Using Association Rules for Product Assortment Decisions: A Case Study

Tom Brijs Gilbert Swinnen Koen Vanhoof Geert Wets

Limburg University Centre

Department of Applied Economic Sciences

B-3590 Diepenbeek

BELGIUM

{tom.brijs;gilbert.swinnen;koen.vanhoof;geert.wets}@luc.ac.be

ABSTRACT

It has been claimed that the discovery of association rules is well-suited for applications of market basket analysis to reveal regularities in the purchase behaviour of customers. Moreover, recent work indicates that the discovery of *interesting* rules can in fact only be addressed within a microeconomic framework. This study integrates the discovery of frequent itemsets with a (microeconomic) model for product selection (PROFSET). The model enables the integration of both quantitative and qualitative (domain knowledge) criteria. Sales transaction data from a fully-automated convenience store is used to demonstrate the effectiveness of the model against a heuristic for product selection based on product-specific profitability. We show that with the use of frequent itemsets we are able to identify the cross-sales potential of product items and use this information for better product selection. Furthermore, we demonstrate that the impact of product assortment decisions on overall assortment profitability can easily be evaluated by means of sensitivity analysis.

Keywords

association rules, frequent itemset, product assortment decisions

1. INTRODUCTION

In the past, retailers saw their job as one of buying products and putting them out for sale to the public. If the products were sold, more were ordered. If they did not sell, they were disposed of. Blischok [7] describes retailing in this model as a *product-oriented* business, where talented merchants could tell by the look and feel of an item whether or not it was a winner. In order to be successful, retailing today can no longer be just a product-oriented business. According to Blischok, it must be a *customer-oriented* business and superior customer service comes from superior knowledge of the customer. It is defined as the understanding of all customer's purchasing behaviour as revealed through his or her sales transactions, i.e. *market basket analysis*.

Currently, the gradual availability of cheaper and better information technology has in many retail organisations resulted in an abundance of sales data. Hedberg [17] mentions the American supermarket chain 'Wal-Mart' which stores about 20 million sales transactions per day. This explosive growth of data leads to a situation in which retailers today find it increasingly difficult to obtain the right information, since traditional methods of data analysis cannot deal effectively with such huge volumes of data. This is where knowledge discovery in databases (KDD) comes into play.

Today, among the most popular techniques in KDD, is the extraction of association rules from large databases. While many researchers have significantly contributed to the development of efficient association rule algorithms [1-3, 10, 21, 26], literature on the use of this technique in concrete real-world applications remains rather limited [4, 5, 25]. Nevertheless, the widespread acceptance of association rules as a valuable technique to solve real business problems will largely depend on the successful application of this technique on real-world data. Moreover, it has been claimed recently [18] that the *utility* of extracted patterns (such as association rules) in decision-making can only be addressed within the microeconomic framework of the enterprise. This means that a pattern in the data is interesting only to the extent in which it can be used in the decision-making process of the enterprise to increase utility.

In this paper, we tackle the problem of product assortment analysis and introduce a concrete microeconomic integer-programming model for product selection (PROFSET[1]) based on the use of frequent itemsets. We demonstrate its effectiveness on real-world sales transaction data obtained from a fully-automated convenience store.

The paper is organised as follows. Section 2 provides an overview of association rules. In section 3, we present the product selection problem and introduce a product selection model based on the use of frequent itemsets. In section 4, we present the results of the empirical study. Finally, section 5 summarises our work and presents directions for future research.

