

January 2012

## Selection of Optimal Discount of Retail Assortments with Data Mining Approach

Padmalatha Eddla

CBIT, Hyderabad, A.P, India, padmalathae.cbit@gmail.com

R.Ravinder Reddy

CBIT, Hyderabad, A.P, India., ravi.ramasani@gmail.com

Mamatha Ramasani

Computer Science Department, CBIT, Gandipet, Hyderabad, A.P, India., mamatha.ramasani@gmail.com

Follow this and additional works at: <https://www.interscience.in/ijcsi>



Part of the [Computer Engineering Commons](#), [Information Security Commons](#), and the [Systems and Communications Commons](#)

---

### Recommended Citation

Eddla, Padmalatha; Reddy, R.Ravinder; and Ramasani, Mamatha (2012) "Selection of Optimal Discount of Retail Assortments with Data Mining Approach," *International Journal of Computer Science and Informatics*: Vol. 1 : Iss. 3 , Article 6.

DOI: 10.47893/IJCSI.2012.1032

Available at: <https://www.interscience.in/ijcsi/vol1/iss3/6>

This Article is brought to you for free and open access by the Interscience Journals at Interscience Research Network. It has been accepted for inclusion in International Journal of Computer Science and Informatics by an authorized editor of Interscience Research Network. For more information, please contact [sritampatnaik@gmail.com](mailto:sritampatnaik@gmail.com).



## Selection of Optimal Discount of Retail Assortments with Data Mining Approach



**Padmalatha Eddla, Ravinder Reddy, Mamatha**

Computer Science Department, CBIT, Gandipet, Hyderabad, A.P, India.

E-mail : [padmalathae.cbit@gmail.com](mailto:padmalathae.cbit@gmail.com), [ravi.ramasani@gmail.com](mailto:ravi.ramasani@gmail.com), [mamatha.ramasani@gmail.com](mailto:mamatha.ramasani@gmail.com)

**Abstract** - Recently, the capabilities of generating and collecting data have been increasing rapidly. Widespread use of bar codes for most commercial products, the computerization of many business, and the advance in data collection tools have provided us with huge amount of retail data. This explosive growth in data and databases has generated an urgent need for data mining techniques and tools that can extract implicit, previously unknown and potentially useful information from data in data storages. One of the most popular data mining approaches is "association rules", which is commonly applied to analyze market baskets to help managers to determine which items are frequently purchased together by customers.

Affinity analysis is a data analysis and data mining technique that discovers co-occurrence relationships among activities performed by (or recorded about) specific individuals or groups. In general, this can be applied to any process where agents can be uniquely identified and information about their activities can be recorded. In retail, affinity analysis is used to perform market basket analysis, in which retailers seek to understand the purchase behavior of customers. This information can then be used for purposes of cross-selling and up-selling, in addition to influencing sales promotions, loyalty programs, store design, and discount plans.

**Keywords**— Association rules, Affinity analysis, temporal support threshold.

### 1. INTRODUCTION

Since almost all mid to large size retailers today possess electronic sales transaction systems, retailers realize that competitive advantage will no longer be achieved by the mere use of these systems for purposes of inventory management or facilitating customer check-out. In contrast, competitive advantage will be gained by those retailers who are able to extract the knowledge hidden in the data, generated by those systems, and use it to optimize their marketing decision making. In this context, knowledge about how customers are using the retail store is of critical importance and distinctive competencies will be built by those retailers who best succeed in extracting actionable knowledge from these data. Association rule mining [2] can help retailers to efficiently extract this knowledge from large retail databases. We assume some familiarity with the basic notions of association rule mining. In recent years, a lot of effort in the area of retail market basket analysis has been invested in the development of techniques to increase the interestingness of association rules. Currently, in essence three different research tracks to study the interestingness of association rules can be distinguished. First, a number of objective measures of interestingness have been developed in order to filter out non-interesting association rules based on a number of

