Product Assortment Decisions for a Network of Retail Stores using Data Mining with Optimization

Sudip Bhattacharjee, Fidan Boylu and Ram Gopal
Department of Operations and Information Management, University of Connecticut, Storrs, U.S.A.

Keywords: Product Assortment, Retail Operations, Optimization, Association Rule Mining.

Abstract: This paper presents a model for product assortment optimization for a network of retail stores operating in various locations of a company. Driven by the local market information of each retail store, the model determines the right products to include in a store's assortment and which stores to ship from in the store network. The model first learns the global patterns of the frequent itemsets based on association rule mining to extract patterns of products with corresponding sales benefits. It then encodes the pattern information into the development of a global optimization formulation, which maximizes the revenue of the company in aggregate and identifies the optimal solution for each local store by taking into account the possibility of shipments in the network. We use the transactional level data from an industry leading plastics manufacturer and retailer in the United States to demonstrate the utility of the model.

1 INTRODUCTION

Retailers with multiple stores of commodity products commonly face the problem of product assortment selection, especially when these stores are spread over several different geographic locations. Given a network of stores, the problem becomes how to incorporate varying market conditions of store locations and at the same time take into account the possible transportation of products between stores, to decide on the right product assortment for a given store location.

We approach the problem using data mining an optimization tools. We use the conflict resolution method (Bai et al., 2012) for itemsets extracted using Association Rule Mining to extract global patterns. We then use these patterns in a network optimization problem to find the optimal product assortment and shipment decisions for local stores. We will make use of the measures developed in Bai et al. (2012) to identify the concentration of clients interested in certain patterns around a given location and the retailer's global sales performance on such patterns.

For our analysis, we use the data of an industry leading plastics manufacturer and distributor with over seventy branch locations in the US, Canada and Mexico. It offers a wide range of products such as plastic sheet, rod, tube and film in various shapes and sizes. These products are used to serve various industries such as appliances, automotive, electrical, manufacturing, medical, printing and transportation. Companies such as these, with a large number of branch locations, strive for solutions that provide data-driven, quantitative decision support tools that would help decide on the right product assortment and shipment decisions for maximum profitability of branch and company-wide profitability.

This paper presents a model for product assortment optimization for multiple stores of a company where these stores, while holding a local assortment of products, can also receive shipments of products outside their portfolio from other stores in the network. We contribute to the literature by providing a network model of product assortment decision using data mining and optimization tools. We differentiate form earlier studies by providing a more realistic framework where there is the possibility of product shipments in a network of stores which hasn't been considered in the prior studies.

2 LITERATURE REVIEW

Association Rule Mining (Agrawal et al., 1993, Agrawal and Srikant, 1994, Brin et al., 1997) is the

Bhattacharjee S., Boylu F. and Gopal R..
Product Assortment Decisions for a Network of Retail Stores using Data Mining with Optimization.
DOI: 10.5220/0004106703190323
In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval* (KDIR-2012), pages 319-323
ISBN: 978-989-8565-29-7

extraction of patterns of interest among different items in a collection of transactions. Association rule mining techniques are based on the several measures of interestingness such as support, confidence and lift to gather possible association rules from the databases. However, due to the increase in the size and complexity of modern databases, the number of the association rules mined can become tremendously large. To narrow down and to optimize the list of association rules, advanced techniques have been proposed (Agarwal et al. 2000, Han 2000, Malik 2006). We adopt the methodology by Bai et al. (2012) which develops procedures for eliminating conflicting patterns for the problem under study, and differentiate between the frequent patterns and beneficial patterns.

Product assortment decision has been the attention of several studies. Caro and Gallien (2007) study the problem of dynamically optimizing the assortment of a store as more information becomes available during the selling season. They focus on one store and predict sales data to optimize the assortment. Gaur and Honhon (2006) developed an assortment planning model under the locational choice demand model. In this model, products are perceived as bundles of attributes, and each product is defined by its location in an attribute space. Individual preferences are defined by associating a point in the attribute space with each consumer.