KDD-99 San Diego CA USA

[1] PROFSET uses the PROFitability per frequent SET to determine the optimal selection of products in terms of maximal total profit

2. ASSOCIATION RULES: AN OVERVIEW

A recent data mining technique for market basket analysis is *association rules* introduced by Agrawal et al. [1]. They provided the following formal description of this technique:

Let $I = \{i_1, i_2, \ldots, i_k\}$ be a set of literals, called items. Let D be a set of transactions, where each transaction T is a set of items such that $T \subseteq I$. Associated with each transaction is a unique identifier, called its *TID*. We say that a transaction *T contains X*, a set of some items in I, if $X \subseteq T$. An *association rule* is an implication of the form $X \Rightarrow Y$, where $X \subset I$, $Y \subset I$, and $X \cap Y = \varnothing$. The rule $X \Rightarrow Y$ holds in the transaction set D with *confidence c* if c% of transactions in D that contain X also contain Y. The rule $X \Rightarrow Y$ has *support s* in the transaction set D if s% of transactions in D contain $X \cup Y$. Given a set of transactions D, the problem of mining association rules is to generate all association rules that have support and confidence greater than a user-specified minimum support (*minsup*) and minimum confidence (*minconf*).

Generating association rules involves looking for so-called *frequent itemsets* in the data [19]. Indeed, the support of the rule $X \Rightarrow Y$ equals the frequency of the itemset $\{X, Y\}$. Thus by looking for frequent itemsets, we can determine the support of each rule.

Definition 1 Frequency of an itemset (adapted from [19])

s(X, D) represents the frequency of itemset X in D, i.e. the fraction of transactions in D that contain X.

Definition 2 Frequent itemset (adapted from [19])

An itemset X is called frequent in D, if $s(X, D) \geq \sigma$ *with* σ *the minsup.*

A typical approach [2] to discover all frequent sets X is to use the knowledge that all subsets of a frequent set are also frequent. This insight simplifies the discovery of all frequent sets considerably. Once all frequent sets are known, finding association rules is easy. Namely, for each frequent set X and each $Y \in X$ verify whether the rule $X \setminus \{Y\} \Rightarrow Y$ has sufficiently high confidence.

To summarise, the technique of association rules produces a set of rules describing underlying purchase patterns in the data, like for instance *bread* ⇒ *cheese* [support = 20%; confidence = 75%]. Informally, support of an association rule indicates how frequent that rule occurs in the data. The higher the support of the rule the more prevalent the rule is. Confidence is a measure of the reliability of an association rule.

3. PROFSET: A PRODUCT SELECTION MODEL

3.1 Problem Situation

Determining the *ideal* product assortment has been (and still is) the dream of every retailer. From the marketing literature [24] it is known that the optimal product assortment should meet two important criteria.

Firstly, the assortment should be *qualitatively* consistent with the store's image. A store's image distinguishes the retailer from its competition and is projected through its design, layout, services and of course its products. Therefore, retailers often distinguish between *basic* products and *added* products. Basic products are products that should not be deleted from the assortment because they are the core materialisation of the retailer's store formula. For example, for a typical convenience store, customers expect at least beverages, cigarettes, food and candy products in the assortment. Therefore, such products should not be removed, otherwise the assortment will not meet the basic expectations of customers who visit the store. In contrast, *added* products are chosen by the retailer to confirm the store image even more and should be selected as to maximise cross-sales potential with *basic* products. Indeed, retailers are interested in adding items whose sales will not be made at the expense of currently stocked items but may help increase the sales of other items (sales complements) [22]. For the convenience store, examples may include cigarette lighters, coffee whitener or tea warmers. This means that *added* products should be selected by the model based on their purchase affinity with *basic* products.

Secondly, because retailing organisations are profit seeking companies, the product assortment should be *quantitatively* appealing in terms of the profitability it generates for the retailer (i.e. the microeconomic framework). In section 3.2 this quantitative element will be further defined.

