Analyzing Association Rule Mining and Clustering on Sales Day Data with XLMiner and Weka

A. M. Khattak, A. M. Khan, Sungyoung Lee*, and
Young-Koo Lee

Department of Computer Engineering, Kyung Hee University, Korea
{asad.masood, kadil, sylee}@oslab.ac.kr, yklee@khu.ac.kr

Abstract

In the era of intense competition among organizations, retaining a customer is a collaborative process. Business organizations are adopting different strategies to facilitate their customers in verity of ways, so that these customers keep on buying from them. Association Rule Mining (ARM) is one of the strategies that find out correspondence/association among the items sold together by applying basket analysis. The clustering technique is also used for different advantages like; recognizing class of most sold products, classifying customers based on their buying behavior and their power of purchase. Different researchers have provided different algorithms for both ARM and Clustering, and are implemented in different data mining tools. This paper is extended version of [4], we have compared the results of Apriori and K-Mean algorithms against their implementation in Weka and XLMiner. For this comparison we have used the transaction data of Sales Day (a super store). The results are very encouraging and also produced valuable information for sales and business improvements. We have also analyzed the data for hidden knowledge and the results showed some very interesting patterns in user buying behavior and buying timings.

Keywords: *Association Rule Mining, Clustering, Visualization, Weka, XLMiner.*

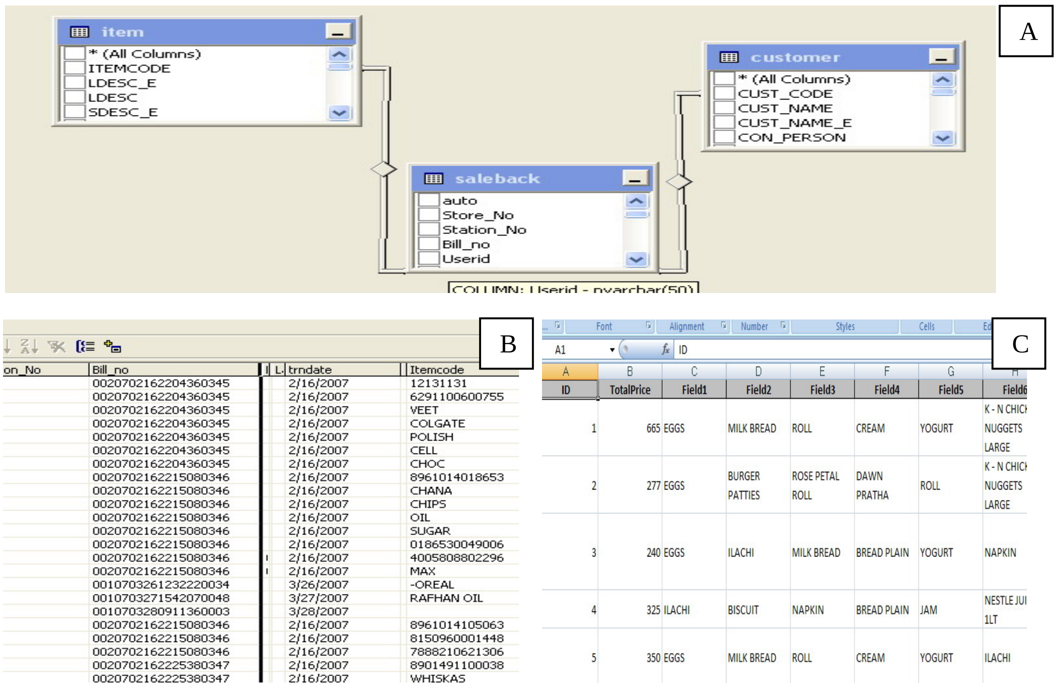
1. Introduction

Now-a-days, retaining old customers is preferred more than attracting new customers. Business organizations are adopting different strategies to facilitate their customers in verity of different ways, so that these customers keep on buying from them. Association Rule Mining (ARM) and data clustering is a particular kind of data mining problem for large set of multidimensional data points, In ARM we search for relationship among different items in the dataset while the data spare is usually not uniformly occupied so will produce different clusters. Data clustering identifies the sparse and the crowded places, and hence discovers the overall distribution patterns of the dataset. Association Rule Mining (ARM) [1] is one of the strategies that have two fold advantages to the business organization after applying the basket analysis. 1) It helps customers to get all the related items from one place and that save their time from visiting different places of the store. 2) It helps organization in more selling of items by placing items closer that are sold together. Different business organizations around the world have used basket analysis technique; among these; *Wal Mart*[1] is the most famous

[1] http://walmartstores.com/
[2] The original publication is available at: http://www.springerlink.com/content/p20785m486k387p5/

example. Clustering technique is used for classifying data based on some of its characteristics into different classes that eventually help users/organizations to further smoothen their business process. Clustering results have different advantages for business organizations; 1) to recognize the class of most sold products, 2) classifying customers based on their buying behavior and their power of purchase, 3) classifying customer's arrivals in different time slots based on customer's arrival time, and 4) identifying item(s) source for major trade.