statistical properties of the rules, such as support and confidence [2], interest [4], intensity of implication, J-measure [5], and correlation [8]. Other measures are based on the syntactical properties of the rules [10], or they are used to discover the least redundant set of rules. Second, it was recognized that domain knowledge may also play an important role in determining the interestingness of association rules. Therefore, a number of subjective measures of interestingness have been put forward, such as unexpectedness [3], action ability [1] and rule templates [2]. Finally, the most recent stream of research advocates the evaluation of the interestingness of associations in the light of the micro-economic framework of the retailer [2]. More specifically, a pattern in the data is considered interesting only to the extent in which it can be used in the decision-making process of the enterprise to increase its utility. It is in this latter stream of research that the authors have previously developed a model for product selection called PROFSET [3], that takes into account both quantitative and qualitative elements of retail domain knowledge in order to determine the set of products that yields maximum cross selling profits. The key idea of the model is that products should not be selected based on their individual profitability, but rather on the total profitability that they generate.

## 2. THE PROBLEM AND DEFINITIONS

The procedure of finding optimal discount for each product begins with an algorithm that discovers temporal association rules with discount information. Next, the procedure proceeds to product discount selection in which the profits of frequent itemsets are considered. Thus, the best discount for each product is selected based on the total profitability that it generates, including profits from cross-selling.

Thus the generalized methodology adopted for the implementation of the aforesaid problem includes- A variation of the Apriori algorithm is developed to mine the association rules by taking into consideration the item- discount pairs and the temporal effects of discount. The variant is termed as the “Apriori -Td” algorithm. It adds new filtering mechanisms including “local support” and “temporal support threshold”.

Large Local 1-D Itemsets -

{(Bread, 2)}, {(Jam, 1)}, {(Milk, 2)}

, where, {(Bread, 2)} => Bread at discount rate 2 is an itemset.

Large Local 2-D Itemsets-

{(Bread, 2) (Jam, 3)} {(Jam, 1) (Milk, 2)}

, where, {(Bread, 2) (Jam, 3)} => Bread at discount rate 2 and Jam at discount rate 3 is an itemset.

- a.) The estimation of the profits of the frequent itemsets outputted from the Apriori-Td algorithm.

Results-

{(Bread, 2)(Jam, 3)}=> Gross Margin : 32.25

- b.) A microeconomic optimization model for discovering the best discount combination for products in which the overall profitability of the frequent itemsets.

Optimal Product Discount Selection –

(Bread, 2), (Jam, 1), (Cereal, 2), (Milk, 2)

,specifying, Bread at discount rate 2,Jam at discount rate 1, cereal at discount rate 2 and milk at discount rate 2 produces the highest overall profit(overall gross margin).

Thus, by doing so not only are the frequency and quantity of a purchased product with a specified discount level is taken into account, but also the effect of such a discount level on selling other products is also factored into the final conclusions drawn from the entire method.

Thus, the optimal discount selection model ensures increases overall profits for the retailer by suggesting optimal discount rates which have been gleaned from interpreting transaction data and evolving a scientific method of maximizing the profit function for the store at large.

## 3. THE APRIORI - Td ALGORITHM

Apriori is a classic algorithm for learning association rules. Apriori is designed to operate on databases containing transactions (for example, collections of items bought by customers, or details of a website frequentation). Apriori uses breadth-first search and a tree structure to count candidate item sets efficiently. It generates candidate item sets of length k from item sets of length k - 1. Then it prunes the candidates which have an infrequent sub pattern. The Apriori-TD differs from Apriori and Apriori-like algorithms for mining frequent itemsets in the fact that it considers the temporal characteristics of discount on products.

### Inputs

- a.) Discount information on products at that particular time period.
- b.) Transaction sales data
- c.) Thresholds for local support, global support, positive effect, negative effect and time for minimum amount of time where the item, discount pair occurs for.

### Functionality

In the first step, we find global-large-1 itemsets, which consists of the items which exceed the global threshold. Then, the local-candidate-1 sets are generated by determining every different discount levels for every item and determining the time interval in which item appears with each discount level. These local candidates thus generated are subjected to pruning using the thresholds for local supports and amount of time in which the item - discount pair appears. Then iteratively larger sets (size:K) are derived from (K-1) size sets by merging them and the real support is calculated, effectiveness on sales is determined.