In contrast to the criteria described above, product assortment decisions are often taken based on rules of thumb; if products generate high sales, more of them is ordered, in contrast, when sales of some products do not meet the expectations, they are deleted from the assortment. In the past, marketing researchers have attempted to construct optimisation models for more *objective* product selection [16, 27, 28] and shelf space allocation [6, 8, 11-16, 23, 27, 28]. Unfortunately, the estimation of shelf space elasticities (for instance obtained through expert interviewing), estimation of substitution and complementarity effects between products and the high cost of conducting experiments to measure these parameters have rendered systematic use of these models impossible in practice. According to [8] the number of parameters required for their shelf management model is $2n + n^2$ where n is the number of items in the assortment. Thus, even for a moderately sized assortment the number of required parameters becomes large very quickly. Moreover, rapidly changing market situations require parameter estimates to be updated continually. While these models, from a theoretical point of view, have contributed a lot to the understanding of product assortment decisions, the above limitations urge for a more flexible and easy to carry out method for product selection.

3.2 Product selection based on frequent sets

According to the problem situation described above, a model must be constructed that is able to select a *hitlist* of products, i.e. a selection of a user-defined number of products from the assortment which yields the maximum overall profit, taking into account background knowledge of the retailer. A simple solution to this problem, which is often used in practice, is to calculate the total profit contribution generated *per product* and then select those products, in addition to the *basic* products that have already been selected by the retailer, that contribute the most to the overall profitability. We call this the product-specific profitability heuristic. Although easy to calculate, it does not take cross-selling effects of products into account. In contrast, the PROFSET model, introduced in this paper, implicitly takes into account cross-selling effects by using frequent itemsets. Before specifying the microeconomic optimization model formally, we will first introduce the parameters and components of the model.

3.2.1 Model parameters

Gross Margin:

Let: T_j be an individual sales transaction at time j

SP_i be the selling price of product i

PP_i be the purchase price of product i

f_i be the number of times product i was purchased in T_j

Definition 3 m_{Tj} is the gross margin generated by sales transaction T_j

$$\forall\, T_j : \; m_{Tj} = \sum_{\in\, T_j} (SP_i - PP_i) * f_i$$

Definition 4 M_X is the gross margin of frequent itemset X

$$\forall\, X : \; M_X = \sum_{j=1}^{\#\,transactions} m_j \quad \text{with} \quad \begin{cases} m_j = m_{Tj} & \text{if } X = T_j \\ m_j = 0 & \text{otherwise} \end{cases}$$

It is important to understand why X must equal T_j for m_j to be non-zero. The reason is that we will use the sum of all M_x to approximate the total profitability of the assortment. Now, suppose that $m_j \neq 0$ when $X \subseteq T_j$ instead of $X = T_j$ with $\{i_1, i_2\}$ a frequent itemset and $\{i_1, i_2, i_4\}$ a sales transaction. Clearly, $\{i_1, i_2\} \subseteq \{i_1, i_2, i_4\}$ but, because $\{i_1, i_2\}$ is frequent, it is known [2] that $\{i_1\}$ and $\{i_2\}$ must also be frequent. Consequently, $\{i_1\} \subseteq \{i_1, i_2, i_4\}$ and $\{i_2\} \subseteq \{i_1, i_2, i_4\}$ and thus the gross margin generated by sales transaction $\{i_1, i_2, i_4\}$ will add to $M_{\{i1,i2\}}$, $M_{\{i1\}}$ and $M_{\{i2\}}$ even if i_4 is not selected for inclusion in the hitlist. Thus, if $m_j \neq 0$ when $X \subseteq T_j$, then a single sales transaction increases the M_X parameter of *all* the frequent itemsets that are contained in that transaction. To summarise, a single sales transaction is allowed to contribute to the total profitability only once through the M_X parameter of the *frequent itemset* that contains the same items as those included in that transaction.

Cost of products:

Also product handling and inventory costs should be included in the model. Product handling costs refer to costs associated with the physical handling of goods. Inventory costs include costs of re-stocking which are a function of replenishment frequency and the lead-time of the orders. In practice, however, these costs are often difficult to obtain, especially product handling costs. For reasons of simplicity, we assume that a total cost figure c_i per product i can be obtained for all products.