Considering high dimensional data with noise and outliers, ARM and Clustering is a challenging task especially when data is very huge and complex. As discussed above, different researchers have provided different algorithms for ARM [5] and Clustering [2 and 3] that helps user to properly and efficiently achieve their objectives. The Apriori [5] algorithm is used for ARM; it had a problem of candidate set generation. This problem was removed, so the new improved Apriori algorithm reduce the time of scanning candidate set. It uses the hash tree to store the candidate sets that facilitate in solving the frequent set counting problem and is now more optimized based on time factor. The same way K-Means [3] algorithm is used for Clustering of data based on the parameter(s) specified.



on_No	Bill_no	L	trndate	Itemcode
	0020702162204360345		2/16/2007	12131131
	0020702162204360345		2/16/2007	6291100600755
	0020702162204360345		2/16/2007	VEET
	0020702162204360345		2/16/2007	COLGATE
	0020702162204360345		2/16/2007	POLISH
	0020702162204360345		2/16/2007	CELL
	0020702162204360345		2/16/2007	CHOC
	0020702162215080346		2/16/2007	8961014018653
	0020702162215080346		2/16/2007	CHANA
	0020702162215080346		2/16/2007	CHIPS
	0020702162215080346		2/16/2007	OIL
	0020702162215080346		2/16/2007	SUGAR
	0020702162215080346		2/16/2007	0186530049006
	0020702162215080346		2/16/2007	4005808802296
	0020702162215080346		2/16/2007	MAX
	0010703261232220034		3/26/2007	-OREAL
	0010703271542070048		3/27/2007	RAFHAN OIL
	0010703280911360003		3/28/2007	
	0020702162215080346		2/16/2007	8961014105063
	0020702162215080346		2/16/2007	8150960001448
	0020702162215080346		2/16/2007	7888210621306
	0020702162225380347		2/16/2007	8901491100038
	0020702162225380347		2/16/2007	WHISKAS

A	B	C	D	E	F	G	H
ID	TotalPrice	Field1	Field2	Field3	Field4	Field5	Field6
1	665	EGGS	MILK BREAD	ROLL	CREAM	YOGURT	K - N CHICK NUGGETS LARGE
2	277	EGGS	BURGER PATTIES	ROSE PETAL ROLL	DAWN PRATHA	ROLL	K - N CHICK NUGGETS LARGE
3	240	EGGS	ILACHI	MILK BREAD	BREAD PLAIN	YOGURT	NAPKIN
4	325	ILACHI	BISCUIT	NAPKIN	BREAD PLAIN	JAM	NESTLE JUI 1LT
5	350	EGGS	MILK BREAD	ROLL	CREAM	YOGURT	ILACHI

Figure 1. (A) shows the original structure of transection data storage in Sales Day Point of Sale system, (B) shows that for all the items sold in a transection is having a separate row entry, (C) shows the converted transections from multiple rows for one order to a single row. Taken from [4].

This is basically an extended version of our presented paper [4]; here we have compared the results of two Data Mining tools i.e. XLMiner [7] and Weka [6] for ARM and Clustering with Apriori and K-Means algorithms respectively. We have tested these algorithms of both the tools using daily transaction data from Point of Sale system of a super store *Sales Day*

*Corresponding author.
Department of Computer Engineering, Kyung Hee University, Korea. sylee@oslab.khu.ac.kr

(SD). We tested the Apriori and K-Means algorithms from both Weka and XLMiner on data of year 2007 of SD. By varying the parameters (i.e. support and confidence) for these algorithms; we got very interesting results for ARM discussed later. The same way, we also tested both the tools with K-Mean algorithm for clustering of the data to identify the different classes of items sold of particular amount and users. The customers are clustered based on their buying power, time and power of buying, most frequent customers, and transaction with amount of transactions that helps in focused advertisement based on customer arrival time and their buying behavior. We have also clustered the average no of transactions in different times of a day and the buying behavior of customers in different time intervals.

Rest of the paper is organized as follows; Section 2 is related to data formulation for the experiments to be used by the algorithms. Section 3 presents experimental results after applying both ARM and Clustering. Detail discussions are also made about the results. In Section 4 is based on conclusions and future directions.

2. Data Formulation

Before applying the algorithms on data, first we need to normalize the transaction data for the algorithms to work on. The daily transaction data of Sales Day (SD) store as shown in Figure 1 is in organization required format, which is a high dimensional and complex data that is not useable by the algorithms. For this reason, we have first converted the data from the organization required format to the format required to be used for experimentation. Figure 1-A shows the schema of SD Point of Sale (PoS) system where the SD data is redundant as clear from Figure 1-B. For every item sold in a single transaction, there is a complete row for that item and repeating the same order data again and again. We have developed a MS-Visual Basic 6.0 application using MS-SQL 2000 queries to translate the data in to a single row (pivoting) for every ordered transaction as shown in Figure 1-C. We have worked on the transaction data of year 2007 and tested both the tools for ARM and Clustering. The transaction had a variability in number of items contained in them e.g., a person may buy only a milk or a snack pack (i.e. only one item) but a transaction may contain a whole variety of items that range from daily use to occasionally used items that make the item set up to 60-80 products in it. We have fixed the number of items sold in a transaction to 12 items per transaction and any transaction having items more than 12 items then the remaining items are eliminated where transaction having less than 12 items are discarded from the dataset.

