Available online at www.sciencedirect.com
ScienceDirect
Expert Systems with Applications 32 (2007) 976–986
Expert Systems with Applications
www.elsevier.com/locate/eswa

A data mining approach to product assortment and shelf space allocation

Mu-Chen Chen [a,*], Chia-Ping Lin [b]

[a] *Institute of Traffic and Transportation, National Chiao Tung University, 4F, No. 118, Section 1, Chung Hsiao W. Road, Taipei 10012, Taiwan, ROC*
[b] *SAS, Taiwan, Taipei, Taiwan, ROC*

Abstract

In retailing, a variety of products compete to be displayed in the limited shelf space since it has a significant effect on demands. To affect customers' purchasing decisions, retailers properly make decisions about which products to display (product assortment) and how much shelf space to allocate the stocked products (shelf space allocation). In the previous studies, researchers usually employed the space elasticity to optimize product assortment and space allocation models. The space elasticity is usually used to construct the relationship between shelf space and product demand. However, the large number of parameters requiring to estimate and the he non-linear nature of space elasticity can reduce the efficacy of the space elasticity based models. This paper utilizes a popular data mining approach, association rule mining, instead of space elasticity to resolve the product assortment and allocation problems in retailing. In this paper, the multi-level association rule mining is applied to explore the relationships between products as well as between product categories. Because association rules are obtained by directly analyzing the transaction database, they can generate more reliable information to shelf space management.

Keywords: Shelf space management; Data mining; Multi-level association rules; Zero–one integer programming

1. Introduction

Most retailers nowadays face challenges such as how to respond consumer's ever-changing demands and how to adapt themselves to keen competition in dynamic market. Retail management is to develop a retail mix to satisfy customers' demands and to affect customers' purchasing decisions. The factors in retail mix include store location, product assortment, pricing, advertising and promotion, store design and display, services and personal selling (Levy & Weitz, 1995). Shelf space is an important resource for retail stores since a great quantity of products compete the limited shelf space for display. Retailers need frequently make decisions about which products to display (assortment) and how much shelf space to allocate these products (allocation) (Borin & Farris, 1995; Borin, Farris, & Freeland, 1994). Product assortment and shelf space allocation are two important issues in retailing which can affect the customers' purchasing decisions. Through the proficient shelf space management, retailers can improve return on inventory and consumer's satisfaction, and therefore increase sales and margin profit (Yang, 1999).

In the past two decades, numerous models and solution approaches have been developed to deal with product assortment and/or shelf space allocation problems (Anderson & Amato, 1974; Borin & Farris, 1995; Borin et al., 1994; Brijs, Goethals, Swinnen, Vanhoof, & Wets, 2000; Brijs, Swinnen, Vanhoof, & Wets, 1999; Bultez & Naert, 1988; Bultez, Naert, Gijbrechts, & Abeele, 1989; Corstjens & Doyle, 1981; Corstjens & Doyle, 1983; Hansen & Heinsbroek, 1979; Urban, 1998; Yang, 1999). In these previous studies, the individual space elasticity and the

* Corresponding author. Tel.: +886 2 2771 2171x3417; fax: +886 2 2776 3964.
E-mail address: bmcchen@ntut.edu.tw (M.-C. Chen).

0957-4174/$ - see front matter © 2006 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2006.02.001

cross-elasticity between products are usually applied to estimate the relationship between shelf space and demands. Traditionally, researchers apply the space elasticities to determine which products to stock and how much shelf space to display these products. However, there are two major limitations that reduce the effectiveness of the space elasticity (Borin & Farris, 1995; Borin et al., 1994). First, due to the non-linear nature of space elasticity, the space elasticity based models are very complicated, and the specific solution approach is developed for each model. Additionally, it is necessary to estimate a large number of parameters by using the space elasticity.

Recently, the progress of information technology makes retailers easily collect daily transaction data at very low cost. Through the point of sale (POS) system, a retail store can collect a large volume of transaction data. From the huge transaction database, a great quantity of useful information can be extracted to support the retail management. Data mining is frequently adopted to discover the valuable information from the huge database. In data mining, association rule mining is widely applied to market basket analysis or transaction data analysis (Agrawal, Imielinski, & Swami, 1993; Srikant & Agrawal, 1997). This study proposes a data mining approach to make decisions about which products to stock, how much shelf space allocated to the stocked products and where to display them. Association rules are generated by directly analyzing the transaction database, and these rules can be used to effectively resolve the product assortment and shelf space allocation problems. This study applies the association instead of the space elasticity to formulate the mathematical model for product assortment. In this paper, multi-level association rules are generated to express the relationships between products and product categories to allocate the products selected in the assortment stage.

2. Literature review

In retailing, shelf space management refers a routine decision-making on product assortment and space allocation (Borin & Farris, 1995; Borin et al., 1994). Product assortment planning is the process to determine the number and types of products in a line, which is accomplished by retailers (Rajaram, 2001). Product assortment should meet the marketing strategy of retailers, and maintain the sustainable competitive advantages that retailers build up. After the stage of product assortment, the display spaces for the products selected from assortment are then determined. Shelf space is one of the most essential resources in logistic decisions and shelf space management (Yang & Chen, 1999), and the high-quality space allocation can attract more consumers. In practice, product assortment and shelf space allocation are usually resolved simultaneously.

Previously, several models and solution approaches have been developed to resolve the product assortment and/or the shelf space size determination problems (Anderson & Amato, 1974; Borin & Farris, 1995; Borin et al., 1994; Brijs et al., 2000; Brijs et al., 1999; Bultez & Naert, 1988; Bultez et al., 1989; Corstjens & Doyle, 1981; Corstjens & Doyle, 1983; Hansen & Heinsbroek, 1979; Urban, 1998; Yang, 1999). In the literature, the space elasticity has been widely used to estimate the relationship between sales and allocated space. Space elasticity is a ratio of relative change of sales to relative change of display space.