3.2.2 Model components

The PROFSET optimisation problem is operationalized by means of an integer-programming model containing two important components:

Objective function:

The objective function represents the goal of the optimisation problem and therefore reflects the microeconomic framework of the retail decision maker. It is constructed in order to maximise the overall profitability of the hitlist. Frequent itemsets X and their associated gross margins M_X contribute in a positive sense to the objective function. In contrast, individual products i and their associated cost c_i contribute in a negative sense.

Constraints:

(1) Because the final decisions need to be taken on the product level instead of on the *frequent itemset* level, we must specify which products i are included in each frequent itemset X. This information can be obtained from association rule mining.

(2) *Basic* products can be specified by forcing the model to select certain products.

(3) The *size* of the hitlist is specified by the *ItemMax* constraint.

3.2.3 Model specification

$$\text{Max } Z = \sum_{X=1}^{\#\,frequent\ sets} M_X * X \; - \sum_{i=1}^{\#\,products} c_i * i$$

s.t.

$$\forall\, X, \forall\, i \in X : i \geq X \tag{1}$$

$$\forall\, i \mid i \text{ is } basic \text{ product} : i = 1 \tag{2}$$

$$\sum_{i=1}^{\#\,products} i = ItemMax \tag{3}$$

where i and X are booleans.

By using frequent itemsets the objective function will give a lower bound, i.e. the *observed* amount of profit will be higher than indicated by the value of the objective function. The reason is that we consider frequent itemsets and thus *in*frequent itemsets will not add to the total profit amount in the objective function. This is however justified because it is highly probable that infrequent itemsets exist because of random purchase behaviour. Consequently, we claim that the objective function only measures the profit from structural, underlying purchase behaviour.

	Total margin	Position	Prod profit.	PROFSET
Tobacco brand 'x'	10353 BEF	17	•	•
Cigarette paper brand 'y'	3258 BEF	66	°	•

Table 1: Total margin, position and selection for tobacco and cigarette paper

4. EMPIRICAL STUDY

The empirical study is based on a data set of 27,148 sales transactions acquired from a fully- automated convenience store over a period of 5.5 months in 1998. The concept of the fully-automated convenience (FAC) store is closely related to that of the vending machine. However, as opposed to the product assortment of the typical vending machine, this new retail store offers a wider variety of products. Typically, a selection of about 200 products is included ranging from the typical product categories such as beverages, food, candy and cigarettes, to products like healthcare, petfood, fruit, batteries, film supplies (camera, roll of film), which are presented to the customer by means of a 8 m² window display. The product assortment of the store under study consists of 206 different items. However, the average sales transaction contains only 1.4 different items because in convenience stores customers typically do not purchase many items during a single shopping visit. With regard to the costs of each individual product in the assortment, detailed information on handling and inventory costs could not be obtained so these will be considered equal for all products and therefore these costs are not included in the model.

Basically, the empirical study involves two important phases. In the first phase, structural purchase behaviour under the form of frequent itemsets is discovered by using the data mining technique of association rules (section 4.1). Then, in the second phase, the PROFSET method is used to select a hitlist of products from the assortment (see section 4.2).

4.1 Mining For Association Rules

As the objective function in the PROFSET method requires frequent itemsets as input, frequent itemsets and association rules were discovered from the database. An *absolute* support of 10 was chosen. This means that no item or set of items will be considered frequent if it does not appear in at least 10 sales transactions. As a consequence, we consider all itemsets *X* as non-frequent, i.e. describing random purchase behaviour, if the itemset appears in less than 10 rows in the sales-transaction database. It could be argued that the choice for this support parameter is rather subjective. This is partially true, however, domain knowledge from the retailer can often indicate what level of support may be considered as relevant. Furthermore, within relatively small intervals, the model will be insensitive to alterations of the minimum support threshold. The reason is that when gross margins of products are within a relatively small range, frequent itemsets with relatively low support will not be able to influence the objective function. From the analysis, 523 frequent itemsets were obtained of size 1 or 2 with absolute *support* ranging from 10 to 2833. The size of the *frequent* itemsets is rather small; this can however be explained by the small size of the average sales transaction. Although the model does not use association rules as input, i.e. it uses only frequent itemsets, the discovery of association rules will be helpful for interpreting the output of PROFSET, as will be explained in the next section.