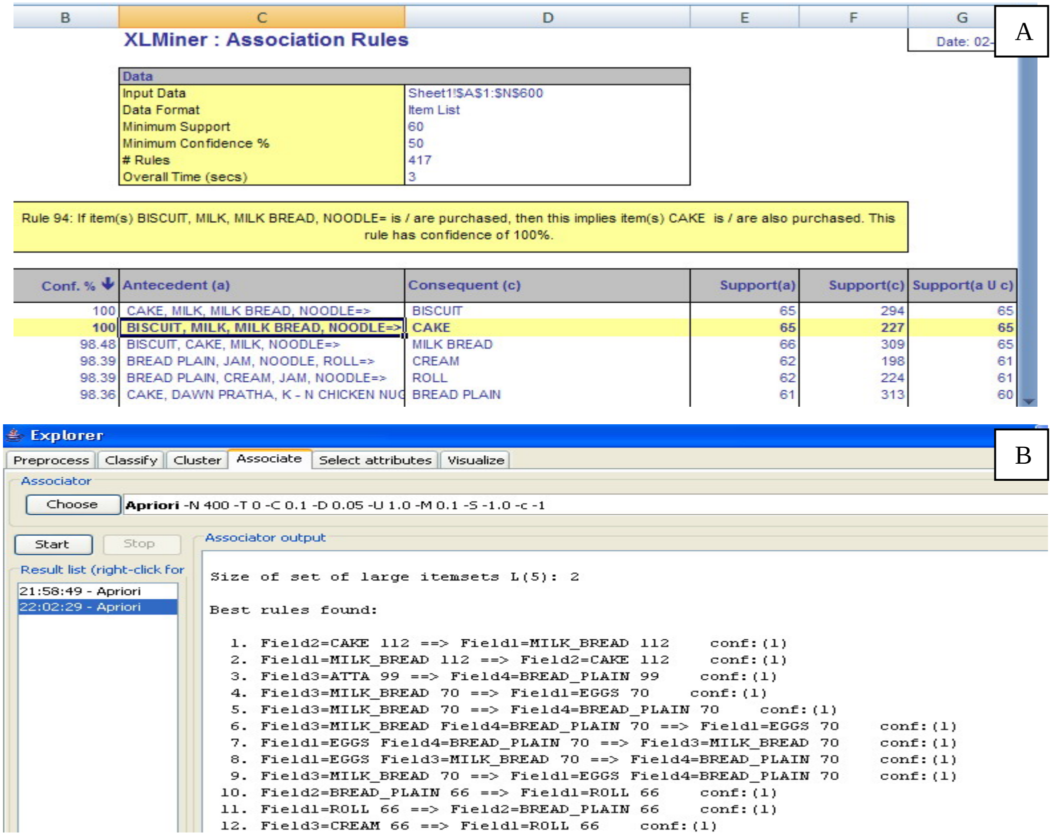
3. Experimental Results

Here we discuss in detail the comparative study of both the tools (i.e. Weka and XLMiner) for ARM and Clustering results over the transaction data normalized in the previous section. We have divided this work into two sections where first one focus on ARM and the second is focusing on Clustering. In Section 3.1 we will discuss ARM using both Weka and XLMiner and after that in Section 3.2 we will discuss Clustering results. The results from both tools will be compared in these two subsections. In Section 3.3 we have a general discussion on the results of these experiments and their benefits.

3.1. Association Rule Mining (ARM)

ARM finds interesting associations and/or correlation relationship among large set of data items. It infers attribute value conditions that occur frequently together in a given dataset e.g. Market Basket Analysis. Our goal is also to mine the Association Rules among data items from the transactions data of a super store Sales Day. The association rules provide information in the form of "if-then" statements where these are probabilistic in nature.

Different parameters are used in ARM for rule generations where their detail is: *Minimum support (# transactions):* Here it basically specifies the minimum number of transactions in which a particular item-set must appear so to qualify for inclusion in an association rule. We selected 60 as the minimum support for our work. *Minimum confidence (%):* It basically specifies the minimum confidence threshold for the rule generation. And our minimum confidence that we selected is 50. *Lift:* Lift is very important parameter. It is a parameter of interest in the association analysis. Lift is nothing but the ratio of *Confidence to Expected Confidence* i.e. the number of transactions that include the consequent divided by the total number of transactions. Lift is a value that gives us information about the increase in probability of the "then" (consequent) given the "if" (antecedent) part.



XLMiner : Association Rules

Data	
Input Data	Sheet1!$A$1:$N$600
Data Format	Item List
Minimum Support	60
Minimum Confidence %	50
# Rules	417
Overall Time (secs)	3

Rule 94: If item(s) BISCUIT, MILK, MILK BREAD, NOODLE= is / are purchased, then this implies item(s) CAKE is / are also purchased. This rule has confidence of 100%.

