CSCanada
Canadian Social Science
Vol. 10, No. 3, 2014, pp. 75-83
DOI:10.3968/4541

ISSN 1712-8056[Print]
ISSN 1923-6697[Online]
www.cscanada.net
www.cscanada.org

Association Rules in Data Mining: An Application on a Clothing and Accessory Specialty Store

Mutlu Yüksel Avcilar[a]; Emre Yakut[a],*

[a]Department of Management Information Systems, Faculty of Economics & Business Administration Sciences, Osmaniye Korkut Ata University, Osmaniye, Turkey.
*Corresponding author.

Received 5 January 2014; accepted 15 April 2014
Pulished online 18 April 2014

Mutlu Yüksel Avcilar, Emre Yakut (2014). Association Rules in Data Mining: An Application on a Clothing and Accessory Specialty Store. *Canadian Social Science, 10*(3), 75-83. Available from: http://www.cscanada.net/index.php/css/article/view/4541
DOI: http://dx.doi.org/10.3968/4541

Abstract

Retailers provide important functions that increase the value of the products and services they sell to consumers. Retailers value creating functions are providing assortment of products and services: breaking bulk, holding inventory, and providing services. For a long time, retail store managers have been interested in learning about within and cross-category purchase behavior of their customers, since valuable insights for designing marketing and/or targeted cross-selling programs can be derived. Especially, parallel to the development of information processing and communication technologies, it has become possible to transfer customers shopping information into databases with the help of barcode technology. Data mining is the technique presenting significant and useful information using of lots of data. Association rule mining is realized by using market basket analysis to discover relationships among items purchased by customers in transaction databases. In this study, association rules were estimated by using market basket analysis and taking support, confidence and lift measures into consideration. In the process of analysis, by using of data belonging to the year of 2012 from a clothing and accessory specialty store operating in the province of Osmaniye, a set of data related to 42.390 sales transactions including 9.000 different product kinds in 35 different product categories (SKU) were used. Analyses were carried out with the help of SPSS Clementine packet program and hence 25.470 rules were determined.

Key words: Specialty retailer store; Data mining; Association rules

1. INTRODUCTION

Retailing is the set of business activities that adds value to the products and services sold to consumers for their personal or family use. Retailers provide important functions that increase the value of the products and services they sell to consumers. These value creating functions are providing assortment of products and services, breaking bulk, holding inventory and providing services and experiences. Retailers are business that manages to satisfy the consumers' needs and wants by offering the right product assortment, at the right quantity, at the reasonable price, at the desired time and place. Thus, they adds value to products and services (Levy & Weitz, 2007, p.7).

Today, retailing industry shows a diversified and partial structure more than ever. Thus, retailers must offer customers too many alternatives. Retailers should add value to the products, offer services to the customers, and provide further product diversity so that they can compete in such diversity (Liao, Chen & Wu, 2008). Because the retailers are the closest marketing unit to consumers within the supply chain, the need for retailers and the importance of retail stores further increase in the process of obtaining customer information, sharing such information in the supply chain and developing strategies which offer high value to consumers. Matters such as how the customer information obtained, processed and used for adding high value strategies to target consumers have begun to be important for the retailers (Liao, Chen, & Wu, 2008).

The improvement in the information technology allows retailers to obtain daily transaction data with very low costs. Thus, large amount of useful data to support retail management can be extracted from large transaction databases. Data mining is used to obtain valuable and useful information from large databases (Chen & Lin, 2007). The way association rule mining, is used in the applications is the market basket analysis. Market basket analysis is one of the important applications of data mining. The products' place in the store can create very important differences in its sales. Therefore, the information regarding which products are sold will show which products should be put side by side in the store shelves. Data mining techniques are used to find the product groups, which are purchased together (Aloysius & Binu, 2013).

When we examine literature, we can see that association rules mining, which is one of the data mining technique is used in researches in various areas in retailing. In their study, Liao and Chen (2004) offered cross sales recommendations for electronic catalogue design, by using association rules based data mining method. In their study, Brijs et al. (2004) examined the cross-sales potential of the products by using frequently repeated product groups in order to increase effectiveness in the selection of the products to promote in the store.

Chen, Chiu and Chang (2005) tried to establish a method, which analyzes the change in the customer behaviors. They used association rules in order to determine the associations between the sold products and purchasing customer specifications. Because of the analysis, they determined what kind of products should be promoted by the store manager in developing more effective marketing strategies and offered recommendations. Chen and Lin (2007) tried to resolve the shelf area allocation and product diversity problems by using association rules mining.

