

Data Mining in Supermarket: A Survey

Daljeet Kaur

*Student, Department of Computer Engineering, Punjabi University,
Patiala, Punjab, India.*

Jagroop Kaur

*Department of Computer Engineering, Punjabi University,
Patiala, Punjab, India.*

Abstract

One of the most motivating areas of research is data mining that becomes continuously popular in supermarket analysis. For discovering new trends mining plays an important role in supermarket analysis which is helpful for all parties associated with this field. The process of data mining is to extract data by automatic or semi-automatic means. Data mining consists of artificial intelligence, machine learning and database management to extract new patterns for huge data sets and the knowledge associated with these patterns. So, we can use data mining in supermarket application, through which management of supermarket get converted into knowledge management.

Keywords: Data Mining, Supermarket, Association Rule, Cluster Analysis.

1. INTRODUCTION

Data mining is a process to find out interesting patterns, correlations and information from databases which is useful to make efficient future decisions [1]. It is used in wide range of area to predict future trends and behaviour analysis. It is most useful technique for motivating area of research because it has various application areas [2]. One of the main application areas of data mining is supermarket analysis. Supermarket analysis is used to discover patterns or correlations with in the set of items. It is used to invent the system for analyzing the habits of the buyer to find out

the correlations between one item to another item [3, 4]. The discovery of these correlations can help the determiner to promote the sales strategy by considering items frequently purchased together by customers [4].

2. DATA MINING MODELS: KEY TECHNIQUES

There are various types of databases and information repositories on which data mining can be performed. So, to perform data mining several techniques are available to mine data from data warehouse. The main function of data mining is either to create a descriptive model or a predictive. A brief overview of these two techniques is given as follows:

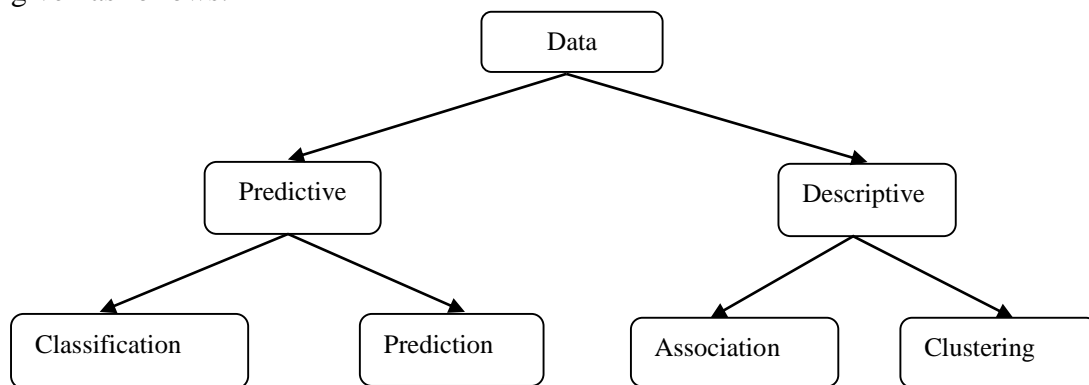


Fig. 1 Techniques of Data Mining

2.1 Predictive Model

Predictive models allow data miner to predict an unknown value of target variable. It covers following two basic data mining techniques:

- **Classification:** Classification is the process of finding a model (or function) that describes and distinguishes data classes or concepts, for the purpose of being able to use the model to predict the class of objects whose class label is unknown.
- **Prediction:** As many real world applications in the field of data mining are required to predict the future data states based on the past and current data stored in database.

Prediction is all about predicting the future state rather than a current state.

2.2 Descriptive Model

Descriptive models present main characteristics of data sets in concise form. It covers following two basic data mining techniques:

- **Association:** Association is the most popular data mining techniques and fined

most frequent item set. Association strives to discover patterns in data which are based upon relationships between items in the same transaction.

- Clustering: It is a data mining technique and a cluster is defined as a correlation of objects which are similar between them and dissimilar to the objects belonging to other cluster. This technique is used to make cluster of objects which have similar characteristics using automatic or semi-automatic means [5].

3. ASSOCIATION RULE

An association rule mining is introduced in data mining to find out hidden patterns in large data sets and drawing inferences on how a subset of items impact the presence of another subset. An association rule is one of the forms $A \rightarrow B$, where A is an “antecedent” (if part) and B is the “consequent” (then part). Here variables A and B are the item sets and the rule $(A \rightarrow B)$ means that customer who purchase an item set A are expected to purchase an item set B with the probability %c, where c is called confidence[6]. Interestingness measures of association rules are support and confidence.

Support(S): It is defined as the ratio of occurrence of two items and total number of transactions.

$$S(A, B) = \frac{\text{No. of Transactions containing both A and B}}{\text{Total No. of Transactions}}$$

Confidence(C): It is defined as ratio that how many instances satisfy the rule of an antecedent.

$$C\left(\frac{B}{A}\right) = \frac{\text{No. of Transactions containing both A and B}}{\text{No. of Transactions containing A}}$$

Let us consider the following example:

Customer ID	Item Purchased	Item Purchased
1	Milk	Bread
2	Bread	Butter
3	Coffee	Newspaper
4	Snacks	Cold Drink

Now, if A is “Purchased Milk” and B is “Purchased Bread”, then support and confidence are calculated as:

Support = $S(A \text{ and } B) = 1/4$

Confidence = $C(B/A) = 1/1$

Item sets that satisfy minimum support and minimum confidence are known as strong association rules [7, 8].