Conf. %	Antecedent (a)	Consequent (c)	Support(a)	Support(c)	Support(a U c)
100	CAKE, MILK, MILK BREAD, NOODLE=>	BISCUIT	65	294	65
100	BISCUIT, MILK, MILK BREAD, NOODLE=>	CAKE	65	227	65
98.48	BISCUIT, CAKE, MILK, NOODLE=>	MILK BREAD	66	309	65
98.39	BREAD PLAIN, JAM, NOODLE, ROLL=>	CREAM	62	198	61
98.39	BREAD PLAIN, CREAM, JAM, NOODLE=>	ROLL	62	224	61
98.36	CAKE, DAWN PRATHA, K - N CHICKEN NUG	BREAD PLAIN	61	313	60

Apriori -N 400 -T 0 -C 0.1 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

```
Size of set of large itemsets L(5): 2

Best rules found:

  1. Field2=CAKE 112 ==> Field1=MILK_BREAD 112    conf:(1)
  2. Field1=MILK_BREAD 112 ==> Field2=CAKE 112    conf:(1)
  3. Field3=ATTA 99 ==> Field4=BREAD_PLAIN 99    conf:(1)
  4. Field3=MILK_BREAD 70 ==> Field1=EGGS 70    conf:(1)
  5. Field3=MILK_BREAD 70 ==> Field4=BREAD_PLAIN 70    conf:(1)
  6. Field3=MILK_BREAD Field4=BREAD_PLAIN 70 ==> Field1=EGGS 70    conf:(1)
  7. Field1=EGGS Field4=BREAD_PLAIN 70 ==> Field3=MILK_BREAD 70    conf:(1)
  8. Field1=EGGS Field3=MILK_BREAD 70 ==> Field4=BREAD_PLAIN 70    conf:(1)
  9. Field3=MILK_BREAD 70 ==> Field1=EGGS Field4=BREAD_PLAIN 70    conf:(1)
 10. Field2=BREAD_PLAIN 66 ==> Field1=ROLL 66    conf:(1)
 11. Field1=ROLL 66 ==> Field2=BREAD_PLAIN 66    conf:(1)
 12. Field3=CREAM 66 ==> Field1=ROLL 66    conf:(1)
```

Figure 2. (A) Association rules identified using XLMiner. The yellow strip above display the complete rule selected in the table. (B) Association rules mined using Weka. Modified results of [4]

Among the different ARM algorithms available like; 1) Apriori, 2) Filtered Associations, 3) Predictive Apriori, and 4) Tertus, we choose to implement Apriori despite its multi scan drawback but the rules generated by Apriori are the most appropriate and finer granulized. To start working with Apriori for ARM, we have specified the environment variables as: 6994 instances of transactions with 12 attributes. The minimum support for ARM is set to 0.6 while minimum confidence is 0.9 with 20 numbers of cycles performed. Based on this input data for Apriori in both Weka and XLMiner, the association rules are mined (see Figure 2).

Results from both the tools depict same rules, while the representation of rules in these tools is different. For instance as shown in Figure 2-A, the confidence for CAKE to be purchased by the customer is 100 % if that customer is going to purchase BISCUIT, MILK, MILK BREAD, and NOODLE. The same rule is also represented in Figure 2-B of Weka where all the items are separately mapped with all the other items, and it also gives the confidence as 1 (which is equivalent to 100 %).

From Figure 3, it can be seen that how the associations are distributed over a plot. The circle points indicate that how the two transactions are related on the confidence level and products (items) occurring in their basket. This plot depicts the transactions with their respective basket items.

The associations/correspondence among sold items is one of the most useful source/results for the business organization. These associations are used by business organizations to resort their products in a way to place the most frequent sold items together. This also facilitates customers in quick checkout. One other strategy is to place the most sold items in different places. In this case the customer will have to visit different places in the store and will have a look at other different items available, that will increase their probability of been purchased by the customers.