Liao, Chen and Wu (2008) researched product line and brand expansion matters in a retail store. They reached certain information patterns and clusters by using association rules and apriori algorithm. They offered solutions for product line and brand expansion problems in the store and recommended that the store has to perform brand expansion in special brand products for low income customer group. In their study, Sohn and Kim (2008) used the association rules to improve mobile service market by finding consumption behavior patterns of the consumers. In their study, Ay and Çil (2008) offered in store place layout recommendation to the food retailer by using apriori algorithm.

The change in product prices can also change the associations between the products. In their study, Chen, Huang and Chang (2008) added price factor to the association rules and took the necessary data from the database of a retail store chain. At the end of the study, they concluded that the sales of two products which have association between each other and various purchase combinations is based on their prices; and that reducing the prices of both products increased their sales but the price change in one product was more effective. In their study, Nafari and Shahrabi (2010) found the associations between the products group in terms of their prices by using association rules. Based on the results of the study, shelf area was allocated to increase cross-sales profit and total profit by selecting the most profitable products and the most affordable price of these products. In their study, Demiriz et al. (2011) reanalyzed the data in order to explain the reason of positive and negative associations by adding product price, purchase time and customer specifications to the association mining in clothing retailer industry.

1.1 Data Mining

Data mining, which is also referred to as knowledge discovery in databases (KDD), means a process of nontrivial extraction of implicit, previously unknown and potentially useful information (such as knowledge rules) from data in databases (Agrawal et al., 1993; Agrawal & Srikant, 1994; Srikant & Agrawal, 1995; Chen et al., 1996).

The KDD process involves using the database along with any required selection, pre-processing, sub-sampling, and transformations of it; applying data-mining methods (algorithms) to enumerate patterns from it; and evaluating the products of data mining to identify the subset of the enumerated patterns deemed knowledge. The data-mining component of the KDD process is concerned with the algorithmic means by which patterns are extracted and enumerated from data. At the end of this process, the user is offered interesting patterns which are discovered among the data; and these interesting patterns are stored in the database as new knowledge (Fayyad, Piatetsky-Shapiro & Smyth, 1996, pp.40-41).

Data mining is the process of discovering interesting patterns, new rules and knowledge from large amount of sales data in the transactional and relational databases (Han, Kamber, & Pei, 2012, p.8). Data mining, in another definition, is the process of automatically discovering useful information in large data repositories. Data mining techniques are deployed to scour large databases in order to find novel and useful patterns that might otherwise remain unknown (Tan, Steinbach & Kumar, 2006, p.2). In the business perspective, data mining is a business process for exploring large amounts of data to discover meaningful patterns and rules and management uses the discovered knowledge and rules in making strategic decisions. Thus, data mining is a business process that interacts with other business processes. Data mining starts with data, then through analysis informs business action, interesting patterns and rules are presented to the user and stored as new knowledge in the knowledge base (Linoff & Berry, 2011, p.2).

In the literature, some researchers consider data mining as a one step in the knowledge discovery process, which is the overall process of converting raw data into useful information. On the other hand, some other researchers consider the term data mining is often used to refer to the entire of the knowledge discovery process (Linoff & Berry, 2011, p.6).

Recently, the progress of information and communication technology have increasingly made retailers easily collect daily transaction data at very low cost. Through the point of sale (POS) system and barcode technology, a retail store can collect a large volume of customer transaction data. From the huge transaction database, a great quantity of useful information can be extracted to support the retail management business strategies (Chen & Lin, 2007, p.977).

Data mining modeling are generally divided into two major categories as predictive and descriptive tasks. The objective of the first modeling is to predict the value of particular attribute based on the values of other attributes (classification, regression etc.). The objective of the second modeling is to derive patterns that summarize the underlying relationships in data (association rules, clustering analysis and anomaly detection etc.). Descriptive data mining modeling are often exploratory in nature and frequently require post processing techniques to validate and explain the results (Tan, Steinbach, & Kumar, 2006, p.7).

Data mining techniques are intensely used in several fields of science, mainly astronomy, telecommunication, business management, marketing, particularly in the fields of retailing, finance, production and internet commerce (Fayyad, Piatetsky-Shapiro and Smyth, 1996, p.38). The primary goal of data mining is discovery of new patterns and deeper insights within the data. New pattern discovery is used in marketing to make predictions about consumer behavior, to understand consumer preference, and manage customer relationship (Seng & Chen, 2010, s.8042). The main areas of business applications of data mining are shown in Table 1.