## 4. PRODUCT DISCOUNT SELECTION PROCEDURE

In this study, by the use of Apriori-Td algorithm, we can specify which products with how much discount is frequently bought by customers at the same market baskets. In this section, a microeconomic zero-one integer programming model is introduced for product discount selection. That is, the proposed model provides us the products that should be on sale and of course, the discount rates of these selected products. The results of

the optimization problem may express that some products should not be sold with discount.

**4.1. Estimation of Gross Margins**

In this sub-module, we analyze the profits due to each frequent d-itemset as generated by the Apriori-TD algorithm and allot these profits to the corresponding d-itemsets.

**Input**

- a.) The frequent d- itemsets which are outputted from the Apriori TD- Algorithm.
- b.) The pricing details of the products, which include, purchasing price, selling price and the discount rate.
- c.) The transaction sales information, including the time period in which it occurred and the quantity of items purchased.

**Functionality**

In this function, the margin (profit) generated due to the purchase of the frequent itemsets are calculated. This is based on the transaction data, which helps us determine the intentions of the customer when making the transaction.

**Output**

After the execution of this sub-module, we have the profits due to each frequent d-itemset.

**4.2 Optimal Discount Selection procedure**

In this step, the zero-one integer programming methodology is used to find the optimal product discount for maximizing the profits of frequent d-itemsets.

The frequent d-itemsets with their corresponding gross margins.

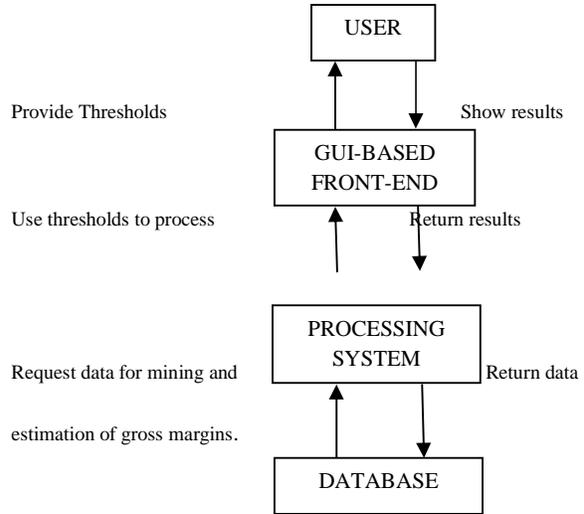
In this function, we try to maximize the overall gross margins by making combinations of the various d-itemsets under certain constraints, most important of which is – any selected item can have only one discount rate. All possible combinations are considered before the best combination is selected.

After execution of this sub-module, we get the final d-itemset combination. The (item, discount) pairs are now extracted from this combination.

**5. ARCHITECTURE OF THE SYSTEM**

The system consists of a front-end provided by the GUI which enables the user to interact better with the system. This is connected to the processing system wherein all the calculations are performed. The user inputs the thresholds and selects to run particular threshold analysis or graphical analysis wherein results for a range of local support threshold values are

provided. These inputs are fed to the Apriori-TD algorithm for processing. It accesses the database to mine for frequent itemsets and d-itemsets. Once Apriori-TD completes execution, we need to assign gross margins to the d-itemsets. Here, we need the sales information like cost price and market price along with the exact discount rate at which each item was sold in every transaction. Once this is completed, database access is no longer required. The maximization of gross margins in performed and the results are returned to the GUI which displays the result.



**Figure 5.1:** System Architecture

**5.1. The experiment**

The product optimal discount selection model was used to analyze market data consisting of 40 transactions for 15 items. The support thresholds were as follows –

- Global support = 0.2
- Local support = 0.06
- Time support = 3
- Effect high threshold = 1.3
- Effect low threshold = 0.7
- Global Large 1-itemsets  
BREAD BUTTER SUGAR CHOCOLATE
- Large local k-d-itemsets
- Large local 1-d itemsets -  
{(BREAD , 1)} {(BREAD , 2)} {(BUTTER , 2)}  
{(BUTTER , 5)} {(SUGAR , 1)} {(SUGAR , 2)}  
{(CHOCOLATE , 4)}
- Large local 2-d itemsets -