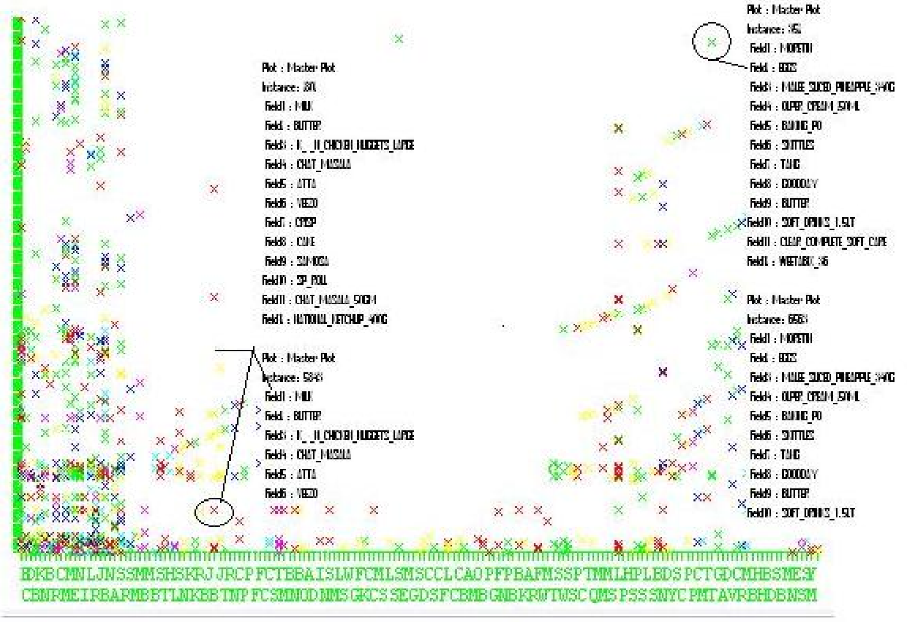


Figure 3. Visualization of association rules over a plot

3.2. Clustering

Clusters are often computed using a fast, heuristic method that generally produces good (but not necessarily optimal) solutions. The K-Means algorithm is one such method. We applied K-Means Clustering algorithm on the transaction data using XLMiner and Weka. In

XLMiner, to do clustering, we enter the data range that needs to be processed and move the variables of interest to the selected variables box. Here, it is visible in Figure 4-B that the numbers of clusters are 4 and the attribute selected for the clustering is *TotalPrice* of transactions. It also represents the mean of clusters. Figure 4-A shows different transactions that are classified in different clusters, while Figure 4-C represents the objects (clusters of transactions) and their total price against each transaction cluster.

Weka uses the centroids positions for calculation of clusters. We have tested the data with K-Means using Weka and got 3 clusters as number of clusters depends on choice of initial centroids, choice of distance measure, and stopping criterion that we defined.

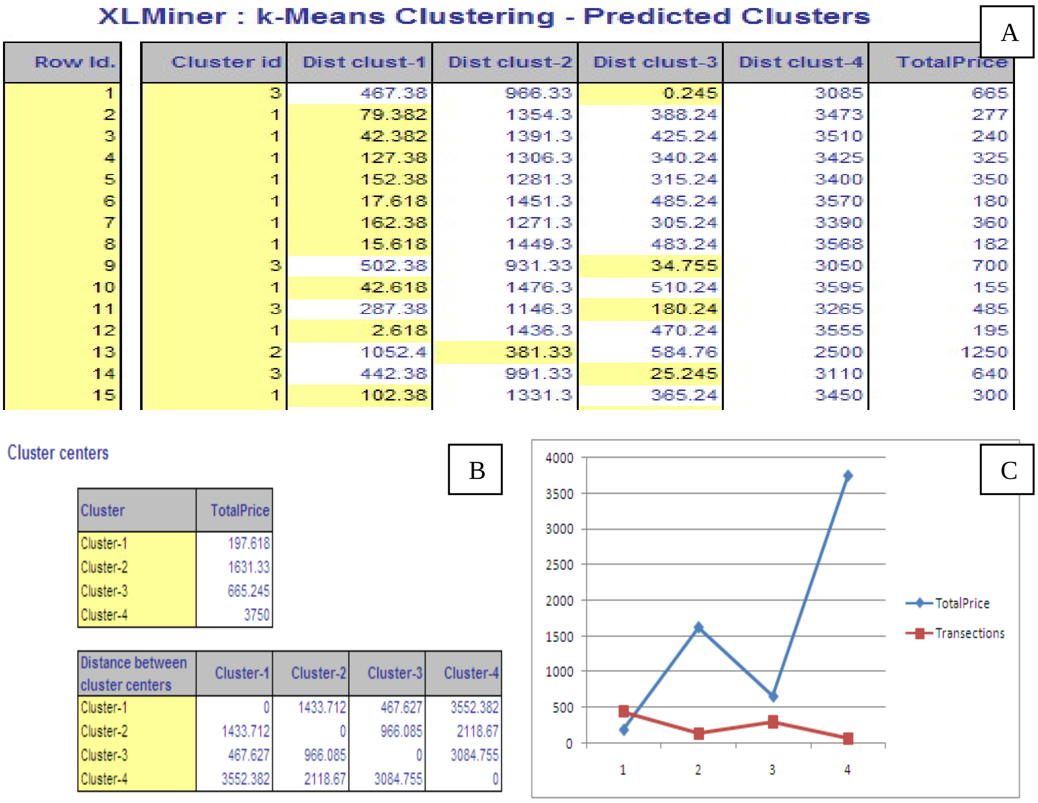
The runtime information is: Instances: 6994, Attributes: 14, and number of iterations: 7

```
Within cluster sum of squared errors: 73261.9474844484
Cluster centroids:
Cluster 0
Mean/Mode: 4195.3426 92045240804.8539 JAM JAM K_-
_N_CHICKEN_NUGGETS_LARGE EGGS NESTLE_JUICE_1LT MILK_BREAD
BREAD_PLAIN CAKE BREAD_PLAIN MILK_BREAD ROSE_PETAL_ROLL NOODLE
Std Devs: 1970.2377 4244911752738.16

Cluster 1
Mean/Mode:  2591.3249      4001780.4705 EGGS CAKE ROSE_PETAL_ROLL
BREAD_PLAIN JAM K_-_N_CHICKEN_NUGGETS_LARGE LAYS JAM
NESTLE_JUICE_1LT BURGER_PATTIES K_-_N_CHICKEN_NUGGETS_LARGE
LAYS
Std Devs:  1853.0708   196916696.0349

Cluster 2
Mean/Mode: 3581.0426 61689328101.9945 CAKE DAWN_PRATHA ATTA
MILK_BREAD CAKE BISCUIT BISCUIT BISCUIT BISCUIT ATTA NAPKIN
BREAD_PLAIN
Std Devs:  1810.0948 2483713762607.912

The clustered instances are:
0     2951 (42%)
1     2422 (35%)
2     1621 (23%)
```