Problem of efficient shipment of products between the stores in retail networks has also been the center of research around inventory management and distribution systems. In Caro and Gallien (2010), authors address the problem of distributing a limited amount of inventory across all the stores in a fast-fashion retail network. They formulate a stochastic model predicting the sales of a product in a single store. They formulate a mixed-integer program embedding a piecewise-linear approximation of the first model applied to every store in the network to compute store shipment quantities maximizing overall predicted sales, subject to inventory availability and other constraints.

Although, closely related to these studies, we focus on the product assortment decisions for individual stores in the network when product shipment is an option between the stores. We maximize the revenue function of the company as a whole taking into account the discount factors of possible shipments. None of the above papers just cited has brought together product assortment using patterns of beneficial product combinations with the availability of shipments in a network of stores.

3 MODEL

Let the firm offer a total of P products indexed as $p \in \{1, \dots, P\}$ in B store locations indexed as $b \in \{1, \dots, B\}$ to clients in S different industry segments indexed as $s \in \{1, \dots, S\}$. Let $i \in \{1, \dots, N\}$ index the client companies in a given industry segment across all stores. Note that each client i may operate in one or more of the stores and N represents the collection of all clients in all stores. Let π_i denote the revenue of client i from its own sales. Let $r_{i,p}$ denote the sales of product p from the firm we model to client i. We define a transaction as the vector $t_i = \{r_{i,1}, \dots, r_{i,P}\}$ which indicates sales to the client of various products.

Let $A = \{a_1, \dots, a_{|A|}\}$ denote the set of all frequent itemsets with lift values greater than 1. A transaction is denoted as supporting a frequent itemset if $r_{i,p} > 0 \ \forall \ p \in a_j$. Let $\theta_{t_i}, \theta_{t_i} \subseteq A$, denote the set of frequent itemsets that are supported by transaction t_i. Following the notation in Bai et al. (2012), we introduce two key measures: Industry coverage: $\Upsilon(p, a_j) = \frac{\sum_{\forall i, a_j \epsilon \theta_{t_i}} \pi_i}{\sum_{\forall i} \pi_i}$ and Revenue Capture Ratio: $\Phi(p, a_j) = f(\pi_i, r_{i,p} \ \forall i: a_j \epsilon \theta_{t_i})$.

Industry coverage denotes the ratio of revenues of clients who purchased the frequent itemset a_j without any conflicts with other frequent itemsets to the revenue of all potential clients around the stores of the firm. Revenue capture ratio describes the ratio of the firm's sales value of product p to the revenue of clients that purchase a_j. As a common approach, we use $\Phi(p, a_j) = \text{average}\left(\frac{r_{i,p}}{\pi_i} \ \forall i: a_j \epsilon \theta_{t_i}\right)$.

We first begin the analysis to remove conflicts in the frequent itemsets and in the process prune the number of itemsets for further analysis. Consider two frequent itemsets a_j and $a_{j'}$ in θ_{t_i} such that $a_j, a_{j'} \in \theta_{t_i}$. They are said to conflict if $a_j \cap a_{j'} \neq \emptyset$ (where $\emptyset$ denotes null or empty set). In other words, there exists at least one product that belongs to both itemsets, which raises the question of which itemset influences the sales of the products common to both itemsets for a given transaction. Clearly, a conflict removal strategy is needed to identify the main itemset that influence the sales of a set of products. Adopting the methodology by Bai et al. (2012), we remove the conflicts in frequent itemsets. We start with one frequent itemset and compare it with every other frequent itemset for the transaction. For each transaction, we terminate with a set of non-conflicting itemsets, and prune the number of itemsets for the whole set of transactions. At the end of the process, for each product in each transaction, at most one frequent itemset is utilized. This process ensures that the frequent itemset that is identified uniquely influences the products present in that itemset.