The measurement of space elasticity can be divided into two types: direct elasticity (main effect) and cross-elasticity (cross-effect) (Borin & Farris, 1995; Borin et al., 1994; Bultez & Naert, 1988; Bultez et al., 1989; Chrhan, 1973; Corstjens & Doyle, 1981; Corstjens & Doyle, 1983; Hansen & Heinsbroek, 1979; Urban, 1998). Direct elasticity is designed to measure the effect on demand by changing the display space for an individual product. The increase of display space for a product may stimulate the demand of products, but in turn, it may decrease the demand of substitute and/or complementary products. Cross-elasticity is used to measure the effect on demand of substitute and/or complementary products by changing the display space of an individual product. The mathematical form of space elasticity is then transformed into the optimization model to select products to display and determine shelf space size to these products. Experimental designs have been applied to measure the space elasticity. Due to the estimation of a large number of parameters, only one or a small number of products cab be considered in most experiments in a store (Borin & Farris, 1995; Borin et al., 1994; Corstjens & Doyle, 1981; Corstjens & Doyle, 1983; Urban, 1998).

Anderson and Amato (1974), took only the direct elasticity into their model to simultaneously optimize the product assortment and shelf space allocation. Anderson and Amato formulated the shelf space management model as a knapsack problem. Hansen and Heinsbroek (1979) also estimated the demand of products by direct space elasticity, and constructed optimization models to select and allocate products. In their models, profit, inventory cost, and cost for allocating a product on a shelf were taken into consideration. The total profit of a retail store was taken as the objective function.

The models presented by Corstjens and Doyle (1981, 1983) took advantage of both direct space elasticity and cross-space elasticity to estimate demands. Corstjens and Doyle (1981) applied a polynomial functional form of demand, and they found a set of solutions by using signomial geometric programming. Zufryden (1986) extended the concept of Corstjens and Doyle (1981) and applied the dynamic programming to solve the shelf management problem. In Zufryden's model, the integer solutions can be provided because it allows the consideration of general objective function requirements.

Borin et al. (1994) and Borin and Farris (1995) simultaneously optimized the product assortment and space allocation problems in which the cross-elasticity effects are considered. In their constrained optimization models, objective function is the return on investment of inventory. Due to the complexity of model and non-linearity of objective function, a meta-heuristic, simulated annealing, was adopted to generate solutions. A critical drawback for applying this model is that it needs to estimate a large number of parameters. The number of estimated parameters in Borin et al. (1994) is $2n + n^2$, in which n is the number of possible products. Rajaram (2001) applied demand forecasts derived from historical sales patterns, and also constructed a non-linear integer-programming model to make the product *assortment* planning. Due to the high complexity in the model, heuristics were by Rajaram developed to resolve this problem.

Although some existing product assortment and space allocation models (e.g., Borin & Farris, 1995; Borin et al., 1994) use return on inventory as the objective and take stockouts into consideration, they do not explicitly include the conventional inventory control decisions as variables (Urban, 1998). Urban (1998) integrated the inventory control model with the product assortment and space allocation models. In addition, a greedy search and a genetic algorithm were developed to resolve the integrated model. Hwang, Choi, and Lee (2005) also proposed an integrated mathematical model, which combines the shelf space allocation model and inventory-control model with the objective of maximizing the retailer's profit. Due to the complexity of the integrated model, Hwang et al. proposed a gradient search heuristic and a genetic algorithm to resolve the model. Using a series of field experiments to study the impact of shelf positioning and facing allocations on sales of individual items, Drèze, Hoch, and Purk (1994) concluded that location had a large impact on sales, whereas changes in the number of facings allocated to a brand had much less impact as long as a minimum stock is maintained.

Except for shelf space allocated to products, other factors such as product price and shelf location have effects on sales. McIntyre and Miller (1999) simultaneously determined what items to stock and how to price the stocked items in retailing. McIntyre and Miller developed a non-parametric approach to deal with the product assortment and pricing problems. Hwang et al. (2005) assumed that the level of shelf on which the product is displayed significantly influences the sales of products. Yang (1999) constructed a knapsack model for the shelf space allocation problem in which factors of display space and shelf location are taken into consideration. Yang additionally proposed a heuristic for the solution to the knapsack problem, which allocates shelf space with respect to a descending order of sales profit for each item per display length. Nogales and Suarez (2005) specifically studied the influence of store brand in shelf space management through a case study using direct shelf observation. Additionally, other aspects of assortment, prices and promotions have also been analyzed to construct their relationship with shelf space management.

To overcome the high cost of conducting experiments to measure parameters in space elasticities, Brijs et al. (1999) proposed an association rule based approach, namely PROFSET (PROFitability per Set), to resolve the product assortment problem in convenience stores. Brijs et al. took the advantage of association rules to develop the product assortment model. Additionally, they considered the profit of cross-selling and store image in terms of basic products in the model. The products that conform to the store's image and characteristics are called basic products, while the other products are called added products. Brijs et al. (2000) further generalized the PROFSET model to deal with large baskets and category management in practice. However, Brijs et al. (1999) and Brijs et al. (2000) only explored the product assortment problem. Therefore, they did not take the shelf space requirement of selected products.

By using space elasticity for solutions to product assortment and shelf space allocation, it needs to estimate a great quantity of parameters to obtain space elasticity. Such an estimation procedure results in high cost and errors in the mathematical model. The previous space elasticity based models do not take the shelf location into consideration. Although the shelf location is considered in Yang's approach (1999), it allows one product to appear on two or more locations on different shelves. It is different from the practice of retailing which usually displays product according to product category.

3. The development of shelf space management model

With the rapid development of information technology, transaction data can be easily collected through POS system. The relationships between products hidden in transaction data can be discovered through association rule mining to assist product assortment and shelf space allocation. By using association rule mining, the shelf space management model can directly apply transaction data in a retail store for analysis. It is not necessary to conduct a series of experiments to estimate a great quantity of parameters in space elasticities. This study develops a data mining approach to make decisions about which products to stock, how much shelf space allocated to the stocked products and where to display them.

The proposed procedure of shelf space management begins with multi-level association rule mining from transaction data to obtain relationships between product items, between product subcategories and between product categories. Next, the procedure proceeds to product assortment in which the profits of frequent itemsets are considered. The products and categories frequently bought together can be displayed together. Finally, the product display locations are determined by considering the relationships between categories, subcategories and between items. The flowchart of the proposed approach is schematically illustrated in Fig. 1.