These are overlapping clusters that are obtained by the range of the mean and standard deviation specified. By analysis of the visual plot obtained and shown in Figure 5. It is clear from Figure 5 that the clusters are roughly distributed. The color distribution for clusters is; Blue: Cluster 0, Green: Cluster 1, and Red = Cluster 2. The circle points show that at that point which cluster value it is.

3.3. Discussions

The ARM and Clustering work conducted here in this paper is basically for the purpose of comparative analysis of Weka and XLMiner with Apriori and K-Means algorithms. We have tested both the tools and for ARM they gave exact answers but during different experiments we performed for Clustering generated different clusters.



XLMiner : k-Means Clustering - Predicted Clusters

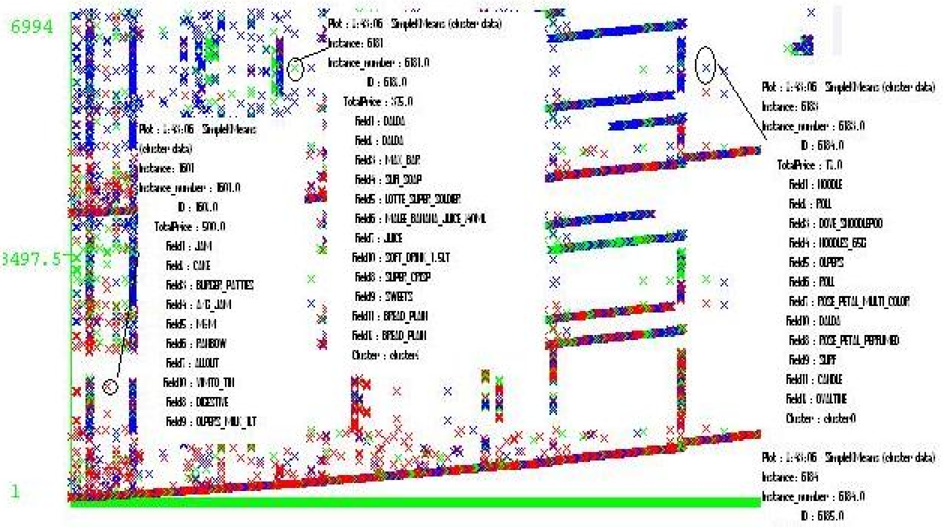
Row Id.	Cluster id	Dist clust-1	Dist clust-2	Dist clust-3	Dist clust-4	TotalPrice
1	3	467.38	966.33	0.245	3085	665
2	1	79.382	1354.3	388.24	3473	277
3	1	42.382	1391.3	425.24	3510	240
4	1	127.38	1306.3	340.24	3425	325
5	1	152.38	1281.3	315.24	3400	350
6	1	17.618	1451.3	485.24	3570	180
7	1	162.38	1271.3	305.24	3390	360
8	1	15.618	1449.3	483.24	3568	182
9	3	502.38	931.33	34.755	3050	700
10	1	42.618	1476.3	510.24	3595	155
11	3	287.38	1146.3	180.24	3265	485
12	1	2.618	1436.3	470.24	3555	195
13	2	1052.4	381.33	584.76	2500	1250
14	3	442.38	991.33	25.245	3110	640
15	1	102.38	1331.3	365.24	3450	300

Cluster centers

Cluster	TotalPrice
Cluster-1	197.618
Cluster-2	1631.33
Cluster-3	665.245
Cluster-4	3750

Distance between cluster centers	Cluster-1	Cluster-2	Cluster-3	Cluster-4
Cluster-1	0	1433.712	467.627	3552.382
Cluster-2	1433.712	0	966.085	2118.67
Cluster-3	467.627	966.085	0	3084.755
Cluster-4	3552.382	2118.67	3084.755	0

Figure 4, (A) Shows classification of transactions in clusters, (B) shows number of clusters and their distance, and (C) visually represent transactions in cluster and their total price. Modified results of [4]