4.2 Product Selection (PROFSET)

In order to make the comparison between PROFSET and the product specific profitability heuristic straightforward, we chose not to specify *basic* products in the model. Consequently, the model will be able to fully exploit cross-sales potential between items in the assortment without any restrictions. Furthermore, any *ItemMax* value can be selected to constrain the number of items for inclusion in the hitlist. Experiments with the PROFSET method indicate that, for this data set, important results can be identified:

1. Using PROFSET, some products with relatively low product-specific profitability but considerably high cross-selling effects are selected for the hitlist.
2. The PROFSET method enables to assess the sensitivity of product assortment decisions and, as a result, allows to identify the importance of the impact of such decisions on the total profitability of the hitlist.

Observation 1

We demonstrate observation 1 by a concrete example obtained from the empirical results with both product selection methods. Consider the products *tobacco brand 'x'* and *cigarette paper brand 'y'*. Table 1 (above) illustrates the total margin[2] of each product and its according position (with regard to the total margin) within the entire product assortment. Table 1 shows that from the product-specific point of view, *tobacco* is the 17th most profitable product in the assortment and *cigarette paper* ends up 66th. So, when the maximum number of products allowed in the hitlist is less than 66, according to the product-specific profitability heuristic, *cigarette paper* will not be included (see column 4). In contrast, experiments showed that for *ItemMax* equal to 35, the PROFSET method selects both products for inclusion in the hitlist (see column 5). This indicates that *cigarette paper* must have considerable cross-selling opportunities with one or more products that are included in the hitlist. In fact, this information can easily be obtained from examination of the association rules (section 4.1) as provided on the following page:

[2] Total profit margin = number of items sold x unit profit margin of the product in Belgian Francs (BEF)

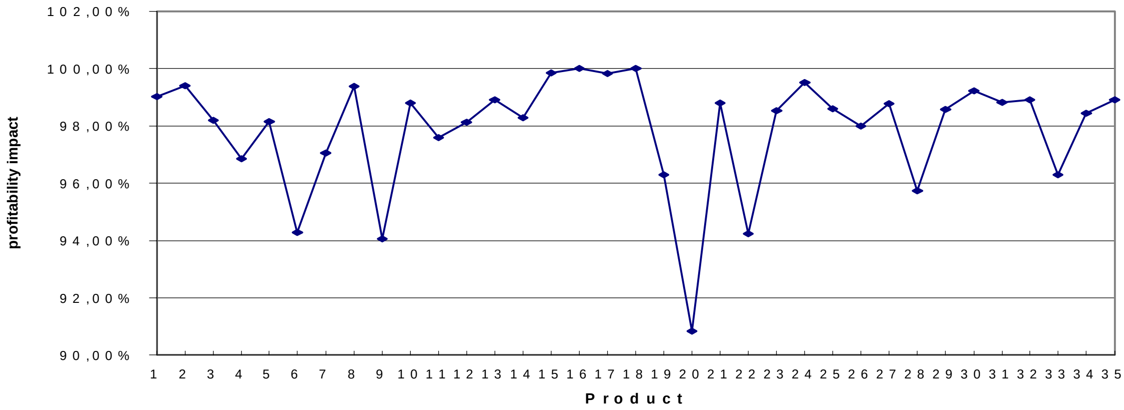

Figure 1: Profitability impact of replacement decisions

Cigarette paper ⇒ tobacco [absolute sup = 291, conf = 1.00]

Tobacco ⇒ cigarette paper [absolute sup = 291, conf = 0.82]

These rules demonstrate that whenever a customer buys *cigarette paper*, he/she also buys *tobacco* (confidence = 100%) and that when a customer buys *tobacco* he will often also buy *cigarette paper* with it (confidence 82%). A more formal method [9] to assess the dependence between two or more products is *interest*.