Table 1
Data Mining Businesses Application

Application categorization	Industry sector	Business applications
Cross-Sales	Financial	Product cross-sales
	Retailing	Member customer cross-sales
Direct Marketing	Financial	Telemarketing and Direct mail marketing
	Retailing	Ads names list analysis
	Retailing	Direct marketing
	Telecommunication	Direct marketing
	Security	Direct marketing
Segmentation Analysis	Financial	Market segmentation analysis
	Retailing	Market segmentation
	Telecommunication	Customer segmentation
Product Mix Analysis	Financial	Product mix analysis
	Retailing	Promotional product mix
Fraud Detection and Bad Debt Collection	Financial	Credit card fraud detection and bad debt collection
Lost Customer Analysis	Financial	Lost customer analysis
	Telecommunication	Customer retention analysis
Credit Rating	Financial	Credit card applications and personal debt rating
Share Certificate and Foreign Exchange Prediction	Financial	Share certificate and exchange rate prediction
Stock Analysis	Retailing	Stock analysis
Price Optimization	Telecommunication	Price Optimization
Telecommunication Connectivity Analysis	Telecommunication	Connectivity Analysis
Production Efficiency Analysis	Manufacturing	Production efficiency analysis
	Manufacturing	Production process control

Source: Seng and Chen, 2010: 8044.

1.2 Association Rules: Market Basket Analysis

Association rule mining (ARM), was first introduced by Agrawal, Imielinski and Swami in 1993, is one of the most important and well researched techniques of data mining. ARM, frequent item set mining, aims to extract interesting correlations, frequent patterns, associations or casual structures among sets of items in the transaction databases or other data repositories. ARM has attracted the attention and interest of a great number of researchers (Agrawal et al., 1993; Agrawal & Srikant, 1994; Srikant

& Agrawal, 1995; Han & Fu, 1995; Liu et al., 1998; Changchien & Lu, 2001; Tsai & Chen 2004; Chen et al., 2005; Zhou & Yau, 2007; Tang et al., 2008; Romero et al., 2011; Ahn, 2012).

ARM analysis is to use to discover patterns that describe strongly associated features in the data. The discovered patterns are typically represented in the form of implication rules or feature subsets. The goal of ARM analysis is to extract the most interesting patterns in an efficient manner (Tan, Steinbach, & Kumar, 2006, p.9). ARM analysis (also known as market basket analysis) is a method of discovering customer purchasing patterns by extracting associations or co-occurrences from retailer' transactional databases (Tang, Chen, & Hu, 2008, p.150). Market basket analysis is known as association rule mining. This technique analyzes customer buying habits by finding associations between the different items that customers place in their shopping baskets. The discovery of these associations can help retailers develop marketing strategies by gaining insight into which items are frequently purchased together by customers (Han, Kamber, & Pei, 2012, p.244). ARM analysis is also described as finding the strong association rules which meet minimum support and confidence level pre-determined by the user (Srikant & Agrawal, 1995). ARM analysis is a data mining method that examines a large transactional database to determine which items are most frequently purchased jointly. By using ARM analysis, marketing analyst try to explore sets of products which are frequently bought together (Chen, 2007).

1.3 Association Rules Mathematical Model

The fundamental logic of analysis is based on the hypotheses that the customer who purchases a certain product set from a certain store will probably purchase another product set during shopping. The basic task of the association rules analysis is to derive a set of strong association rules in the form of ("$A_1 \wedge A_2 \wedge ....A_m \Rightarrow B_1 \wedge B_2 \wedge ....B_n$") where A_i and B_j are sets of attribute values, from relevant data sets in a database (Chen, Han and Yu, 1996, p.868).

An association rule is an implication expression of the form $A \Rightarrow B$, where A and B are disjoint item sets, i.e., $A \cap B = \varnothing$. The strength of an association rule can be measured in terms of its support and confidence values. Support determines how often a rule is applicable to a given data set, while confidence determines how frequently items in B appear in transactions that contain A (Tan, Steinbach, & Kumar, 2006, pp.329-330).