Plot : 1:43:05 SimpleKMeans (cluster data)
Instance: 6181
Instance_number : 6181.0
ID : 6182.0
TotalPrice : 375.0
Field1 : DALDA
Field2 : DALDA
Field3 : MAX_BAR
Field4 : SUN_SOAP
Field5 : LOTTE_SUPER_COLDER
Field6 : MALEE_BANANA_JUICE_200ML
Field7 : JUICE
Field10 : SOFT_DRINK_1.5LT
Field8 : SUPER_CRISP
Field9 : SWEETS
Field11 : BREAD_PLAIN
Field12 : BREAD_PLAIN
Cluster : cluster4

Plot : 1:43:05 SimpleKMeans (cluster data)
Instance: 1601
Instance_number : 1601.0
ID : 1602.0
TotalPrice : 500.0
Field1 : JAM
Field2 : CAKE
Field3 : BURGER_PATTIES
Field4 : A/G_JAM
Field5 : M&M
Field6 : RAINBOW
Field7 : ALLOUT
Field10 : VIMTO_TIN
Field8 : DIGESTIVE
Field9 : OLPERS_MILK_1LT

Plot : 1:43:05 SimpleKMeans (cluster data)
Instance: 6183
Instance_number : 6183.0
ID : 6184.0
TotalPrice : 75.0
Field1 : NOODLE
Field2 : ROLL
Field3 : DOVE_SHAMPOO
Field4 : NOODLES_65G
Field5 : OLPERS
Field6 : ROLL
Field7 : ROSE_PETAL_MULTI_COLOR
Field10 : DALDA
Field8 : ROSE_PETAL_PERFUMED
Field9 : SURF
Field11 : CANDLE
Field12 : OVALTINE
Cluster : cluster0

Plot : 1:43:05 SimpleKMeans (cluster data)
Instance: 6184
Instance_number : 6184.0
ID : 6185.0

Figure 5. Visualization of clustered transactions using Weka. Taken from [4]

Beside this comparison objective, we also got some very interesting results based on the transactions data. We got the most frequent sold items in a basket that helped the organization in re-organizing the sale strategy for these items to improve their sale. As shown above, in Clustering we have also classified the transactions data on different parameters like; age group of customers, age group plus purchase power, most sold items, time of high customer traffic, and time of high purchase power customer's arrival. These classified (Clustered) transactions are used by the business organization to great advantage. These results are very useful, for instance; with the help of cluster i.e. Time of high purchase power customer arrival, the organization can lunch new high cost products for these customers. Knowing the arrival time of particular type of customers can also be used for focused advertisement of product of interest to the arriving customers. These results also avoid the out-of-stock situation as it gives information about most sold items.

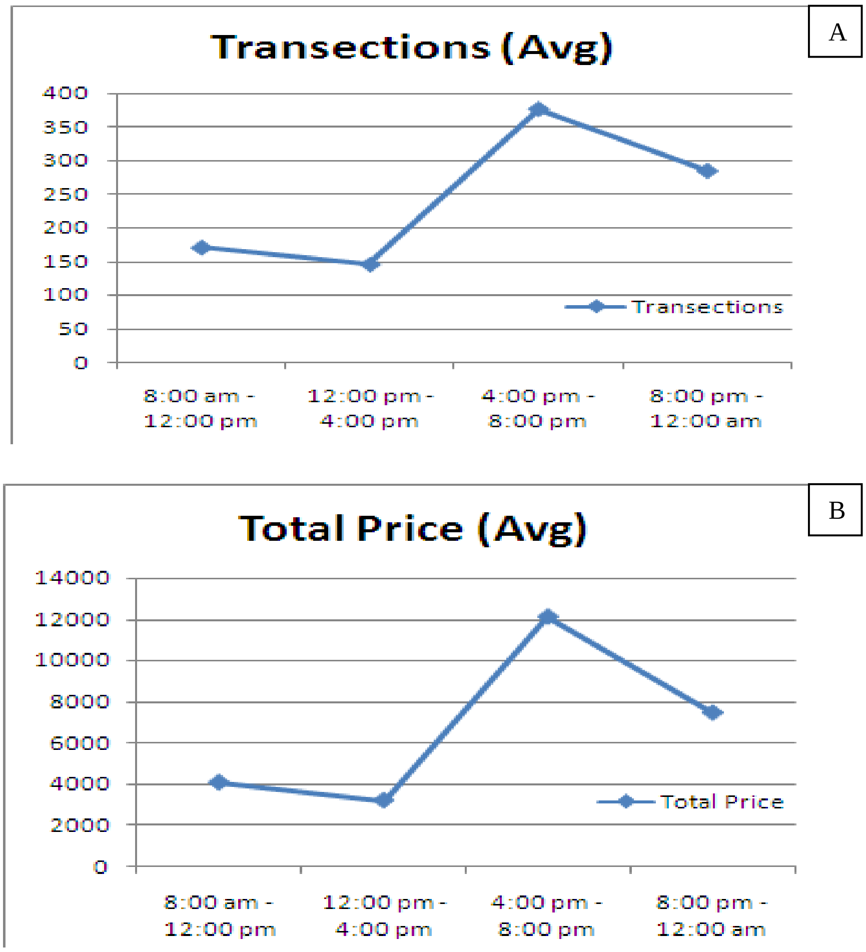


Figure 6. (A) Shows the average no of transactions in different times of a day where (B) shows the amount (money) spent by customers in different times of a day.

