Market Basket Analysis of Consumer Buying Behaviour of a Lifestyle Store

Grishma Kapadia  
Kavita Kalyandurgmath  
Prin. L. N. Welingkar Institute of Management Development and Res  
(Grishma.kapadia22@gmail.com)  
(Kavita.Kalyandurgmath@welingkar.org)

Market Basket Analysis is the discovery of relations among various items. The proposed method can analyze and predict the sales of individual product category on a daily basis and predict most likely items sold. The study area is Mumbai. Average Data per store is one year. Primary data was provided by the store itself from the retail software. It predicts using Microsoft Associative algorithm with help of Business Intelligence Development studio software The type of Research is exploratory. Results can provide a valuable reference for cross-sell, up-sell, devising promotions and placing the merchandise in the store for improved sales.

1. Introduction

1.1 Concept of Market Basket Analysis
Market basket analysis is the study of items that are purchased independently or together in a single or multiple, transactions. This study of behaviour of customers helps to understand the relationships of the products. Thus the strength of those relationships is valuable information and can be used to cross-sell, up-sell, offer coupons, and make other recommendations. Market Basket Analysis is a modelling technique based on the theory that if a customer buys a certain products or contents, there are probable chances that he or she might buy other set of products or contents. For example if a customer buys tea and bag it is most likely that he buys Memento. The set of items a customer buys is referred as an itemset and market basket analysis becomes handy to find relationships between itemsets. Market basket analysis helps a retailer with clues as to what a customer might think of buying to constitute his basket.

1.2 Literature Review
Data mining has taken an important part of marketing literature for the last several decades. Market basket analysis is one of the oldest areas in the field of data mining and is the best example for mining association rules. Various algorithms for Association Rule Mining (ARM) and Clustering have been developed by researchers to help users achieve their objectives. Rakesh Agrawal and Usama Fayyad are one of the pioneers in data mining. They account for a number of developed algorithms and procedures .According to Shapiro, rule generating procedures can be divided into procedures that find quantitative rules and procedures that find qualitative rules.

As proposed by Gregory Piatetsky-Shapiro, William Frawley, 1991 tested the effectiveness of the algorithm by applying it to data obtained from a large retailing company. They used attributes for mining qualitative rules for categorical data using association rules. Association rules found application in many research areas such as: market basket analysis, recommendation systems to Advertising and Strategy formulation. In marketing literature market basket analysis has been classified into two models: explanatory and exploratory. The main idea behind exploratory models is the discovering of purchase patterns from POS (point-of-sale) data. Exploratory approaches do not include information on consumer demographics or marketing mix variables summarise a vast amount of data into a fewer meaningful rules or measures. Such methods are quite useful for discovering unknown relationships between the items in the data. Moreover, these methods are computationally simple and can be used for undirected data mining. However, exploratory approaches are not appropriate for forecasting and finding the cause-roots of complex problems. They are just used to uncover distinguished cross-category interdependencies based on some frequency patterns for items or product categories purchased together. A typical application of these exploratory approaches is identifying product category relationships by simple association measures. Pairwise associations are used to compare entities in pairs and judge which entity is preferred or has greater amount of some quantitative property.

Julander, 1992 compared the percentage of shoppers buying a certain product and the percentage of all total sales generated by this product. By making such comparisons, one can easily find out the leading products and what is their share of sales. Examining which the leading products are for consumers is extremely important since a large number of shoppers come into contact with these specific product types every day. As the departments with leading products generate much in-store traffic, it is crucial to use this information for placing other specific products nearby. Another significant stream of research in the field of exploratory analysis is the process of generating association rules.

According to M.J.Zaki, M.Ogihara, S. Parthasarathy, 1996 there are few algorithms developed that are not based on the Apriori, i. Robert J. Hilderman, Colin L. Carter, Howard J. Hamilton, and Nick Cercone developed a framework for knowledge discovery from market basket data. Combining Apriori and Attribute Oriented Generalization (AOG) D.W. Cheung, A.W. Fu, and J. Han., 1994) developed the methodology, using the algorithm not only to explain how to discover customer purchase patterns, but to find out customer profiles by dividing customers into distinct classes. According to Qiankun Zhao, Sourav S. Bhowmick, 2003, there are two basic parameters of Association Rule Mining (ARM): support and confidence. They both measure the strength of an association rule .Many Data Scientists have developed their own
methodology using the associative algorithm for example to segregate the customers into heavy users, moderate users and lighter users. Heavy user households are found to be less price sensitive, visiting the store less often, in most cases high income customers. While, on the other hand, lighter users are mainly students or people that visit the store very often and are very price sensitive. The results all showed that households that have identical behaviour across product categories tend to be lighter users than households that behave independently. Also households with identical behaviours are said to be more price sensitive, less sensitive to store advertising, also showing weaker loyalty in terms of brand names.