The formal statement of the association rule can be stated as, I= $\{I_1, I_2, ..., I_m\}$ is set of binary attributes called items. D is a market basket database, where contains transactional records, each transaction T is a nonempty itemset such that $T \subseteq I$. Each transaction t is represented as a binary vector, with t [k] =1 if t bought the item I_k, and t [k] =0 otherwise. There is one tuple in the database for each transaction and each transaction associated with an identifier, called a TID (Agrawal, Imeelinski, & Swami, 1993, p.208). Let A and B are a set of items. A and B, transaction T is said to contain A and B if and only if $A \subseteq T$ and $B \subseteq T$. An association rule is an implication of the form $A \Rightarrow B$, where $A \subset I$, $B \subset I$, $A \neq \varnothing$, $B \neq \varnothing$, and $A \cap B = \varnothing$. The rule $A \Rightarrow B$ holds in the transaction set D with support s, where s is the percentage of transactions in D that contain $A \cap B$. The rule $A \Rightarrow B$ has confidence c in the transaction set D, where c is the percentage of transactions in D containing A that also contain B. Support indicates the frequencies of the occurring patterns in the rule and confidence denotes the strength of implication. Association rules that satisfy both a minimum support threshold and a minimum confidence threshold are called strong rules (Han, Kamber, & Pei, 2012, p.246).

The task of association rules mining is essentially to discover strong association rules in large database. In general, the problem of mining association rules is decomposed into the following two steps (Agrawal, Imeelinski, & Swami, 1993, p.208):

a) Discover the large itemsets, i.e, the sets of itemsets that have transaction support above a predetermined minimum support s. Itemsets with minimum support are called frequent itemsets.
b) Use the large itemsets to generate the association rules for the database.

Agrawal, Imeelinski and Swami (1993) stated that the overall performance of mining association rules is determined by the first step. After the large itemsets are identified, the corresponding association rules can be derived in a straightforward manner.

1.4 Apriori Algorithm

Apriori algorithm is proposed by R. Agrawal and R. Srikant in 1994 for mining frequent itemsets for Boolean association rules. Apriori algorithm is one of the most widely used techniques for finding association rules (Agrawal & Srikant, 1994). The name of the algorithm is based on the fact that the algorithm uses prior knowledge of frequent itemset properties. Apriori employs an iterative approach known as a level-wise search, where k-itemsets are used to explore (k+1) itemsets. First, the set of frequent 1-itemsets is found by scanning the database to accumulate the count for each item, and collecting those items that satisfy minimum support. The resulting set is denoted by L_1. Next, L_1 used to find L_2, the set of frequent 2-itemsets, which is used to find L_3, and so on, until no more frequent k-itemsets can be found. To improve the efficiency of level-wise generation of frequent itemsets, apriori property (all nonempty subsets of a frequent itemset must also be frequent) is used to reduce the search space (Han, Kamber, 7 Pei, 2012, p.248). The basic notations which are used in apriori algorithm are shown in Table 2.

Table 2
Notations Used in Apriori Algorithm

k-itemset	An itemset having k itemset
L_k	Set of large k-itemsets (those with minimum support).
C_k	Set of candidate k-itemsets (potentially large itemsets).

Source: Agrawal and Srikant, 1994.

The apriori property is based on the following observation. By definition, if an itemset *I* does not satisfy the minimum support threshold, min_sup, then *I* is not frequent. If an item A is added to the itemset *I,* then the resulting itemset cannot occur more frequently than *I.* Therefore, I∪A is not frequent either, that is , P(I∪A) min_sup. Based on this property, if a set cannot pass the minimum support threshold, all of its supersets will fail the same test as well. Thus, if a itemset is not a frequent itemset, this itemset will not used to create large itemset (Han, Kamber, & Pei, 2012, p.248-249).

In order to understand how is the apriori property used in the algorithm, it should be considered how L_{k-1} is used to find L_k for k≥2. A two-step process is followed, which consisting of join and prune steps in the apriori algorithm (Han, Kamber, & Pei, 2012, p.249-250). These steps are as shown below:

a. The join step: To find L_k, a set of candidate k-itemsets is generated by joining L_{k-1} with itself. This set of candidates is denoted C_k.

b. The prune step: C_k is a superset of L_k , that is, its members may or may not be frequent, but all of frequent k-itemset are included in C_k. A database scan to determine the count of each candidate in C_k would result in the determination of L_k. C_k, however, can be huge, and so this could involve heavy computation. To reduce the size of C_k, the apriori property is used. Any (k-1) itemset that is not frequent cannot be a subset of a frequent k-itemset. Hence, If any (k-1) subset of a candidate k-itemset is not in L_{k-1}, then the candidate cannot be frequent either and so can be removed from C_k.