{{(BREAD , 2) (BUTTER , 5)}}

Gross Margins of itemsets -

{{(BREAD , 1)}=> Gross Margin: 78.75

{{(BREAD , 2)}=> Gross Margin: 12.5

{{(BUTTER , 2)}=> Gross Margin: 15.75

{{(BUTTER , 5)}=> Gross Margin: -7.0

{{(SUGAR , 1)}=> Gross Margin: 52.5

{{(SUGAR , 2)}=> Gross Margin: 50.0

{{(CHOCOLATE , 4)}=> Gross Margin: 85.0

{{(BREAD , 2) (BUTTER , 5)}=> Gross Margin: 44.25

Optimal Product Discount Selection -

(BREAD , 1)

(BUTTER , 2)

(SUGAR , 1)

(CHOCOLATE , 4)

Overall Gross Margin - 232.0

Profit Analysis -

Old Profits - 1480.3

New Profits - 1616.6

Improved Profits - 136.3 which in % terms is 9.2

Sales - 7760.34

Improved Profits with respect to sales - 1.8

The above observations show a spike in net profits of about 9.2 % for the given thresholds.

## 6. RESULT ANALYSIS

The above results indicate that if the merchant had sold the items at the above prescribed discount rates, it would have generated increased profits of 136.3 units. It also, tells us that the increased profits would have been about 2 % of the total sales, which means you would get 2 units more for every 100 units worth of products sold. The Apriori-TD identifies only 4 items with good global support, namely – BREAD, BUTTER, SUGAR and CHOCOLATE. Combinations of these items with their various discount rates are made to get local 1-d itemsets. In this example, the 1-d itemsets total to 7. These are further combined to get local 2-d itemsets. In this example, only 1 local 2-d itemset makes the cut.

Next, we estimate the gross margins of each of these large local sets over all the transactions. The result of this step is listed above. Next, we try to maximize the overall gross margins. It can be clearly seen that the derived (item, profit) optimized values give the maximum gross margin.

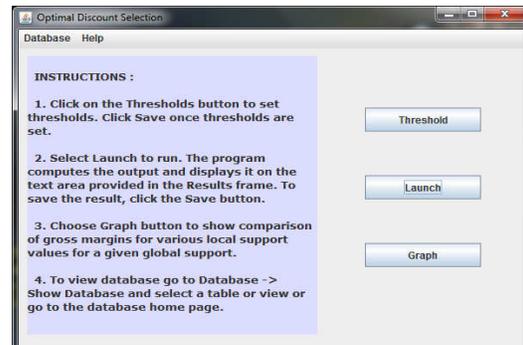


Figure 6.1: GUI Home Screen

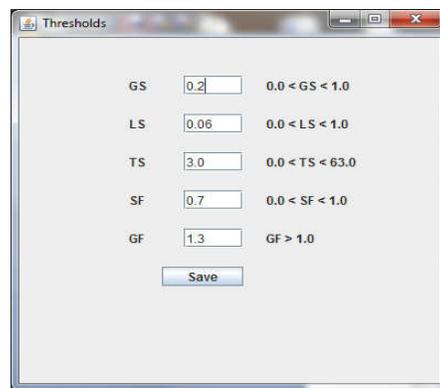


Figure 6.2: Setting the thresholds

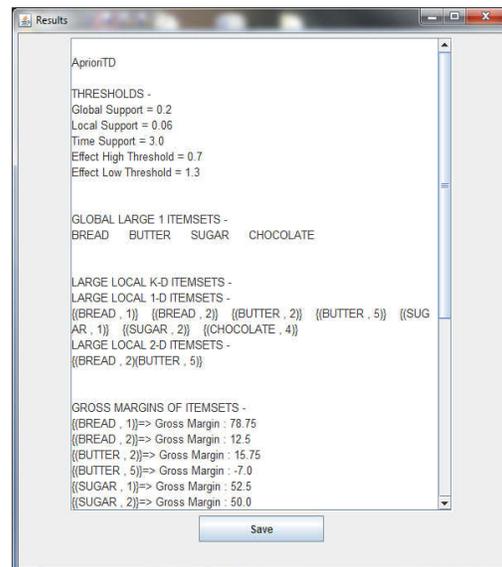


Fig 6.3: Result analysis for given data and threshold values

Local support	Gross Margin
0.1	198.8
0.2	85
0.3	0
0.4	0
0.5	0
0.6	0
0.7	0
0.8	0
0.9	0

