami: "Mining association rules between sets of items in large database", In *Proc. Of the ACM SIGMOD Conf*, 03.
- [14] H. Mastsuzawa, and T. Fukuda: "Mining structured association patterns from databases", In *Proceeding of the 4th Pacific-Asia Conference, PAKDD*, 2000.
- [15] D. Xing, and J. Shen: "Efficient data mining for web navigation patterns", *Information and Software Technology*, Vol.46, PP.55-63, 2004.
- [16] J. Borges, and M. Levene: "Data mining of user navigation patterns, in: *Web Usage Analysis and User Profiling*", *Lecture Notes in Computer Science*, Vol. 1836, Springer, Berlin, pp. 92-111, 2000.
- [17] M.Kuramochi, and G.Karypis: "Frequent sub graph discovery, *Proceedings of IEEE International Conference on Data Mining (ICDM)*, San Jose, USA, pp. 313-320, November 2001.
- [18] M. Kubat, A. Hafez, and V.V. Raghavan: "Itemset trees for targeted association querying", *IEEE Transactions on Knowledge and Data Engineering*, Vol.15, 2003.
- [19] M.J. Zaki: "Efficiently mining frequent trees in a forest", In *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, Edmonton, Alberta, Canada.
- [20] U. Ckert, and S. Kramer: "Frequent free tree discovery in graph data", In *Proceedings of the 2004 ACM Symposium on Applied Computing*.
- [21] [http://en.wikipedia.org/wiki/Online\\_shopping](http://en.wikipedia.org/wiki/Online_shopping).
- [22] Y. Kai and M. Yuan: "A Fast Algorithm for Discovering Maximum Frequent Item sets" *IEEE 3rd International Conference on Communication Software and Networks (ICCSN)*, 2011.
- [23] H. Wang, C. Hu, and Y. Chen: "Mining Maximal Patterns Based on Improved FP-lattice structure and Array Technique" *2nd International Conf. on Future Computer and Communication*, 2010.
- [24] K. Gouda, M. Zaki: "A New Method for Mining Maximal Frequent Itemsets," In *Proceedings. Of the World Congress on Engineering* London, U.K, 2008.
- [25] D. Burdick, M. Calimlim, and J. Gehrke: "GenMax: An Efficient Algorithm for Mining Maximal Frequent Itemsets," *Data Mining and Knowledge Discovery*, the Netherlands, 2005.
- [26] L. Jian, and W. Chong: "Prediction of E-shopper's Behavior Changes Based on Purchase Sequences" *International Conf. on Artificial Intelligence and Computational Intelligence (AICI)*, 2010.
- [27] A. Meenakshi, and Dr. K. Alagarsamy: "Efficient Storage Reduction of Frequency of Items in Vertical Data Layout" *International Journal on Computer Science and Engineering*, Vol. 3 No. 2 Feb 2011.
- [28] C.Wang, J. Liu, and Y. Wang: "Mining Online customers' Purchase Rules Based on K-trees Frequent Pattern" *7th International Conf. on Fuzzy Systems and Knowledge Discovery (FSKD)*, 2010.
- [29] K. S.Beyer and R.Ramkrishnan "Bottom-up computation of sparse and Iceberg CUBEs" In *Proc. of the ACM SIGMOD*, 1999.