3.1. Multi-level association rules

In the stage of product display, the relationships between categories, between subcategories and between product

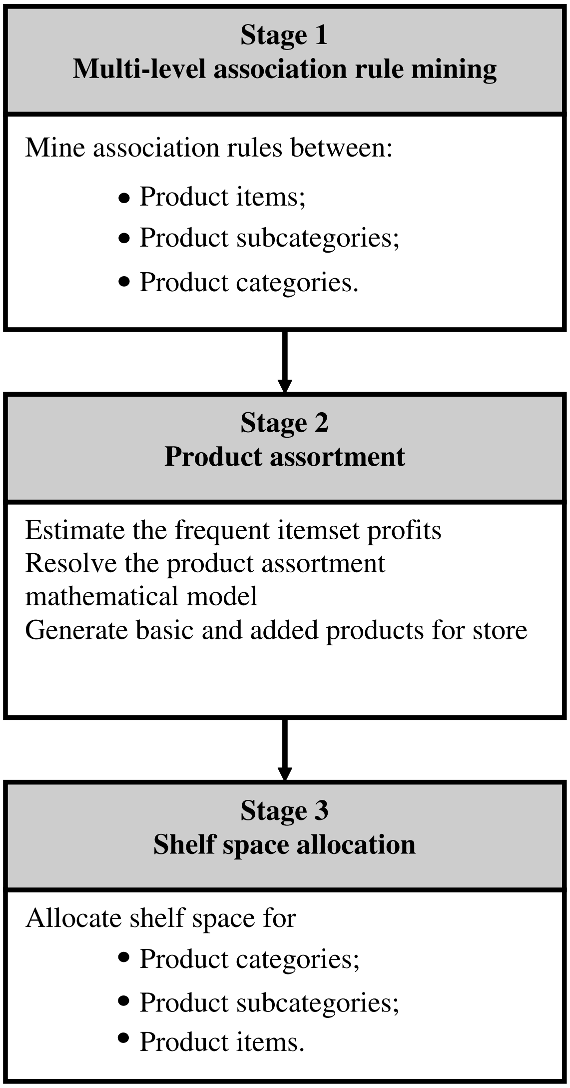

Fig. 1. Flowchart of the proposed approach.

items are utilized to plan product display. Therefore, multi-level association rules between product items, subcategories and categories are discovered in this paper. The problem of mining association rules involves generating all association rules that have support and confidence greater than the user-specified minimum support and minimum confidence, respectively (Agrawal et al., 1993). With a huge quantity of data being constantly collected and stored in business, since they are easy to comprehend and implement (Agrawal et al., 1993; Srikant & Agrawal, 1997). Due to the huge valuable data stored in enterprise information system, the applications of association rules in marketing (Chen, Chiu, & Chang, 2005; Cho, Cho, & Kim, 2005; Wang, Chuang, Hsu, & Keh, 2004), logistics (Chen, Huang, Chen, & Wu, 2005; Chen & Wu, 2005), medicine (Tang, Jin, & Zhang, 2005) and manufacturing (Chen, 2003; Del Castillo Sobrino & Barrios, 1999) are increasing.

The frequent itemset, frequent subcategory and frequent category are utilized in the shelf space management model. For many real world applications, due to the sparsity in retail transaction data, there exist relatively few frequent itemsets for products. The level of product category is higher than subcategory, and the level of subcategory is higher than item. The strategy of reduced minimum support is generally used in mining multi-level association rules (Han & Kamber, 2001). The lower the abstraction level, the smaller the corresponding minimum support. Therefore, the support threshold of category is largest; the support threshold of item is least. The product assortment model in the study takes the association between items as the basis for selecting products to fill in shelf space, while it additionally takes the associations between categories and between subcategories as the basis for determining product display locations.

The problem of mining association rules was first presented in Agrawal et al. (1993). The problem is formally stated as follows (Agrawal et al., 1993; Srikant & Agrawal, 1997). Let $I = \{i_1, i_2, \ldots, i_m\}$ denote a set of literals, namely items. Moreover, let $\mathbf{D}$ represent a set of transactions, where each transaction T is a set of items such that $T \subset I$. A unique identifier, namely TID, is associated with each transaction. A transaction T is said to contains $\mathbf{X}$, a set of some items in I, if $\mathbf{X} \subseteq T$. An association rule is an implication of the form $\mathbf{X} \Rightarrow \mathbf{Y}$, where $\mathbf{X} \subset I$, $\mathbf{Y} \subset I$ and $\mathbf{X} \cap \mathbf{Y} = \varnothing$. The rule $\mathbf{X} \Rightarrow \mathbf{Y}$ holds in the transaction set $\mathbf{D}$ with *confidence*, c, if c% of transactions in $\mathbf{D}$ that contain $\mathbf{X}$ also contains $\mathbf{Y}$. The rule has *support*, s, in the transaction set $\mathbf{D}$ if s% of transactions in $\mathbf{D}$ contain $\mathbf{X} \cup \mathbf{Y}$.

The proposed shelf space management approach based on multi-level association rules applies the Apriori algorithm to extract *frequent itemsets*, *frequent subcategory sets* and *frequent category sets*. The Apriori algorithm is an efficient algorithm for mining association rules. It is implemented in a specific way in the shelf space management in this paper. The details of mining association rules can be found in Agrawal et al. (1993), Srikant and Agrawal (1997) and Han and Kamber (2001).

3.2. The product assortment procedure

In this study, association rule mining is conducted with the store transaction data. The association rules obtained from the analysis can specify which products are frequently bought by customers at the same market basket (frequent itemsets). With the estimated gross margin of frequent itemsets, the profit of selected product mix can be obtained. This study maximizes the profit of selected product mix under the constraint of available shelf space. The product assortment model is constructed as a zero–one integer program.