For a given store, let Π denote the total revenues of the client companies belonging to a given industry segment in that store location. We focus on deciding which products to offer (and as a result which itemsets come into play) and the corresponding branch sales figures. We define the following variables:

$x_p^{\beta b} = 1$ if the product p is shipped from branch β to b, 0 otherwise, for $\mathrm{p} = 1, \dots, \mathrm{P}$, $b = 1, \dots, \mathrm{B}$ and $\beta = 1, \dots, \mathrm{B}$

$y_j^b = 1$ if itemset $\mathrm{a_j}$ applies for branch b, 0 otherwise, for $\mathrm{j} = 1, \dots, |\mathrm{A}|$ and $b = 1, \dots, \mathrm{B}$

$y_{jp}^{\beta b} = 1$ if product $p \in \mathrm{a_j}$ is shipped from branch β to b, 0 otherwise, for $\mathrm{j} = 1, \dots, |\mathrm{A}|$ and $b = 1, \dots, \mathrm{B}$, $\beta = 1, \dots, \mathrm{B}$

$\Delta_p^{\beta b} \leq 1$ is the discount factor when product p is shipped from branch β to b, for $\mathrm{p} = 1, \dots, \mathrm{P}$, $b = 1, \dots, \mathrm{B}$ and $\beta = 1, \dots, \mathrm{B}$ where $\Delta_p^{\beta b} = \Delta_p^{b\beta}$.

Notice that when $\beta = b$, $x_p^{\beta b} = 1$ if the product p is kept in branch b and 0 otherwise. Similarly, when $\beta = b$, $\Delta_p^{\beta b} = 1$ since no discount is necessary. The total expected sales for a branch b can be written as:

$$\sum_{\mathrm{p}=1}^{\mathrm{P}} \Upsilon(\mathrm{p}, \emptyset)\Phi(\mathrm{p}, \emptyset)\Pi \sum_{\beta=1}^{B} \Delta_p^{\beta b} x_p^{\beta b} + \sum_{\mathrm{p}=1}^{\mathrm{P}} \sum_{\mathrm{j}=1}^{|\mathrm{A}|} \Upsilon(\mathrm{p}, \mathrm{a_j})\Pi\left[\Phi(\mathrm{p}, \emptyset)\left(\sum_{\beta=1}^{B} \Delta_p^{\beta b}(x_p^{\beta b} - y_{jp}^{\beta b})\right) + \Phi(\mathrm{p}, \mathrm{a_j})\sum_{\beta=1}^{B} \Delta_p^{\beta b} y_{jp}^{\beta b}\right]$$

Next, we develop the optimization model using the above sales formula for each branch to formulate the objective function.

<u>Problem K^b (no more than K^bproducts kept at each store):</u>

$$Max \sum_{b=1}^{B}\left\{\sum_{\mathrm{p}=1}^{\mathrm{P}} \Upsilon(\mathrm{p}, \emptyset)\Phi(\mathrm{p}, \emptyset)\Pi \sum_{\beta=1}^{B} \Delta_p^{\beta b} x_p^{\beta b} + \sum_{\mathrm{p}=1}^{\mathrm{P}} \sum_{\mathrm{j}=1}^{|\mathrm{A}|} \Upsilon(\mathrm{p}, \mathrm{a_j})\Pi\left[\Phi(\mathrm{p}, \emptyset)\left(\sum_{\beta=1}^{B} \Delta_p^{\beta b}(x_p^{\beta b} - y_{jp}^{\beta b})\right) + \Phi(\mathrm{p}, \mathrm{a_j})\sum_{\beta=1}^{B} \Delta_p^{\beta b} y_{jp}^{\beta b}\right]\right\}$$

s,t For $b = 1$ *to* B:

$\sum_{\beta=1}^{B} x_p^{\beta b} \leq 1$ (A product should be shipped from at most 1 other store)

$x_p^{bb} \geq x_p^{b\beta}$ for $\mathrm{p} = 1, \dots, \mathrm{P}$ $\forall \beta \in \{1, \dots, \mathrm{B}\} - \{\mathrm{b}\}$ (If a store doesn't hold a product, can't ship that product to other stores)