Gary J. Russel, Wagner A. Kamakura, 1997 using long-run market basket data, showed how brand preference segmentation can be discovered without the availability of marketing mix data. A number of simplifying assumptions need to be made in order to permit these cross-category preferences to be estimated. Exploratory models are very useful for uncovering cross-category relations, but not for finding their causes. While the main task of exploratory market basket analysis is to reveal and present hidden relationships between product categories, explanatory models aim at explaining effects. Datasets for such models consist of market basket data, customer attributes and marketing mix variables.

Andreas Mild, Thomas Reutterer, 2003 identified the purpose of explanatory models is to identify and quantify cross-category choice effects of marketing variables, such as price, promotion and other marketing features. Andreas Mild, Thomas Reutterer, 2003 mentioned that the exploratory models rely greatly on regression analysis, logit probit and multivariate logistic model mainly on quantitative data. Mining transactional data along with household data gives retailers and managers space for customised target marketing actions. Analysing past purchases makes it possible for supermarkets to price goods intelligently while still serving heterogeneous consumers. Using the shopping basket as a unit of analysis instead of single articles can provide retailers with consumer-oriented information.

Rakesh Agrawal, Ramakrishnan Srikant elaborated on the concept of mining quantitative rules in large relational tables. Quantitative rules are defined in terms of the type of attributes contained in these relational tables. Attributes can be either quantitative like customer’s age, customers’ income, etc. or categorical type of a product, make of a car. Boolean attributes are such attributes that can take on one of two options (True or False, 1 or 0). They are considered a special case of categorical attributes. The authors call this mining problem the Quantitative Association Rules problem.

S. Prakash, R.M.S. Parvathi, 2011 proposed a qualitative approach for mining quantitative association rules. The nature of the proposed approach is qualitative because the method converts numerical attributes to binary attributes. However, finding qualitative rules is of main interest in their analysis. These rules are most commonly represented as decision trees, patterns or dependency tables.

I.3 Objective of the Study
To stay competitive, retailers must understand their current consumer behaviour and be able to predict future consumer behaviour. Understanding customer behaviour can help retailers to retain customers, improve sales, and extend their relationship with customers.

I. To understand the buying pattern of the products that comprises the customers’ basket
II. To study the most likely products purchased by the customers.
III. To study the most likely products purchased by the customers along with a particular product category
IV. To predict and suggest products to individual consumers.

I.4 Research Methodology
The input is transactional data captured from the customers billing served as a primary source. Hence the data is reliable and has an authentic source of consumer buying behaviour. The data was extracted by the retailer software s like , Tally shopper Version 9 from which was later connected to the database of the retail store. The researcher used categorical data of transaction records and with the help of Structured Query language (SQL) the database was connected to Microsoft softwares such as Business Intelligence Development Studio (BIDS) software in server 2008 R2. Business Intelligence Development Studio is the primary environment which is used to develop business solutions that includes Analysis Services, Integration Services and also for creating the objects required for business intelligence solutions. The average data is of one year duration for four outlets of a lifestyle retail store at Mumbai, India. The data is analysed using advanced data mining technology i.e. Association Rules. This transactional data is used to build market basket scenario in order to predict the most likely products that a customer may add to their basket.

2. Building a Market Basket Scenario
The ability to predict products that a customer might want to purchase based on the other products that are already in the customer's shopping basket. Data mining is especially useful for this kind of market basket analysis and data mining model is developed. A mining model that shows groups of items from historical customer transactions is used. Also, mining model can be used to predict additional items that a customer may want to purchase.

A mining model that is required to be built to do market basket analysis for which first of all, data source is created and then by adding a data source view that contains tables about the customer is created in Business Intelligence development software on server 2008 R2. A mining model uses the Microsoft Association Rules algorithm is used for market basket scenarios. The following steps are followed for Market Basket Analysis
1. Adding a Data Source View with Nested Tables
2. Creating a Market Basket Structure and Model  
3. Modifying and Processing the Market Basket Model  
4. Exploring the Market Basket Models  
5. Predicting Associations

2.1 Adding a Data Source View with Nested Tables
In order to build the mining model for market basket analysis a data source view is created. A nested table is a table that contains multiple rows of information about a single row in the case table. This data source contains a nested table. For example, if our model analyses the purchasing behaviour of customers, we would typically use a table that has a unique row for each customer as the case table. However, each customer might make multiple purchases, and might want to analyse the sequence of purchases, or products that are frequently purchased together. To logically represent these purchases in our model, we add another table to the data source view that lists the purchases for each customer. This nested purchases table is related to the customer table by a many-to-one relationship. The nested table might contain many rows for each customer, each row containing a single product that was purchased, perhaps with additional information about the order that the purchases were made, the price at the time of the order, or any promotions that applied.