3.2.1. Profit estimation of frequent itemsets

The profit estimation follows the idea of Brijs et al. (2000). Not only the individual profit generated by that product is considered in evaluating a product value, but the cross-selling effects with other products in the assortment are also taken into account. Brijs et al. developed a profit allocation method to estimate the margin of transaction from various frequent itemsets of that transaction. Their method of estimating profit of frequent itemsets is described as follows:

T_j items included in the jth transaction
F_I the collection of all frequent itemsets of T_j
X a frequent itemset in the jth transaction
$X_{\max}$ the maximal frequent itemset in the jth transaction
$Y_{\max}$ the second maximal frequent itemset in the jth transaction
$\Theta_{T_j}(X)$ the probability of selecting X in T_j to allocate gross margin, $\Theta_{T_j}(X_{\max}) = \frac{\text{Support}(X_{\max})}{\sum_{\forall Y_{\text{MAX}}} \text{Support}(Y_{\max})}$
Support(X) support of X
$T_j \backslash X$ items included in the jth transaction after excluding frequent itemset X
$m(X)$ the profit of products in frequent itemsets X
$M(X)$ the summation of $m(X)$

The procedure of estimating the profit of frequent itemset is described as follows:

Step 1: Input the transaction database, collection of frequent itemsets and gross margin of items.
Step 2: For each transaction T_j in transaction database,
(a) If $X = T_j$, the profit $m(X)$ is the profit of product multiplies numbers bought in transaction record T_j. Set $M(X) = M(X) + m(X)$.
(b) Otherwise, the profit $m(X)$ from frequent itemsets $X_{\max}$ in T_j based on the probability Θ_{T_j}. Set $M(X) = M(X) + m(X)$. Repeat this substep, if $T_j \backslash X$ still has frequent itemsets.
Step 3: Return $M(X)$ for all frequent itemsets.

3.2.2. The product assortment model

The product assortment problem addressed in this study maximizes the total profit of products, which follows the concept of Brijs et al. (2000). However, Brijs et al. only addressed the product assortment problem. The necessary shelf space size of selected products is not taken into account in their model. Furthermore, the model of Brijs et al. cannot ensure the selection of basic products to conveying the store's image. In this paper, the amount of selected products is restricted due to the limit of shelf space in a retail store. A certain amount of product items for each category should be selected for displaying on shelf. Additionally, the basic products should be selected in the assortment stage.

Before introducing the product assortment model, the notation is firstly given below.

Model parameters
C the set of categories
S the set of subcategories
I the set of items
F_{C} the collection of frequent category sets
F_{S} the collection of frequent subcategory sets
SFC_i the set of subcategories included in the ith frequent category set
IFI_i the set of items included in the ith frequent itemset
IC_k the set of items included in kth category
A the set of added products
B the set of basic products
b_{jk} $= \begin{cases} 1, & \text{Item } j \text{ in Category} k \text{ is a basic product;} \\ 0, & \text{Otherwise.} \end{cases}$
$G_i(X)$ the estimated gross margin of the ith frequent itemset
h_{jk} the inventory and handling costs of Item j in Category k
f_{jk} the product facing length of Item j in Category k
q_{jk} the minimum quantity of the selected Item j in Category k
S_k The total shelf space allocated to Category k
N_k the minimum number of items in Category k selected for displaying

Decision variables

p_i $= \begin{cases} 1, & \text{if any item in frequent itemset IFI}_i \text{ is selected;} \\ 0, & \text{otherwise.} \end{cases}$

d_{jk} $= \begin{cases} 1, & \text{if Item } j \text{ in Category } k \text{ is selected;} \\ 0, & \text{otherwise.} \end{cases}$

The mathematical model of product assortment is as follows:

$$\text{Maximize} \quad \sum_i G_i(X)p_i - \sum_k \sum_j h_{jk}d_{jk} \tag{1}$$

Subject to:

$$\sum_{j \in \text{IC}_k} d_{jk}q_{jk}f_{jk} \leqslant S_k, \quad \forall k; \tag{2}$$

$$d_{jk} \geqslant p_i, \quad \forall i, \ \forall k, \ \forall j \in \text{IFI}_i; \tag{3}$$

$$\sum_j d_{jk} \geqslant N_k, \quad \forall k; \tag{4}$$

$$d_{jk} \geqslant b_{jk}, \quad \forall k, \ \forall j \in \text{IC}_k; \tag{5}$$

$$p_i \in \{0, 1\}; \quad d_{jk} \in \{0, 1\}. \tag{6}$$

The model is described as follows. The objective function expressed in Eq. (1) is to maximize the total profit of products. Constraint (2) specifies the maximum shelf space allowed to be displayed for each product category. Constraint (3) ensures that once a frequent itemset is selected, the products in this frequent itemset have to be selected. Constraint (4) specifies the maximum number of items in each category can be selected for displaying on shelf. Constraint (5) ensures that the basic product items (B) of the store have to be selected in the optimization procedure. The basic products are identified by retailers to express their store image. The added product items (A) can be selected by taking the frequent itemsets and profit into account. Constraint (6) limits the decision variables to be binary. The above model for product assortment is a zero–one integer program.

3.3. The shelf space allocation procedure

After the products are selected from the assortment procedure, the allocation procedure is adopted to assign the assorted products to the shelf space. In this paper, the proposed space allocation procedure takes shelf levels and associations between categories, between subcategories and between product items into consideration. Retailers usually adopt the grid display to allocate the shelf space (Levy & Weitz, 1995). The grid display has the longer demonstration shelf and walkway. Therefore, the grid display shown in Fig. 2 is used in this paper. Each shelf is divided into three levels: high profit, middle profit and low profit. One product may have different sales if it is displayed on different levels. In this paper, the profit weights of high, middle and low shelf levels are assumed to be 2/6, 3/6 and 1/6, respectively. The shelf space allocation procedure only takes the facing width into account. The length and depth of shelf and product are disregarded.