$\sum_{p \in a_j} \sum_{\beta=1}^{B} x_p^{\beta b} - (|a_j| - 1) \leq y_j^b, \forall a_j \epsilon A$, (all products of an itemset should be either received or held for the itemset to be in effect)

$\sum_{\beta=1}^{B} x_p^{\beta b} \geq \mathrm{y_j^b}$ $\forall p \in a_j, \forall a_j \epsilon A$ (If any of the products of an itemset are not received or kept then the itemset is not in effect)

$x_p^{\beta b} \geq y_{jp}^{\beta b}$ $, \forall p \in a_j, \forall a_j \epsilon A,$ $\forall \beta \in \{1, \dots, \mathrm{B}\}$ (When a product of an itemset is not received or held, $y_{jp}^{\beta b} = 0$)

$x_p^{\beta b} + y_j^b - 1 \leq y_{jp}^{\beta b}, \forall p \in a_j, \forall a_j \epsilon A,$ $\forall \beta \in \{1, \dots, \mathrm{B}\}$ (When an itemset is in effect, $y_{jp}^{\beta b} = 1$ for $x_p^{\beta b} = 1$)

$\sum_{\beta=1}^{B} y_{jp}^{\beta b} \leq \mathrm{y_j^b}$ $, \forall p \in a_j, \forall a_j \epsilon A$ (When an itemset is not in effect, $y_{jp}^{\beta b} = 0$)

$\sum_{p=1}^{P} x_p^{bb} \leq K^b$ (Maximum number of products that can be kept at the store)

$x_p^{\beta b}, y_j^b, y_{jp}^{\beta b} \in \{0,1\}$

Note that the above formulation is for a single industry segment. However, it can easily be generalized for multiple industry segments by summing the objective function over the industry segments and populating the constraints accordingly.

4 COMPUTATIONAL ANALYSIS AND RESULTS

In order to illustrate our model, we use a prototype example for 10 products and 3 branches for a single industry segment. After conflict removal, we extract 7 frequent non-conflicting itemsets. We randomly select the shipment discount factor $\Delta_p^{\beta b}$ in the

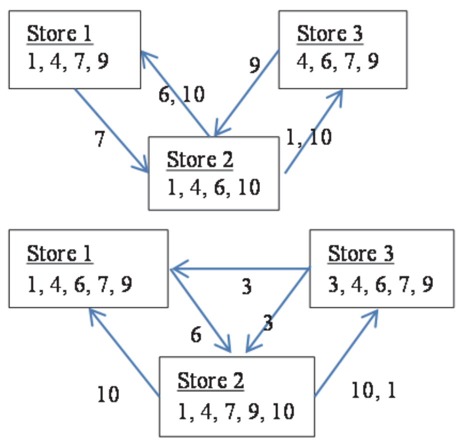

Figure 1: K^b=4 and K^b=5 shipment networks.

interval [0.5, 1] and use the same value of Π for all stores. We solve the model for different values of K^b. In Figure 1, as an example, we display the product shipment graphs for the two values of K^b=4 and 5. We observe that the shipments of products dramatically change with an increase in K^b.

Table 1 reports the total revenue and individual store revenues when maximum number of products for all stores are set to increasing values of K^b starting from $K^b = 7$ for ease of presentation.

Notice that the product shipments are not reported in Table 1 due to space limitations. When the product assortments in Table1 for increasing values of K^b are examined, we observe that the assortments show much variability from one $K^b$to another. This can be attributed to the random shipment discount factor $\Delta_p^{\beta b}$ which does not have a certain pattern. This results in variability in optimal shipment of products as the number of products per store is increased

Table 1: Revenues and product assortments.