Definition 5: Interest

$$\frac{s\,(X \Rightarrow Y)}{s\,(X) * s\,(Y)}$$

The nominator s ($X \Rightarrow Y$) measures the observed frequency of the co-occurrence of the items in the antecedent (X) and the consequent (Y) of the rule. The denominator s (X) * s (Y) measures the expected frequency of the co-occurrence of the items in the antecedent and the consequent of the rule if both itemsets were conditionally independent. Table 2 illustrates the three possible outcomes for the interest measure and their associated economic interpretation for the dependence between the items in the antecedent and consequent of the rule.

Outcome	**Interpretation**
Interest > 1	Complementarity effects between X and Y
Interest = 1	Conditional independence between X and Y
Interest < 1	Substitutability[3] effects between X and Y

Table 2: Economical interpretation of *interest*

[3] Recall that substitutability indicates less than the expected level of mutual support.

With regard to the products *tobacco* and *cigarette paper* interest is equal to 76.7 >> 1 which indicates very strong complementarity effects between the two products. As a consequence, when treated together, these two products may represent a high total profit contribution indicating that it may be advised to select both products for the hitlist instead of selecting only *tobacco*. Indeed, the total profit contribution of the frequent itemset {*tobacco*, *cigarette paper*} equals 8875 BEF, which makes this sales combination the 10th most profitable frequent itemset, and therefore PROFSET selected both products for inclusion in the hitlist.

However, not all product combinations with high cross-selling potential are necessarily included in the hitlist. The profit contribution of the sales combination must be sufficiently high for the items to be included in the hitlist. For instance, the itemset {toothpaste, toothbrush} has an interest of 2468 >> 1 (extremely high) and, according to the association rules generated in section 4.1, they are always bought together. However, the support count of the itemset is equal to 11 (slightly above the minimum absolute support threshold). As a consequence, the total profit contribution of this itemset (415 BEF) is insufficient to influence the product selection process. Indeed, this again illustrates that the microeconomic framework of the retailer directly determines the *interestingness* of the associations. Some associations (see example toothpaste and toothbrush) are very interesting from the statistical point of view (i.e. strong dependency between both items) but it is the microeconomic framework of the retailer that ultimately determines the profitability and thus the *real* interestingness of the association.

Observation 2

Concerning observation 2, the impact on total profitability caused by product assortment decisions can easily be assessed by means of sensitivity analysis. When for instance product i is deleted from the optimal set, and it is replaced by the best product i' outside the hitlist, its impact on profitability can easily be observed through the *shadow prices* of the optimisation model.

Figure 1 (presented above) illustrates the profit impact of the replacement of each product in the hitlist. While most product replacements have only minor profit implications (-2 %), some products (6, 9, 20, 22) represent major profit drivers that should not be deleted from the hitlist. This insight can help retailers to quantitatively evaluate product assortment changes. Furthermore, the replacement of items in the PROFSET model is based on dynamic reselection of products whereas for the product-specific profitability heuristic the product that replaces the exiting product will always be the one with the highest product-specific margin outside the list. If this entrant happens to have no or small cross-selling effects with the items inside the hitlist, selecting a product with a lower product-specific profitability but higher cross-selling effects with items inside the hitlist could be more appropriate.

In fact, experiments revealed that for 43% of the products, dynamic reselection of items with the PROFSET model resulted in better overall profit compared to replacement with the product-specific profit heuristic. Obviously, the more cross-selling effects exist in the product assortment, the more impressive the profit improvement of the dynamic reselection will be, when compared to the product-specific profitability heuristic.