Screenshot 1: Many-to-one relationship between the two tables.

So in this case, the case table that contains information about the Document number of the products purchased by the customer is used and nested table that contains information about Category, subcategory, Description of the products purchased by each customer is used. Many-to-one relationship with the help of Transaction control number attribute of the customer's bill is used to relate the tables as shown in screenshot 1.

2.2 Creating a Market Basket Structure and Model
A data source view is created which is used to create a new mining structure. A mining structure and a mining model that is based on the Microsoft Association algorithm are created. It is a technique that assists in understanding what items are likely to be purchased together as per the association rules. A predictive market basket analysis can be used to identify sets of item purchases that generally occur in sequence — which are of interest to direct marketers. Associative prediction can serve many purposes, like recommending items to a customer, or finding relationships amongst products. A group of items in a case is called an item set. An association model consists of a series of item sets and the rules that describe how those items are grouped together within the cases. The rules that the algorithm identifies can be used to predict a customer's likely future purchases, based on the items that already exist in the customer's shopping cart.

Requirements
An association model must contain is key columns, and a single predictable column. Market Basket Analysis Variables:
- Independent Variable considered for this study:
- Doc No: Document number of the products purchased by the customers (Key column)
- Dependent Variable considered for this study:
- Subcategory: Subcategory of the products of a lifestyle store. (Predictable column)
For Example: Bodycare, Haircare, etc. as depicted in screenshot 1.

2.3 Modifying and Processing the Market Basket Model
Before the association mining model is processed, the default values of two of the parameters: Support and Probability is changed as required.
- Support defines the percentage of cases in which a rule must exist before it is considered valid. In this case it is specified that a rule must be found in at least 1 percent of cases.
Probability defines how likely an association must be before it is considered valid. In this case any association with a probability of at least 10 percent is to be considered.

- \[
\text{MINIMUM\_PROBABILITY} = 0.1
\]
- \[
\text{MINIMUM\_SUPPORT} = 0.01
\]

The values for support and probability are adjusted. Thus the structure and parameters for the Association mining model are defined and then the model is processed.

2.4 Exploring the Market Basket Models

Typically the relationship will be in the form of a rule: For example, the probability that a customer will buy a wall decor without a momento is referred to as the support for the rule. The conditional probability that a customer will purchase a body care product is referred to as the confidence.

After the Association model is built, it can be explored by using the rulers tab of the viewer which helps us to see at a glance which products tend to appear together, and get a general idea of the emerging patterns.

The association model can be explored by using Microsoft Association Viewer in the Mining Model Viewer Rule tab. This helps to explore relationships between items, to observe which products tend to appear together, and get a general idea of the emerging patterns.

The Rules tab displays the following information that is related to the rules that the algorithm finds.

- **Rule:** The definition of the rule. For a market basket model, a rule describes a specific combination of items
- **Probability:** The likelihood of a rule, defined as the probability of the right-hand item given the left-hand side item.
- **Importance:** A measure of the usefulness of a rule. A greater value means a better rule.

Each rule can be used to predict the presence of an item in a transaction based on the presence of other items. Importance is provided to help gauge the usefulness of a rule, because probability alone can be misleading. For example, if every transaction contains a tea--perhaps the tea is added to each customer's cart automatically as part of a promotion--the model would create a rule predicting that tea has a probability of 1. Based on probability alone, this rule is very accurate, but it does not provide useful information.

2.5 Predicting Associations

After the models have been processed, the information about associations stored in the model is used to create predictions. Prediction queries are built against the association models. Associative prediction is used for many purposes like recommending items to a customer, or finding relationships amongst products.

To build a prediction query, the association model is selected and then the input data is specified in terms of the table i.e. the parameter or variable to be used as input and the parameter or variable for which the particular output needs to be generated. Hence, some singleton prediction queries are created. Then a query for batch predictions is created that can be used for making recommendations based on a customer's current purchases.