While applying apriori algorithm, the standard measures are used to asses association rules. These rules are the support and confidence value. Both are computed from the support of certain itemsets. For association rules like A⇒B, two criteria are jointly used for rule evaluation. The support s, is the percentage of transactions that contain A∪B (Agrawal, Imeelinski, (Swami, 1993). It takes the form support (A⇒B)=P(A∪B), where support is the percentage of transactions contain A∪B (i.e., the union of sets A and B or both A and B). This is taken to be the probability, P(A∪B). The confidence is the ratio of percentage of transactions that contain A∪B to the percentage of transactions that contain A. It takes the form confidence (A⇒B)=P(B\A)= support (A∪B)/support (A). Rules that satisfy both a minimum support threshold (min_sup) and minimum confidence threshold (min_conf) are called strong. Given a set of transactions, the problem of mining association rules is to generate all the rules that have support and confidence greater than the user specified minimum support and minimum confidence.

2. RESEARCH METHODOLOGY

Association rules analysis was used in this empirical research. Association rules were determined by considering support level, confidence level and lift values. In the process of analysis, by using the data of the sales transactions made between 01.01.2012 and 31.12.2012 from a specialty store which operates in Osmaniye province in Turkey. During the analysis process, dataset including 9.000 different product ranges in 35 different product categories (SK) and 42.390 sales transactions were used. For the analyses, firstly, the sales data was prepared for analysis by giving 0 and 1 codes in Excel. And then, the analyses were conducted by using Apriori algorithm on SPSS Clementine 12.0 software.

2.1 Modeling

In this study, Apriori algorithm, which is the most frequently used algorithm among the association rules algorithms, was used at the analysis phase. During the analysis, minimum support level was determined as 0,05%, minimum confidence level as 50% and antecedent number as 6. The antecedent number 6 will enable to obtain more useful and stronger rules. Figure 1 includes the structure of the model developed for the research in SPSS Clementine 12.0 software.

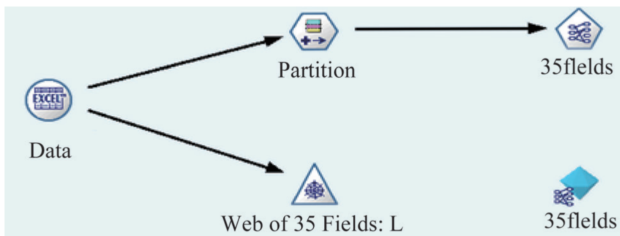

Figure 1
Structure of the Model

The descriptive statistic of the dataset used in the analysis at product category level is given in Table 3. When we examined Table 3, it was observed that the product category in which the most frequent sales transactions are made among total 42.390 transactions at the product category level was the shirt category with GMC product code with 13.148 times.

Table 3
Descriptive Statistics

	Product category level
Total	42.390
Average	1.211
Median	236
Mod	GMC=13.148
Maximum	13.148
Minimum	2

Frequency and percentage values regarding the most frequently sold product groups at product category level are given in Table 4. When we examined Table 4, it was observed that 54% of the total sales within one year were made in 31% shirts (GMC), 14% socks (CRC) and 9% polyester trousers (PNC) product groups respectively.

Table 4
Sales Frequency of Products

Product code	**Sales** (frequency)	**Percentage** (%)
GMC	13.148	31.02%
CRC	5.812	13.71%
PNC	3.841	9.06%
ÇRC	3.666	8.65%
KRG	2.554	6.03%
TKC	1.975	4.66%
PNA	1.970	4.65%
AYC	1.429	3.37%
KMC	1.183	2.79%
TSC	1.153	2.72%
CMC	1.130	2.67%

Within association rules analysis process, with Apriori algorithm, minimum support level 0,05%, minimum confidence level 50% and antecedent number 6, total 25.470 rules were created. Among the rules obtained as a result of operating the models, total 8 major rules with the highest confidence level and lift value is higher than 1 are given in Table 5 .

Table 5
Association Rules

Consequence	**Antecedent**	**Support %**	**Confidence %**	**Lift value**
GMC	KMC&TKC&KRG	4.24	100	1.63
KRG	KMC&TKC&GMC	4.38	97	3.93
TKC	AYC&KRG&GMC	4.85	93	5.02
AYC	KMC& TKC&KRG&GMC	4.24	80	5.65
KMC	AYC& TKC&KRG&GMC	4.51	75	6.40
PNC	CTC&GMC	2.95	68	3.13
CRC	CMC	3.70	55	2.88
ÇRC	ÇMC	3.22	51	3.09

Support and confidence measurements are known to remain insufficient in the elimination of unimportant association rules. To eliminate this problem, it is recommended to check the lift value of the association rules and have lift value equal to or higher than 1 (Han, Kamber and Pei, 2012, p.265-266). Lift value, is a simple correlation measure of the rules. The occurrence of itemset A is independent of the occurrence of itemset B if P(AB)= P(A)x P(B); otherwise, itemsets A and B are dependent and correlated. The lift value between the occurrence of A and B can be measured by computing (Brin et al., 1997, p.260):

Lift (A, B)= P(AB) / P(A)x P(B)

If the lift value is less than 1, then the occurrence of A is negatively correlated with the occurrence of B. If the lift value is greater than 1, then A and B positively correlated, meaning that the occurrence of one implies the occurrence of the other. If the lift value is equal to 1, then A and B are independent and there is no correlation between them (Han, Kamber and Pei, 2012, p.266). As it can be understood from the abovementioned information, Lift value was taken to determine the most important association rules are given in Table 5.