K^b	Store	Rev.(million)	# of itemsets	Products in store	Total Rev. (million)
7	1	5.66	4	1,3,4,5,6,7,9	16.68
	2	5.69	2	1,4,6,8,9,10	
	3	5.33	4	3,4,5,6,7,8,9	
8	**1**	**5.66**	**4**	**1,3,4,5,6,7,8, 9**	**16.69**
	2	**5.70**	**2**	**1,3,4,6,7,8,9,10**	
	3	**5.33**	**4**	**3,4,5,6,7,8,9,10**	
9	1	5.61	7	1,2,3,4,5,6,7,8,9	16.57
	2	5.65	5	1,2,3,4,6,7,8,9,10	
	3	5.30	7	2, 3, 4, 5, 6, 7, 8, 9, 10	

We observe that the optimal revenue occurs at K^b= 8. To obtain values for K^b= 9, we force the K^b constraint in the model to an equality. We see that the objective value for K^b= 9 is less than the optimal K^b= 8. In Figure 2, we display the shipment graph of this optimal network.

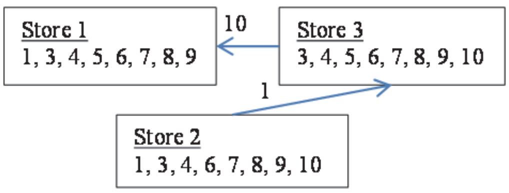

Figure 2: Optimal shipment network.

In Table 1, when the numbers of itemsets are compared, we see that they saturate around the optimal value at 4, 2 and 4 for three stores of the company. We note that for larger number of products and higher number of stores, the optimal graph will look more interesting as the shipment of multiple products will be in effect and the related revenues and assortments will show more variability as multiple industry segments are considered. The results of such a dataset from an industry leading plastics manufacturer and retailer in the United States will be demonstrated at the time of presentation.

5 CONCLUSIONS

In this paper, we present a model for product assortment optimization for a network of retail stores operating in various locations of a company. We combine local information captured from each retail store and use a global frequent itemset analysis. Later, for each retail store, our optimization model determines the right products to include in a store's assortment and which stores to ship from in the store network. The model first learns the global patterns of the frequent itemsets based on association rule mining to extract patterns of products with corresponding sales benefits. It then uses a global optimization formulation maximizing the revenue of the company in aggregate and identifies the optimal assortment for each local store by taking into account the possibility of shipments in the network.

REFERENCES

Agarwal, R., Imielinski, T., Swami, A., 1993, Mining association Rules Between Sets of Items in massive databases, *Proceedings of the ACM/SIGMOD International Conference on Management of Data*, pp. 207-216

Agrawal, R., Srikant. R. (1994) Fast algorithms for mining association rules. *Proceedings of the 20th International Conference on Very Large Data Bases*, VLDB, pp. 487-499.

Agarwal, R., Aggarwal C. C., Prasad V. V. V. (2000). A tree projection algorithm for generation of frequent itemsets. *Journal of Parallel and Distributed Computing*. 61. pp. 350-371

Bai, X., Bhattacharjee, S., Boylu, F., Gopal, R. (2012) A Data Mining and Optimization Methodology for Improving Branch Product Portfolio and Performance, under review

Brin, S., Motwani, R., Ullman, Tsur. J. D., (1997). Dynamic itemset counting and implication rules for market basket data. *In SIGMOD 1997, Proceedings ACM SIGMOD International Conference on Management of Data*, pp. 255-264.

Caro, F., J. Gallien. (2007). Dynamic assortment with demand learning for seasonal consumer goods. *Management Science*. 53(2) 276–292.

Caro, F., J. Gallien. (2010). Inventory Management of a Fast-fashion Retail Network. *Operations Research*. 58(2) 250–273.
Gaur, V., D. Honhon. (2006). Assortment planning and inventory decisions under a locational choice model. *Management Science*. 52(10) 1528-1543.
Han, J. A. (2000). Mining frequent patterns by pattern growth: Methodology and Implications. *ACM SIGKDD Explorations Newsletter*. 2(2) pp. 14-20.
Malik, H. H. (2006). Clustering web images using association rules, interestingness measures, and hypergraph partitions. *Proceedings of the 6th International Conference on Web Engineering* (ICWE '06). pp. 48-55.