3. Findings

3.1 To Understand the buying Pattern of the Products that Comprises the Customers’ Basket

Screenshot 2 shows the Rules tab that gives an idea of emerging patterns, which specific items are related to each other more and association between them along with the importance.

**Screenshot 2:** Products associated with other products that comprises the basket

This is used to predict products that a customer might want to purchase based on the other products that are already in the customer's shopping basket.

Hence using Market Basket Analysis on server 2008 R2 with the help of SQL Server Business Intelligence Development Studio, a sequence is obtained for different category of the products.

As per the analysis, following pattern has emerged that can be used for various purposes like strategic placement of the products on the shelf, to improve sales, for convenience of the customers, devising promotions, to suggest products to the
customers by salesperson, etc.
Interpretation: Tableware, Desktop -> Memento

- A customer buying Tableware and Desktop is likely to buy Memento

Similarly, Other Patterns emerged are as below

<table>
<thead>
<tr>
<th>Pattern</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tableware, Desk Top -&gt; Memento</td>
</tr>
<tr>
<td>Desk Top, Memento -&gt; Tableware</td>
</tr>
<tr>
<td>Furniture, Memento -&gt; Hanging</td>
</tr>
<tr>
<td>Hanging, Memento -&gt; Furniture</td>
</tr>
<tr>
<td>Furniture, Wall Decor -&gt; Memento</td>
</tr>
<tr>
<td>Wall Decor, Memento -&gt; Furniture</td>
</tr>
<tr>
<td>Tea, Memento -&gt; Bag</td>
</tr>
<tr>
<td>Tableware, Furniture -&gt; Memento</td>
</tr>
</tbody>
</table>

3.2 To Study the Top Most Likely Products Purchased by the Customers.
Singleton prediction query is created for the five most likely products purchased by the customers. It can be extended to as many numbers as required. Thus in mining model prediction tab, the query is created to obtain the results.

Screenshot 3: The singleton prediction query for the top most likely products purchased by the customers

The expression column gives the five most likely (subcategory) of the products that are purchased by the customers as shown in screenshot 3. This gives an idea to the retailer of the products that are frequently sold and grabs the attention of the customers.

Screenshot 4: The most likely products purchased by the customers as per the model

3.3 To Study the Most Likely Products Purchased by the Customers with a Particular Product Category
A singleton prediction query with nested table inputs is created. Query for predicting products that are most likely associated with the Body Care is created as shown in screenshot 5. The results of the predictions for products that are most likely associated with the Body Care are obtained as shown in screenshot 6. This gives an idea to the retailer of the likely products purchased by the customers along with the Body care category. This can help in devising promotions, discounts, to tempt the customers to purchase other associated products.
3.4 Creating Multiple Predictions for Suggesting Products to each Customer

The best predictions for individual customers, based on past purchases can be known with the help of multiple predictions query.

Multiple Predictions Query is created as shown in screenshot 7 and predictions for individual customers is obtained as shown in screenshot 8, where party ID is the contact numbers of the customers who make the purchases and expression gives the products that can be suggested to these customers based on past purchases. This can also be helpful for loyalty customers and for making certain suggestions to them.
4. Inferences

4.1 Conclusion
The study analyses the pattern of consumer buying behaviour of products of a lifestyle store. The software proves to be useful for retailers to understand the purchasing behaviour of their customers and gives valuable insights relating to the formation of the basket. It helps in product assortments, refilling of the stocks for the likely items sold, make promotions based on likely items sold with a particular category, bundling of the products, give discounts to prompt the customers to buy the products. The increase in parameters used in the software increases the efficiency of the analysis. Retailers can use the analysis for devising strategies and to give suggestions to loyal customers.

Business Intelligence Development Studio software can be used for analysis based on past sales data by using market basket analysis. It can be effectively used for optimizing the patterns associated with dynamic behaviours of the transactions made by the customers while purchasing the products of a lifestyle store. It can help the company to improve sales, suggest products to the customers, cross-selling, and formulate promotions as per the results obtained form market basket analysis. This data mining tool can be used to improve strategic placement of the product on the shelves. The marketing department of the retailer also can understand customer purchasing behaviour better, so that they can design the Web site in such a way that products that tend to be purchased together appear together. In addition it also helps the advertising and also for forming marketing strategies. It can prove profitable and helps in making decisions that add value to the customer shopping experience as well as to the organization.

4.2 Limitations
Market basket analysis can be more accurate when the dynamic softwares namely alteyx, archibus are used. The study has used static software. The scope of the study is limited for one particular store but can be extended for various stores and areas.

5. References