The most important rules determined as a result of analyses are as seen in Table 5:

❖ **Rule 1:** The most important rule; a customer who purchases a belt (KMC), suit (TKC) and a tie (KRG) purchases a shirt (GMC) with 100% probability. This group has 4.24% probability to be found together within the operations in the dataset.

❖ **Rule 2:** A customer who purchases a belt (KMC), suit (TKC) and shirt (GMC) will purchase a tie (KRG) with 97% probability. This group has 4.38% probability to be found together within the operations in the dataset.

❖ **Rule 3:** A customer who purchases Neolith shoes (AYC), tie (KRG) and shirt (GMC) will purchase suit (TKC) with 93% probability. This product group has 4.85% probability to be found together within the operations in the dataset.

❖ **Rule 4:** A customer who purchases a belt (KMC), suit (TKC), tie (KRG) and shirt (GMC) will purchase Neolith shoes with 80% probability. This product group has 4.24% probability to be found together within the operations in the dataset.

To see the associations among the products, we exploited the web graphic which shows the frequency of being found together among the product categories in SPSS Clementine 12.0 software. As it can be seen in the web graphic, we can say that as the line thickness increases, the association level among the product categories increase. In the web graphic in Figure 2, the line value is set to be between 25% and 50%.

When Figure 2 is examined, it is seen that the line thicknesses intensify in tie (KRG), shirt (GMS) and shirt (GMC) categories from the inter product groups associations. The association level among the product categories are given in Figure 3.

As it can be seen in Figure 3, the highest association is between the groom's suit (TKS) and shirt (GMS) and the association between these two product groups is 81.76%. The second highest association is between the suit (TKC) and tie (KRG) and the association between these two product groups is 69.34%. The third highest association is between the belt (KMC) and neolith shoes (AYC) and the association between these two product groups is 56.23%.

The fourth highest association is between the vest foulard (YFS) and groom's suit (TKS) and the association between these two product groups is 51.16%. The fifth highest association is between the cuff-link (KDC) and shirt (GMS) and the association between these two product groups is 50.83%.

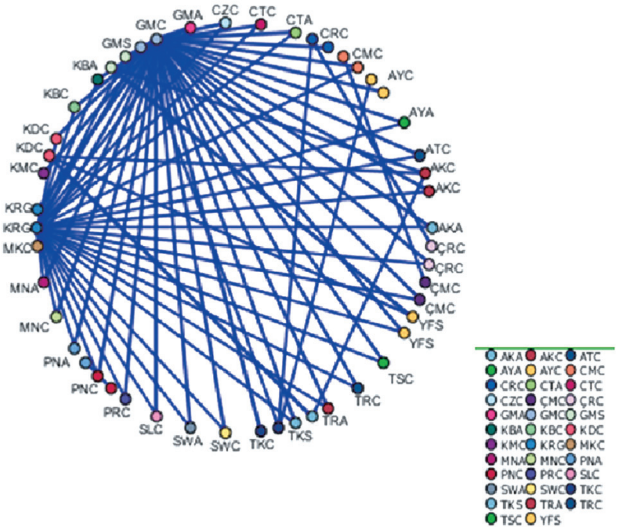

Figure 2
Web Diagram Showing Inter-Products Association Level

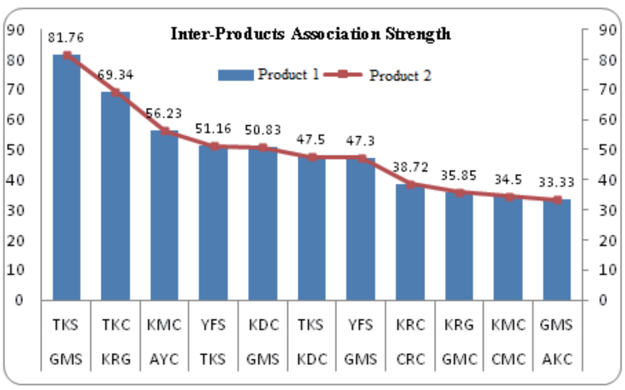

Figure 3
Inter-Products Association Strength

CONCLUSION

In this study, specialty store customers' shopping information were analyzed by using association rules mining with Apriori algorithm. As a result of the analyses, strong and useful association rules were determined between the product groups with regard to understanding what kind of purchase behavior customers exhibit within a certain shopping visit from both in-category and different product categories for the specialty store in question. By utilizing the association rules which are discovered as a result of the analyses, the retail store manager will be able to develop and apply effective marketing and sales promotion strategies.