As shown in Figure 6-A, we have evaluated the average no of transactions in different time intervals of a day per a complete year of 2007 and got very interesting and realistic results. In the morning time the tendency is for items purchased for breakfast while in office times the purchase pattern is decreased. After office hours are the best time for buying different items and the transactions graph goes very high. The same way in Figure 6-B, the graph shows that the high amount purchases are made in after office hours. These results make the organization focus on their most important time intervals of selling products while in other time slots they can focus more on the other related activities.

4. Conclusions and Future Work

In this paper, Association Rule Mining and Clustering algorithms i.e. Apriori and K-Means are evaluated on a transaction data of a super store. ARM and Clustering are well established areas of data mining. These are used for extracting hidden facts from a huge repository of raw data. We used these two techniques with Apriori and K-Means algorithms implemented in Weka and XLMiner to analyze the trend of sale at a super store Sales Day. We have compared these algorithms by using Weka and XLMiner over Sales Day data and got very encouraging results that not only satisfy the implementation of these algorithms in both the tools as same but also support the business organization for customer support and future extension in their business. The results also suggest and/or encourage making some strategy change for organization that might result in better sale. We are planning to extend our work to different tools for more algorithms and use the results to business advantages. More sophisticated visualization of results is also under considerations for better human understandability.

Acknowledgement

This research was supported by the MKE (Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the NIPA (National IT Industry Promotion Agency)" (NIPA-2009-(C1090-0902-0002)). This work was supported by the IT R&D program of MKE/KEIT. [2009-S-033-01, Development of SaaS Platform for S/W Service of Small and Medium sized Enterprises].

References

[1] R. Agrawal, T. Imielinski, and A. Swami, “Mining Associations between seta of items in Massive Databases”. In proc. Of the ACM-SIGMOD 1993 int’l conf. on Management of Data, pages 207-216, Washingtom D.c USA, 1993.

[2] L. Ertoz, M. Steinbach, and V. Kumar, “Finding Clusters of Different Sizes, Shapes, and Densities in Noisy,” SIAM International Conference on Data Mining, Feburary 20, 2003.

[3] L. Kaufman and P. J. Rousseeuw, “Finding Groups in Data: An Introduction to Cluster Analysis,” Wiley Series in Probability and Statistics. John Wiley and Sons, New York, Novemeber 1990.

[4] A.M. Khattak, A.M. Khan, Tahir Rasheed, Sungyoung Lee, and Young-Koo Lee, “Comparative Analysis of XLMiner and Weka for Association Rule Mining and Clustering,” 1st International Conference on Database Theory and Applications (DTA 2009), pp. 82–89, Jeju - Korea, December 2009.

[5] X. W. Liu and P. L. He, “The research of improved association rules mining Apriori algorithm” Proceedings of 2004 International Conference on Machine Learning and Cybernetics, Volume 3, Issue, Page(s): 1577 – 1579, 26-29 Aug. 2004.

[6] Ian H. Witten and E. Frank (2005) "Data Mining: Practical machine learning tools and techniques", 2nd Edition, Morgan Kaufmann, San Francisco, 2005.

[7] Data Mining Add-In for Excel, http://www.resample.com/xlminer/

Authors

Asad Masood Khattak received his BS from Institute of Computing and Information Technology, Gomal University in 2006. He received his MS in Information Technology from School of Electrical Engineering and Computer Science, National University of Sciences and Technology in 2008. Since March 2009, he is working on his PhD degree in Department of Computer Engineering at Kyung Hee University, Korea. His research interests include data management, knowledge representation, semantic web, and ontology. Email: asad.masood@oslab.khu.ac.kr

Adil Mehmood Khan received his BS from NUST Institute of Information Technology, Pakistan in 2005. Since March 2006, he is working on his PhD degree in Department of Computer Engineering at Kyung Hee University, Korea. His research interests include pattern recognition, data mining, knowledge discovery, and biomedical signal processing. Email: kadil@oslab.khu.ac.kr

Sungyoung Lee received his B.S. from Korea University, Seoul, Korea . He got his M.S. and PhD degrees in Computer Science from Illinois In stitute of Technology (IIT), Chicago, Illinois, USA in 1987 and 1991 re spectively. He has been a professor in the Department of Computer Eng ineering, Kyung Hee University, Korea since 1993. He is a founding dir ector of the Ubiquitous Computing Laboratory, and has been affiliated with a director of Neo Medical ubiquitous-Life Care Information Techn ology Research Center, Kyung Hee University since 2006. He is a mem ber of ACM and IEEE. Email: sylee@oslab.khu.ac.kr

Young-Koo Lee got his B.S., M.S. and PhD in Computer Science from Korea advanced Institute of Science and Technology, Korea. He is a professor in the Department of Computer Engineering at Kyung Hee University, Korea. His research interests include ubiquitous data management, data mining, and databases. He is a member of the IEEE, the IEEE Computer Society, and the ACM. Email: yklee@khu.ac.kr