5. CONCLUSIONS AND FUTURE RESEARCH

5.1 Conclusions

In this paper, we proposed a microeconomic model for product selection based on the use of frequent itemsets obtained from association rule mining. More specifically, we integrated the notion of frequent sets with some important microeconomic parameters that are often used by retailers to support their product selection decision-making process into an integer programming model. We used sales transaction data from a fully-automated convenience store to empirically validate our approach against a frequently used heuristic for product selection. Results indicated that our model is able to identify cross-selling effects implicitly by using frequent itemsets, instead of having to estimate cross-selling parameters explicitly (like it is often done in product selection and shelf-space allocation models). We also showed that with our model, sensitivity analysis could easily be carried out, enabling the retailer to quantitatively assess the profitability impact of product assortment decisions. Furthermore, the retailer's background knowledge with regard to *basic* products can easily be incorporated in the model by means of additional constraints.

Yet, the retailer should also consider the following limitation. The presented model is deterministic in nature. This means that the model assumes that when for itemset $\{X, Y\}$ one of the items X or Y is not selected by the model, consequently all sales related to this itemset will be lost. This is of course too simplistic as customers do not always purchase certain product combinations intentionally. Therefore, it may well be that a fraction of the sales related to that itemset may still be recovered.

5.2 Future Research

Three main topics will be issues for further research.

Firstly, we want to assess our model on supermarket sales transaction data. It is expected that cross-selling effects are more manifestly present in supermarket data because consumers typically visit supermarkets to do one-stop-shopping.

Secondly, when sales transaction data from multiple stores with different product assortments but more or less the same underlying purchase behaviour can be obtained, it is possible to use the PROFSET method to construct an *ideal composite* product assortment. Indeed, when certain product combinations demonstrate to be very successful, the best product combinations obtained from multiple stores could be integrated in one *ideal* product assortment.

Finally, instead of including only gross margins from transactions for which the items contained in that transaction equally match the items in the frequent set (i.e. $X = T_j$), an alternate model would be to split the gross margin among all frequent itemsets that are contained in the transaction. While this may not influence the results for the current case study (since the average transaction length was only 1.4), the alternate model may be able to capture a higher percentage of transactions in sales data with higher transaction length (since the model will cover a higher percentage of transactions). However, the crucial point then is how much of the gross margin of a transaction should be allocated to each of the frequent sets that are contained in that transaction. Especially, the problem of frequent sets that are overlapping each-other in the same transaction poses significant problems.

6. REFERENCES

[1] Agrawal, R., Imielinski, T., and Swami, A. Mining association rules between sets of items in large databases. In Buneman, P., and Jajodia, S., (eds.). Proceedings of ACM SIGMOD Conference on Management of Data, 1993 (SIGMOD'93), 207-216.

[2] Agrawal, R., Mannila, H., Srikant, R., Toivonen, H., and Verkamo, A. Fast discovery of association rules. In Fayyad, U., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R., (eds.). Advances in Knowledge Discovery and Data Mining: AAAI Press, 1996, 307-328.

[3] Agrawal, R., and Srikant, R. Fast algorithms for mining association rules. In Bocca, J., Jarke, M., Zaniolo, C., (eds.). Proceedings of the 20th International Conference on Very Large Databases, 1994 (VLDB'94), 487-499.

[4] Ali, K., Manganaris, S., and Srikant, R. Partial classification using association rules. In Heckerman, D., Mannila, H., Pregibon, D., and Uthurusamy, R. (eds.). Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, 1997 (KDD'97), 115-118.

[5] Anand, S., Hughes, J., Bell, D., and Patrick, A. Tackling the cross-sales problem using data mining. In Lu, H., Motoda, H., and Liu, H. (eds.). Proceedings of the First Pacific-Asia Conference on Knowledge Discovery and Data Mining, 1997 (PAKDD'97), 331-343.

[6] Anderson, E. An analysis of retail display space: theory and methods. In Journal of Business, 52, 103-118, 1979.