Today, the success of retail stores is based on the high value strategies to their target customers. From the information obtained as a result of the association rules analyses, the store manager will be able to support quite important strategic decisions of the retailing businesses which effect their success such as market segmentation, cross-sales, determination of the product mix to be offered for sale in the store, effective product assortment management, determination of the product prices and management of product price discounts, planning and management of product promotions, stock management, visual presentation of the products in the store, allocation the products on the shelves in the store.

Consequently, strong associations have been observed among the purchased product groups with regard to the purchase behavior of the customers of the retail store within the scope of this study. In future research, time based consecutive, patterned association rules can be determined through the prospective researches. Additionally, more useful areas which take the customers' specifications into account can be created by creating decision trees from data mining techniques upon adding the demographic and psychographic variables of the customers to the research model.

REFERENCES

Agrawal, R., Imielinski, T., & Swami, A. (1993, May). Mining association rules between sets of items in large databases. ACM SIGMOD Conference, Washington DC, USA.

Agrawal, R., & Srikant, R. (1994). Fast algorithms for mining association rules. VLDB (Very Large Data Base Endowment), 20th VLDB Conference, Santiago, Chile.

Ahn, K. (2012). Effective product assignment based on association rule mining in retail. *Expert Systems with Applications*, *39*, 551-556.

Aloysius, G., & Binu, D. (2013). An approach to products placement in supermarkets using prefixspan algorithm. *Journal of King Saud University-Computer and Information Sciences, 25*, 77-87.

Ay, D., & Çil, İ. (2008). Using association rules for Migros Turk in the development of layout plan. *Journal of Industrial Engineering, 21*(2), s.14-29.

Brijs, T., Swinnen, G., Vanhoof, K., & Wets, G. (2004). Building an association rules framework to improve product assortment decisions. *Data Mining and Knowledge Discovery, 8*, 7-23.

Brin, S., Motwani, R., Ullman, J. D., & Tsur, S. (1997). Dynamic item set counting and implication rules for market basket data. SİGMOD, Acm-sigmod international conference on management of data, New York, NY, USA.

Changchien, S. W., & Lu, T. (2001). Mining association rules procedure to support on-line recommendation by customers and products fragmentation. *Expert Systems With Applications*, *20*, 325-335.

Chen, M., Han, J., & Yu, P. (1996). Data mining: An overview from a database perspective. *IEEE Transactions on Knowledge and Data Engineering, 8*(6), 866-883.

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

Chen, M., & Lin, C. (2007). A data mining approach to product assortment and shelf space allocation. *Expert Systems With Applications, 32*, 976-986.

Chen, Y., Huang, T. C., & Chang, S. (2008). A novel approach for discovering retail knowledge with price information from transaction databases. *Expert Systems With Applications, 34*, 2350-2359.

Demiriz, A., Ertek, G., Atan, T., & Kula, U. (2011). Re-mining item associations: Methodology and a case study in apparel retailing. *Decision Support Systems, 52*, 284-293.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine, 17*(3), 37-54.

Han, J., & Fu, Y. (1995). Discovery of multiple-level association rules from large databases. 21st VLDB (Very Large Data Base Endowment) Conference, Zurich, Switzerland.

Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques* (3rd ed.). San Francisco: Morgan Kaufmann Inc.

Kamakura, W. A. (2012). Sequential market basket analysis. *Mark Lett, 23*, 505-516.

Levy, M., & Weitz, B. A. (2004). *Retailing management* (5th ed.). Boston: Irwin McGraw-Hill Inc.

Linoff, G. S., & Berry, M. J. (2011). *Data mining techniques: For marketing, sales and customer relationship management* (3rd ed.). Indianapolis: Wiley Publishing Inc.

Liao, S., & Chen, Y. (2004). Mining customer knowledge for electronic catalog marketing. *Expert Systems with Applications, 27*, 521-532.