The proposed approach allocates products on shelf according to average profit, association among categories and shelf profit weight. The product with the higher profit is allocated to the shelf with the higher weight in order to increase the sales and profit. Additionally, products are allocated closer if they have higher supports. Before introducing the shelf space allocation procedure, several primary principles are firstly presented as follows:

1. To allocate frequent categories as close as possible, or on the same shelf, if possible;
2. To allocate frequent subcategories as close as possible, or on the same shelf, if possible;
3. To allocate the product items in the same frequent itemset and in the same category as close as possible, or on the same shelf, if possible;
4. To allocate product items of the same category on the same area, if possible;
5. To allocate product items of the same subcategory on the same shelf, if possible;
6. The product with the higher profit is allocated to the shelf with the higher weight, if possible.

The proposed allocation procedure based on multi-level association rules mainly includes three steps. The first step is to allocate the shelf space for product category; the second step is to allocate shelf space for subcategory and the third step is to allocate shelf space for product item. Before presenting the proposed allocation procedure, some additional notation is firstly listed as follows:

SC_k	the set of subcategories included in the kth category
IS_l	the set of items included in the lth subcategory
IFI_i	the set of items included in the ith frequent itemset
f_j	the product facing length of Item j
q_j	the minimum quantity of the jth selected item
p_j	the profit of the jth selected item
PC_k	the average profit per shelf space for the kth category
PS_l	the average profit per shelf space for the lth subcategory
PI_j	the average profit per shelf space for the jth selected item

The proposed allocation procedure is described as follows:

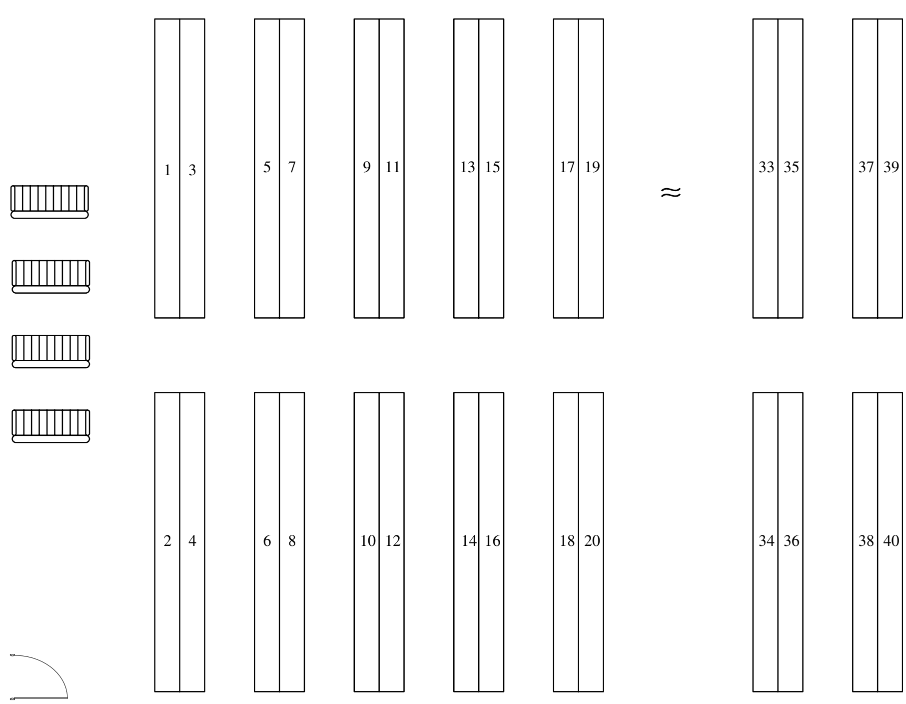

Fig. 2. The grid display of retailing.

Step 1: Sequentially allocate the categories.

(a) Join the categories included in each frequent category set to be a virtual category by considering the support of frequent category set.

(b) Keep those categories that are not included in any frequent category set (non-frequent category).