[7] Blischok, T. Every transaction tells a story. In Chain Store Age Executive with Shopping Center Age, 71 (3), 50-57, 1995.

[8] Borin, N., and Farris, P. A sensitivity analysis of retailer shelf management models. In Journal of Retailing, 71, 153-171, 1995.

[9] Brin, S., Motwani, R., and Silverstein, C. Beyond market baskets: generalizing association rules to correlations. In Peckham, J., (ed.). Proceedings of the ACM SIGMOD Conference on Management of Data, 1997 (SIGMOD'97), 265-276.

[10] Brin, S., Motwani, R., Ullman, J., and Tsur, S. Dynamic itemset counting and implication rules for market basket data. In Peckham, J. (ed.). Proceedings ACM SIGMOD International Conference on Management of Data, 1997 (SIGMOD'97), 255-264.

[11] Brown, W., and Tucker W. The marketing center; vanishing shelf space, Atlanta Econ. Rev., 9-13, 1961.

[12] Bultez, A., and Naert, P. S.H.A.R.P.: shelf allocation for retailers' profit. In Marketing Science, 7, 211-231, 1988.

[13] Bultez, A., Gijsbrechts, E., Naert, P., and Vanden Abeele, P. Asymmetric cannibalism in retail assortments. In Journal of Retailing, 65(2), 153-192, 1989.

[14] Corstjens, M., and Doyle, P. A model for optimizing retail space allocations. In Management Science, 27, 822-833, 1981.

[15] Corstjens, M., and Doyle, P. A dynamic model for strategically allocating retail space. In Journal of the Operational Research Society, 34, 943-951, 1983.

[16] Hansen, P., and Heinsbroek, H. Product selection and space allocation in supermarkets. In European Journal of Operational Research, 3, 474-484, 1978.

[17] Hedberg, S. The data gold rush, BYTE, October 1995, 83-88.

[18] Kleinberg, J., Papadimitriou, C., and Raghavan, P. A microeconomic view of data mining. Data Mining and Knowledge Discovery Journal, 2 (4), 311-324, 1998.

[19] Mannila, H. Methods and problems in data mining. In Afrati, F., and Kolaitis, P. (eds.). Proceedings of the International Conference on Database Theory, 41-55, 1997.

[20] McGoldrick, P. Retail Marketing, McGraw-Hill, 362p, 1990.

[21] Park, J., Chen, M., and Yu, Ph. An effective hash based algorithm for mining association rules. In Carey, M., Schneider, D. (eds.). Proceedings of the 1995 ACM SIGMOD International Conference on Management of Data, 175-186.

[22] Pessemier, E. Retail assortments - some theoretical and applied problems, Technical Report, Marketing Science Institute Research Program, 77p., 1980.

[23] Urban, T. An inventory-theoretic approach to product assortment and shelf-space allocation. In Journal of Retailing, 74, 15-35, 1998.

[24] Van der Ster, W., and van Wissen, P. Marketing & detailhandel, Wolters-Noordhoff, 459 p., 1993.

[25] Viveros, M., Nearhos, J., and Rothman, M. Applying data mining techniques to a health insurance information system. In Vijayaraman, T., Buchmann, A., Mohan, C., Sarda, N. (eds.). Proceedings of the 22nd International Conference on Very Large Data Bases, 1996 (VLDB'96), 286-294.

[26] Zaki, M., Parthasarathy, S., Ogihara, M., and Li, M. New algorithms for fast discovery of association rules. In Heckerman, D., Mannila, H., Pregibon, D., and Uthurusamy, R. (eds.). Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, 1997 (KDD'97), 283-286.

[27] Zufryden, F. A dynamic programming approach for product selection and supermarket shelf-space allocation. In Journal of the Operational Research Society, 37, 413-422, 1986.

[28] Zufryden, F. New computational results with a dynamic programming approach for product selection and supermarket shelf-space allocation. In Journal of the Operational Research Society, 38, 201-203, 1987.