**Table 6.1 Global and Local Support**

Global Support 0.2  
Time Support

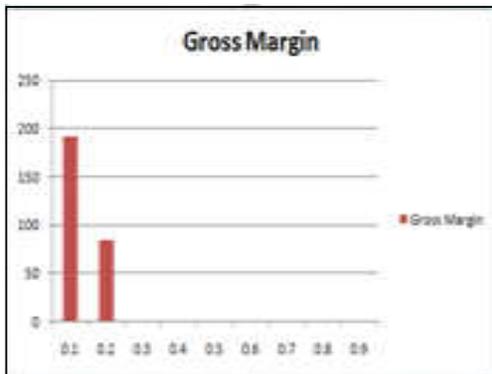


Fig. 6.4: Graph analysis for global support threshold value 0.2

**7. CONCLUSION**

Since almost all mid to large size retailers today possess electronic sales transaction systems, retailers realize that competitive advantage will no longer be achieved by the mere use of these systems for purposes of inventory management or facilitating customer check-out. In contrast, competitive advantage will be gained by those retailers who are able to extract the knowledge hidden in the data, generated by those systems, and use it to optimize their marketing decision making.

The foresaid methodology has been duly adhered to and the problem at hand has been addressed. The Apriori- Td algorithm has been implemented and the gross margins are got retrieved using the Generalized PROFSET model. Then a micro-economic zero-one integer programming model is utilized to get the best product- discount combinations.

**REFERENCES**

- [1] R. Agrawal and R. Srikant, “Fast algorithm for mining association rules”, in Proc.20th International Conference on very large data bases, Santiago, China, 1994, 487-499.
- [2] J.S. Park, M.-S. Chen, and P.S. Yu, “An effective hashbased algorithm for mining association rules”, in Proc. ACM SIGMOD International Conference on Management of Data, San Jose, CA, 1995.
- [3] R. Rastogi, and K. Shim, “Mining optimized association rules with categorical and numeric attributes”, in Proc. IEEE Transactions on Knowledge and Data Engineering, 1998, pp.503-512.
- [4] B. Padmanabhan and A. Tuzhilin. nexpectedness as a measure of interestingness in knowledge discovery. In Decision Support Systems, volume 27, pages 303–318. Elsevier Science, 1999.
- [5] L. Wang, K. Xie, T. Chen, and X. Ma, “Efficient discovery of multilevel spatial association rules using partitions”, Journal of Information and Software Technology, 2005, pp. 829-840.
- [6] K. Koperski, J. Han, “Discovery of spatial association rules in geographic information databases”, in Proc. 4<sup>th</sup> International Symposium on Advance in spatial Databases,1995, pp. 47-66.
- [7] S. Guillaume, F. Guillet, and J. Philipp. Improving the discovery of association rules with intensity of implication. In Principles of Data Mining and Knowledge Discovery, volume 1510 of Lecture Notes in Artificial Intelligence, pages 318–327. Springer, 1998.
- [8] P. Cabena, P. Hadjinian, R. Stadler, J. Verhees, and A. Zanasi. Discovering Data Mining: From Concept to Implementation. Prentice Hall, 1997.
- [9] B. Liu and W. Hsu. Post-analysis of learned rules. In Proceedings of the Thirteenth National Conference on Artificial Intelligence, Lecture Notes in Artificial Intelligence, pages 828–834. AAAI Press/MIT Press, 1996.
- [10] P. Cabena, P. Hadjinian, R. Stadler, J. Verhees, and A. Zanasi. Discovering Data Mining: From Concept to Implementation. Prentice Hall, 1997.

□□□