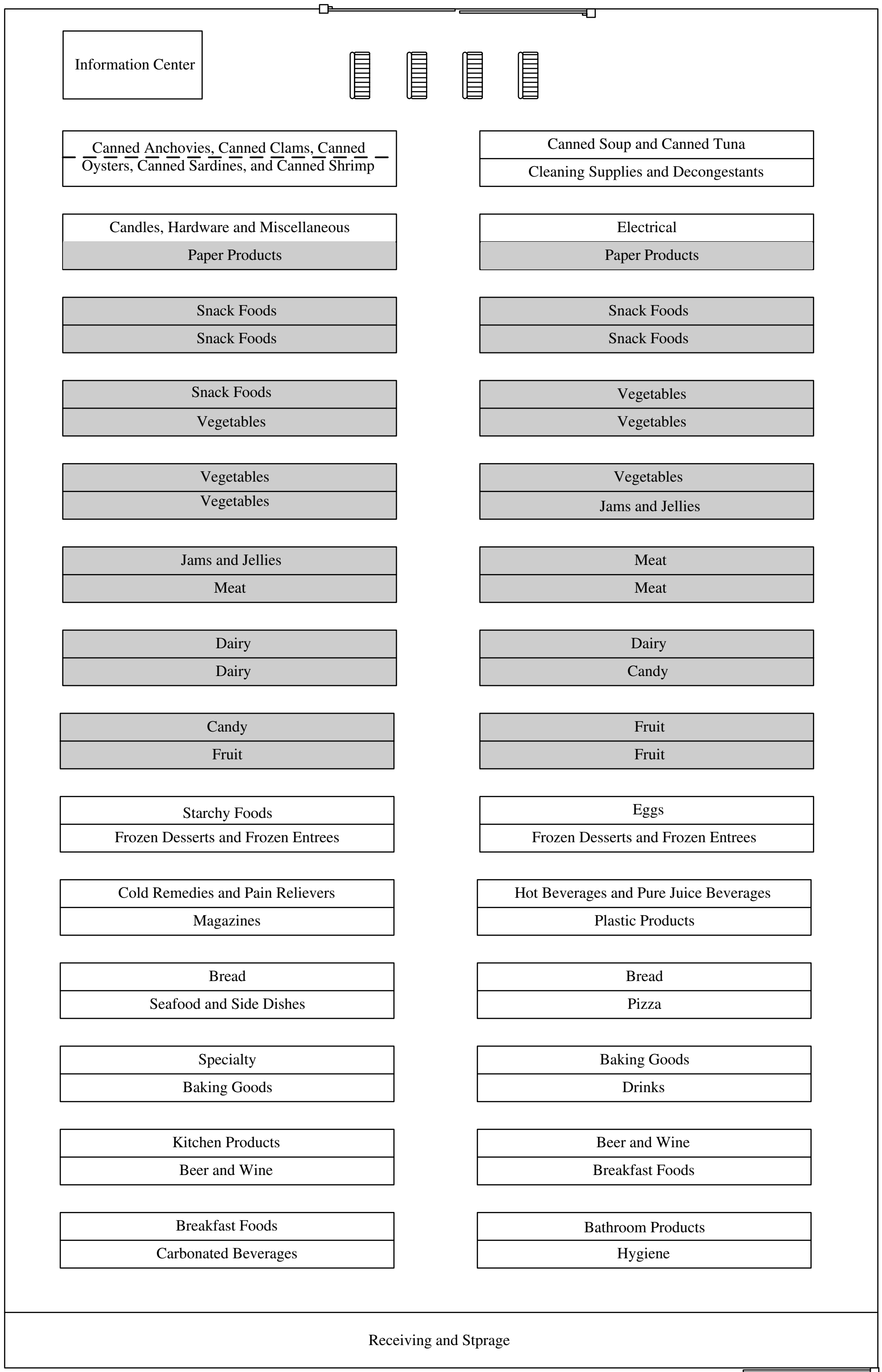

Fig. 3. Allocation of product categories.

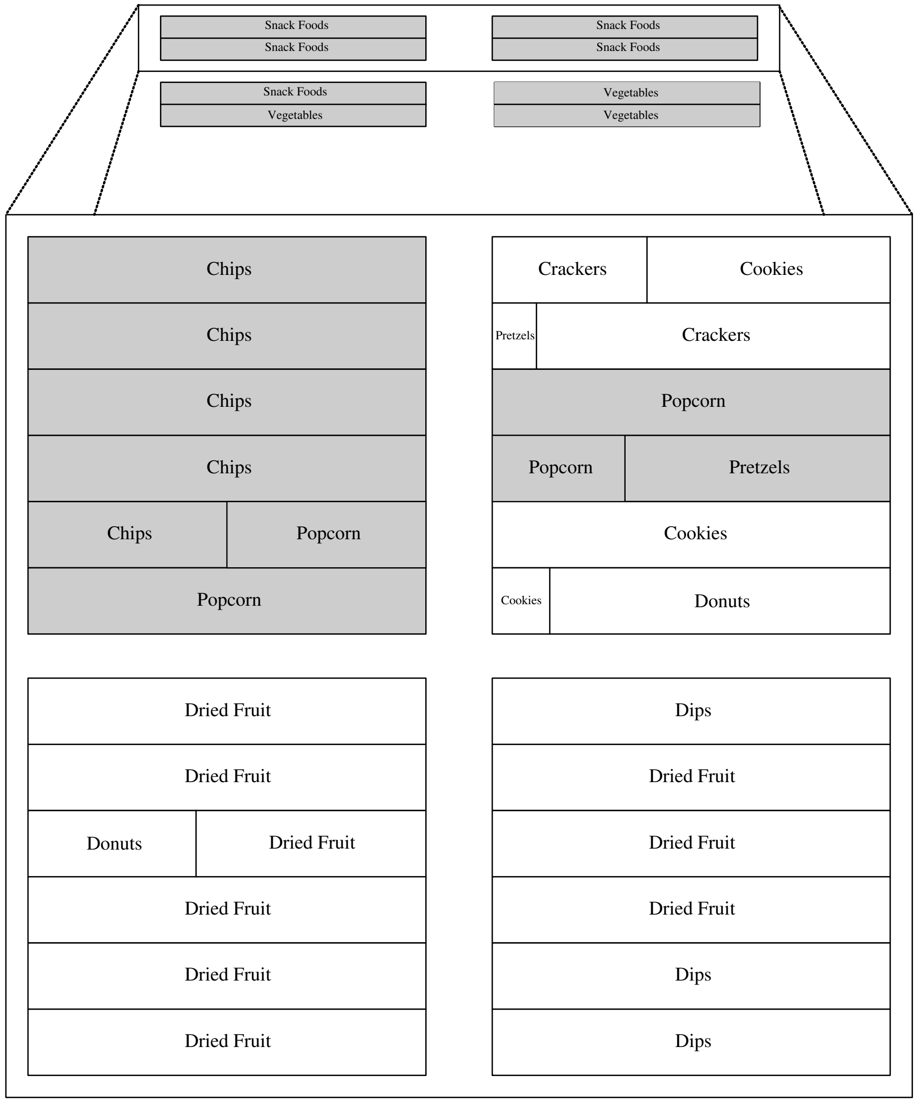

Fig. 4. Partial allocation of product subcategories for Shelves 1–4 in snack food.

(c) For the kth category, the average profit per shelf space is computed as $\mathrm{PC}_k = \frac{1}{|\mathrm{IC}_k|}\left(\sum_{j\in \mathrm{IC}_k} \frac{p_j}{f_j}\right)$.

(d) Sequentially allocate the categories (virtual categories and non-frequent categories) on shelf space with respect to shelf profit weight and average profit per shelf space. The more profitable category is allocated on the shelf with a higher weight.

Step 2: Sequentially allocate the subcategories.

(a) Within each category, join the subcategories included in each frequent subcategory set to be a virtual subcategory by considering the support of frequent subcategory set.

(b) Keep those subcategories that are not included in any frequent subcategory set (non-frequent subcategory).

(c) For the lth subcategory, the average profit per shelf space is calculated as $\mathrm{PC}_l = \frac{1}{|\mathrm{IS}_l|}\left(\sum_{j\in \mathrm{IS}_l} \frac{p_j}{f_j}\right)$. Sequentially allocate the subcategories (virtual subcategories and non-frequent subcategories) on shelf space with respect to shelf profit weight and average profit per shelf space. Within each category, the more profitable subcategory is allocated on the shelf with a higher weight.

Step 3: Sequentially allocate the items.