Liao, S., Chen, C., & Wu, C. (2008). Mining customer knowledge for product line and brand extension in retailing. *Expert Systems with Applications, 34*, 63-76.

Liu, B., Hsu, W., & Ma, Y. (1998). Integrating classification and association rule mining. *Knowledge Discovery and Data Mining*, 80-86.

Nafari, M., & Shahrabi, J. (2010). A temporal data mining approach for shelf-space allocation with consideration of product price. *Expert Systems with Applications, 37*, 66-72.

Romero, C., Luna, J. M., Romero, J. R., & Ventura, S. (2011). RM-Tool: A framework for discovering and evaluating association rules. *Advances in Engineering Software, 42*, .566-576.

Seng, J., & Chen, T. C. (2010). An analytic approach to select data mining for business decision. *Expert Systems with Applications, 37*, 42-57.

Sohn, S. Y., & Kim, Y. (2008). Searching customer patterns of mobile service using clustering and quantitative association rule. *Expert Systems With Applications, 34*, 70-77.

Srikant, R., & Agrawal, R. (19950. Mining generalized association rules. 21st VLDB (Very Large Data Base Endowment) Conference, Zurich, Switzerland.

Tan, P., Steinbach, M., & Kumar, V. (2006). *Introduction to data mining*. Boston: Pearson Education.

Tang, K., Chen, Y., & Hu, H. (2008). Context-based market basket analysis in a multiple-store environment. *Decision Support Systems, 45*, 150-163.

Tsai, P. S. M., & Chen, C. (2004). Mining interesting association rules from customer databases and transaction databases. *Information Systems, 29*, 685-696.

Zhou, L., & Yau, S. (2007). Efficient association rule mining among both frequent and infrequent items. *Computers and Mathematics With Applications, 54*, 737-749.

APPENDIX

Table 6
Product Codes Corresponding Product Names

S.N	Product codes	Category	Product name
1	AKA	Shawl	PATTERNED SHAWL
2	AKC	Accessory	CUFF-LINK
3	ATC	Scarf	CACHET SCARF
4	AYA	Shoes - Casual	SVR1010 CASUAL SHOES
5	AYC	Shoes - Neolith	SVR1152 NEOLITH SHOES
6	CMC	Underwear	PATTERNED BOXER
7	CRC	Socks	CHECKERED SOCKS
8	CTA	Jacket	POLYESTER VISCOSE 6 DR PATTERNED JACKET
9	CTC	Jacket	CT12062 6 DROP BLAZER JACKET
10	CZC	Wallet	SVR12345 LEATHER WALLET
11	ÇMC	Underwear	SINGLE LONG
12	ÇRC	Socks – Classic	ÇR2011 CLASSIC SOCKS
13	GMA	Shirt	SLIMFIT DENIM MODEL SHIRT
14	GMC	Shirt	G52067 WONDER EXTRA TIGHT LONG SHIRT
15	GMS	Shirt	G41720 WONDER NECKBAND TIGHT
16	KBA	Topper	WOOL POLYESTER MODEL TOPPER
17	KBC	Topper	WOOL POLYESTER JAM TOPPER
18	KDC	Cuff-link	CUFF-LINK
19	KMC	Belt	KM300 SVR BELT
20	KRG	Tie	WEAVE 3 TIE Article. WEA3
21	MKC	Leather Jacket	MK 3163 MODEL TOP COAT
22	MNA	Coat	EPAULETTE POCKET FLAP COAT
23	MNC	Coat	TRENCH COAT ROUND COLLAR TOP COAT
24	PNA	Trousers – Cotton	P25165 MODEL COTTON TROUSERS
25	PNC	Trousers - Polyester	P12124 UNPLEATED VELVET TROUSERS 2012
26	PRC	Perfume	S.97 PERFUME
27	SLC	Shawl	PATTERNED SHAWL
28	SWA	Sweatshirt	COTTON POLYESTER MOCK-TURTLENECK ZIPPERED SWEATSHIRT
29	SWC	Sweatshirt	SW1014 V COLLAR
30	TKC	Suit	TK12623 6 DROP FIL A FIL SUIT
31	TKS	Suit – Groom’s Suit	TK41728 6 DROP SSN GROOM’S SUIT
32	TRA	Pullover – Turtleneck	SVR1001 O COLLAT TRICOT
33	TRC	Tricot	TR4114 MOCK TURTLENECK SOLID TRICOT TWIN NEEDLE
34	TSC	Mercerized – T-shirt	100% COTTON O COLLAR T-SHIRT
35	YFS	Vest foulard	VEST FOULARD SET