(a) Within each subcategory, join the product items included in each frequent itemset to be a virtual item by considering the support of frequent itemset. Keep those product items that are not included in any frequent itemset (non-frequent item).

(b) For the jth item, the average profit per shelf space is calculated as $\mathrm{PI}_j = \frac{1}{|\mathrm{IFI}_i|}\left(\sum_{j\in \mathrm{IFI}_i} \frac{p_j}{f_j}\right)$.

(c) The required shelf space of each virtual item and non-frequent item is set to $f_j \times q_j$.

(d) Sequentially allocate the items (virtual subcategories and non-frequent subcategories) on shelf space with respect to shelf significance weight and average profit per shelf space. Within each subcategory, the more profitable subcategory is allocated on the shelf with a higher weight.

Step 4: Return the product allocation.

4. The implementation

4.1. Data and assumptions

The proposed data mining based procedure for product assortment and allocation is implemented with an example of a retail store. The database for implementation is derived from Foodmart in Microsoft SQL Sever 2000. The database includes product data, customer data and transaction records. There are 1560 product items, which are divided into 45 categories and 102 subcategories. To conduct the implementation, some additional assumptions are made as follows:

1. In the supermarket, the grid display is adopted. The size of shelf is 300×210 cm (width $\times$ height) with six layers. Fig. 2 illustrates the top-view of shelf display in this implementation.
2. This study neglects the height and depth of products and considers only facing width of products.
3. The product items are divided into two types of basic product and added product. Basic products aims to build up the characteristics of store and have to be selected for displaying. Added products are included to increase the product varieties in addition to basic products. There are 804 basic products as well as 756 added products in the store.

4.2. Product assortment

After the data being transformed into the format, which can be read by Apriori algorithm, the multi-level association rules of product category, subcategory and item are discovered for product assortment and allocation. The transaction data in this study is sparse since there are 6568 transaction records with 1560 product items. There may exist relatively few frequent itemsets for product items. The reduced minimum support is used in mining multi-level association rules. The lower the abstraction level, the smaller the corresponding minimum support and minimum confidence. The minimum supports and minimum confidences for subcategory and item are set to a very low threshold in order to discover enough association rules for further analysis. The minimum supports for

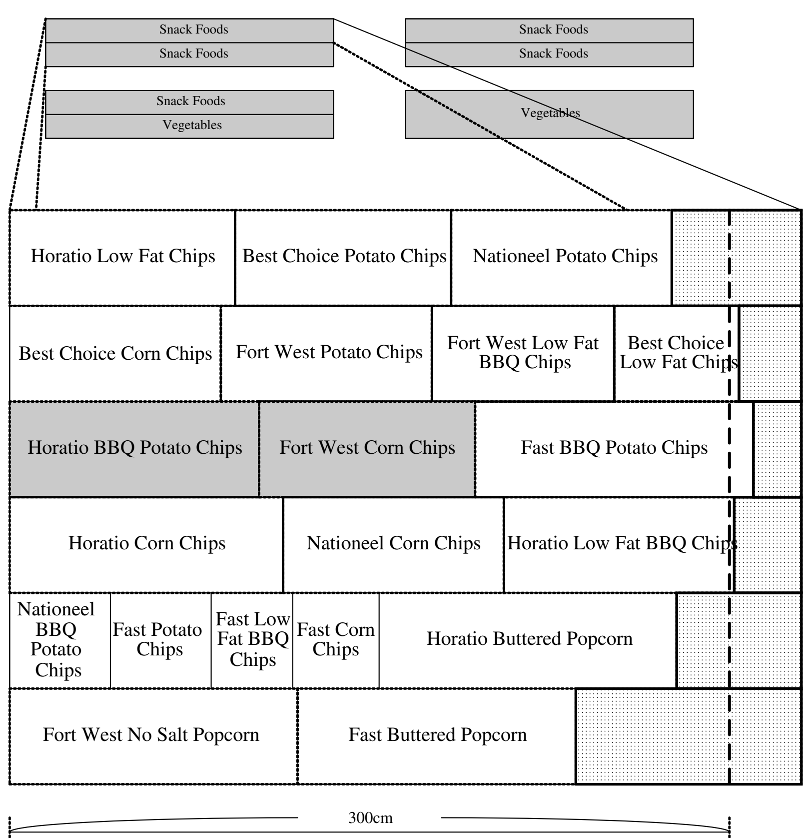

Fig. 5. Partial allocation of product items for Shelf 1 in snack food.

product category, subcategory and item are, respectively, set to 30%, 1% and 1%. The minimum confidences for product category, subcategory and item are set to 10%.

For the above threshold setting, there are 54, 54 and 3238 rules, respectively, for product category, subcategory and item. After mining the multi-level association rules, the margin gross profits of frequent itemsets are estimated by the approach discussed in Section 3. The mathematical model of assortment (refer to Eqs. (1)–(6)) is resolved by using ILOG CPLEX. Totally, 894 items are selected for allocating on shelf.

4.3. Shelf space allocation

From the results obtained from the product assortment model, the shelf space for each category can then be generated in production allocation stage. The product categories, subcategories and items with high associations are allocated as close as possible to increase the cross-selling effects. The product category allocation is schematically illustrated in Fig. 3.

After allocating the shelf space for categories, the product allocation procedure then proceeds to the allocation of subcategories and items. Taking the snack food category as an example, the allocations subcategories and items are partially illustrated in Figs. 4 and 5. In this paper, the width of each shelf space is increased by 10% to easily tolerate and adjust the allocation.

5. Conclusions

To face the keen competition in retail market, retailers need to accurately and quickly respond the dynamic customers' requirements. Shelf space management is an important issue to keep the competitive advantage in retailing sector. Retailers can try to satisfy the diverse customers' demands and to affect customers' purchasing decisions by using the systematic approach for product assortment and allocation. With the rapid development of information technology, retailers have put a huge amount of transaction data in storage, and they potentially can be used to support shelf space management. This paper develops a data mining based approach to simultaneously make decisions about which products to stock, how much shelf space allocated to the stocked products and where to display them. There exist some advantages in the proposed product assortment and space allocation approach. Firstly, because association rules are obtained by directly analyzing the transaction database, therefore they are reliable for shelf space management. Secondly, the massive estimation of parameters in space elasticity can be eliminated, and the estimation error and costly experiment can thus be reduced. Thirdly, association rules can quickly respond to market changes since the transaction data are timely collected by retailer's POS system. Forth, the assortment model ensures to include the basic products for expressing the store's image, and the added products are determined by using the associations between product items. Fifth, by mining the multi-level association rules, retailers can allocate the product categories, subcategories and items with respect to their associations and profits.

Acknowledgment

The authors would like to thank the National Science Council of the Republic of China, Taiwan for financially supporting this research under Contract No. NSC 94-2416-H-027-004.

References

Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In *ACM SIGMOD conference*, *Washington*, *DC*, *USA* (pp. 254–259).

Anderson, E. E., & Amato, H. N. (1974). A mathematical model for simultaneously determine the optimal brand-collection and display-area allocation. *Operations Research, 22*(1), 13–21.

Borin, N., & Farris, P. W. (1995). A sensitivity analysis of retailer shelf management models. *Journal of Retailing, 71*(2), 153–171.

Borin, N., Farris, P. W., & Freeland, J. R. (1994). A model for determining retail product category assortment and shelf space allocation. *Decision Science, 25*(3), 359–384.

Brijs, T., Goethals, B., Swinnen, G., Vanhoof, K., & Wets, G. (2000). A data mining framework for optimal product selection in retail supermarket data: the generalized PROFSET model. In *KDD-2000*, *Boston*, *MA*, *USA* (pp. 300–304).

Brijs, T., Swinnen, G., Vanhoof, K., & Wets, G. (1999). Using association rules for product assortment decisions: a case study. In *KDD-99*, *San Diego*, *CA*, *USA* (pp. 254–260).

Bultez, A., & Naert, P. (1988). SHARP: shelf allocation for retailers' profit. *Marketing Science, 7*(3), 211–231.

Bultez, A., Naert, P., Gijbrechts, E., & Abeele, P. V. (1989). Asymmetric cannibalism in retail assortment. *Journal of Retailing, 65*(2), 153–192.

Chen, M.-C. (2003). Configuration of cellular manufacturing systems using association rule induction. *International Journal of Production Research, 41*(2), 381–395.

Chen, M.-C., Chiu, A.-L., & Chang, H.-H. (2005). Mining changes in customer behavior in retail marketing. *Expert Systems with Applications, 28*(4), 773–781.

Chen, M.-C., Huang, C.-L., Chen, K.-Y., & Wu, H.-P. (2005). Aggregation of orders in distribution centers using data mining. *Expert Systems with Applications, 28*(3), 453–460.

Chen, M.-C., & Wu, H.-P. (2005). An association-based clustering approach to order batching considering customer demand patterns. *Omega—International Journal of Management Science, 33*(4), 333–343.

Cho, Y. B., Cho, Y. H., & Kim, S. H. (2005). Mining changes in customer buying behavior for collaborative recommendations. *Expert Systems with Applications, 28*(2), 359–369.

Chrhan, R. C. (1973). Shelf space allocation and profit maximization in mass retailing. *Journal of Marketing, 37*(1), 54–60.

Corstjens, M., & Doyle, P. (1981). A model for optimizing retail space allocations. *Management Science, 27*(7), 822–833.

Corstjens, M., & Doyle, P. (1983). A dynamic model for strategically allocating retail space. *Journal of Operational Research Society, 34*(10), 943–951.

Del Castillo Sobrino, M. D., & Barrios, B. L. J. (1999). Knowledge acquisition from batch semiconductor manufacturing data. *Intelligent Data Analysis, 3*(5), 399–408.

Drèze, X., Hoch, S. J., & Purk, M. E. (1994). Shelf management and space elasticity. *Journal of Retailing, 70*(4), 301–326.

Han, J., & Kamber, M. (2001). *Data mining: Concepts and techniques*. San Francisco: Morgan Kaufmann.

Hansen, P., & Heinsbroek, H. (1979). Product selection and space allocation in supermarket. *European Journal of Operational Research, 3*(6), 474–484.

Hwang, H., Choi, B., & Lee, M.-J. (2005). A model for shelf space allocation and inventory control considering location and inventory level effects on demand. *International Journal of Production Economics, 97*(2), 185–195.

Levy, M., & Weitz, B. A. (1995). *Retailing management*. Chicago: Irwin.

McIntyre, S. H., & Miller, C. M. (1999). The selection and pricing of retail assortments: an empirical approach. *Journal of Retailing, 75*(3), 295–318.

Nogales, A. F., & Suarez, M. G. (2005). Shelf space management of private labels: a case study in Spanish retailing. *Journal of Retailing and Consumer Services, 12*(3), 205–216.

Rajaram, K. (2001). Assortment planning in fashion retailing: methodology, application and analysis. *European Journal of Operational Research, 129*(1), 186–208.

Srikant, R., & Agrawal, R. (1997). Mining generalized association rules. *Future Generation Computer Systems, 13*(2–3), 161–180.

Tang, Y., Jin, B., & Zhang, Y.-Q. (2005). Granular support vector machines with association rules mining for protein homology prediction. *Artificial Intelligence in Medicine, 35*(1–2), 121–134.

Urban, T. (1998). An inventory theoretic approach to product assortment and shelf-space allocation. *Journal of Retailing, 74*(1), 15–35.

Wang, Y.-F., Chuang, Y.-L., Hsu, M.-H., & Keh, H.-C. (2004). A personalized recommender system for the cosmetic business. *Expert Systems with Applications, 26*(3), 427–434.

Yang, M.-H. (1999). An efficient algorithm to allocate shelf space. *European Journal of Operational Research, 131*(1), 107–118.

Yang, M.-H., & Chen, W.-C. (1999). A study on shelf space allocation and management. *International Journal of Production Economics, 60–61*, 309–317.

Zufryden, F. S. (1986). A dynamic programming approach for product selection and supermarket shelf-space allocation. *Journal of Operational Research Society, 37*(4), 413–422.