4. CLUSTER ANALYSIS

Clustering plays an important role in analysis of data and data mining applications. Cluster analysis is a fundamental operation in data mining to find out similar type of objects from large amount of database. Cluster is defined as a group of data objects that are similar to one another within the same cluster and are dissimilar to the objects in the other clusters. Basically cluster is in the form of ordered list of the data which have the similar characteristics. Cluster analysis is able to find out clusters irrespective of their shape. It falls under unsupervised learning technique. Most important requirements of clustering algorithms are scalability, ability to deal with noisy data, imperceptible to the order of input records.

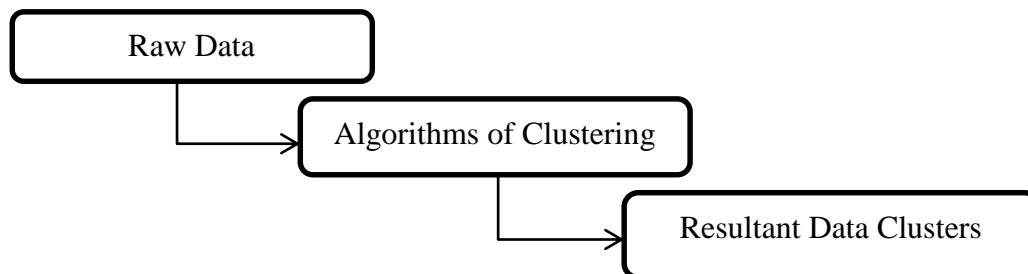


Fig. 2 Stages of Cluster Analysis

Clustering plays an important role, as from a practical point of view it is used in various data mining applications such as marketing, CRM, medical diagnostic, information retrieval and text mining, web analysis and many others. It is a machine learning technique used to put similar data elements into related groups without having any prior knowledge of group definitions.

5. COMPARISON OF ASSOCIATION AND CLUSTERING

The below given table gives a comparison between two basic data mining techniques, association and clustering based on the following parameters:

Table 1: Comparative Analysis of Association Rule and Cluster Analysis

S.No.	Parameters	Association Rule	Cluster Analysis
1	Technique	Supervised learning.	Unsupervised learning.
2	Specification	From the given dataset, find which item inside the dataset predict another item in the dataset.	From the given set of items, find connected subsets of items.
3	Representation	Combinatorial/Symbolic in nature.	Geometrical structure.
4	Data needs	Uses Labelled data.	Uses unlabelled data.
5	Algorithms	Apriori, Eclat, FP growth.	Centroid Clustering, Density based Clustering, hierarchical Clustering, Distribution based Clustering.

6. APPLICATIONS OF SUPERMARKET USING DATA MINING

Supermarket assembles sales data for gathering customer's information. After gathering information regarding sales data, this data get stored into data warehouse including a group of superintendent, buyer, merchant in supermarket expose the customers internal insight to identifying customer's buying patterns and trends for retaining, promote sales strategy and reduces profitable cost.

The applications of supermarket using data mining technique fall into the below given aspects:

1. **Consumer Aspects:** First step of data mining is to pick out the aspect depiction of consumer. Supermarket will never be interested to understand the consumer's information such as address, age, income and other basic information, but also understand consumption level of consumer and consumer trends and also the consumer's home situation will be the goal of data mining.
2. **Auspicious Consumer:** By analyzing consumer behaviour, operators of supermarket will group the consumers into different class with spending and stability. On the basis of the consumption level of the consumer, some of the supermarket operators will incline the "Golden Card" consumer, "Silver Card" consumer and some other grades. On the other hand, they establish profitable marketing strategy in order to retain customers, such as cash bask, discount and other policies.
3. **Consumer Target:** Supermarket operators contact consumers to identify the aspects of the consumer most concern about by analysis to carry out targeted marketing

campaigns. From time to time, they communicate with consumer through text messages and mail to send information regarding supermarket of interest to users in order to promote their sales.

4. Improve Customer Relationship Management: In this era, there is increasingly excessive competition among supermarket operators. The development of supermarkets depends upon stable consumers. In order to attract more consumers, types of promotions are put into use for building a suitable relationship with consumers [9, 10].

7. CONCLUSION

As the development of economy is going on, there is rapid increase in the application of data mining in retail industry, basically in large scale supermarkets. So, it becomes very important to establish appropriate model for data mining technology to provide decision making process in supermarket. So, there is wide usage of data mining technology in number of such type of applications.

REFERENCES

- [1] Qiankun Zhao, Sourav S. Bhowmick. (2003). *Association Rule Mining: A Survey*.
- [2] Jagmeet Kaur, Neena Madan. (2015). Association Rule Mining: A Survey. *International Journal of Hybrid Information Technology*.
- [3] Warnia Nengsih. (2015). A Comparative Study on Market Basket Analysis and Apriori Association Technique. *International Conference on Information and Communication Technology (ICoICT)*. IEEE.
- [4] Djoni Haryadi Setiabudi, Gregorius Satia Budhi, I Wayan Jatu Purnama, Agustinus Noertjahyana. (2011). Data Mining Market Basket Analysis' Using Hybrid-Dimension Association Rules, Case Study in Minimarket X. *International Conference on Uncertainty Reasoning and Knowledge Engineering*. IEEE.
- [5] Phani Prasad J, Murlidher Mourya. (2013). A Study on Market Basket Analysis Using a Data Mining Algorithm. *International Journal of Emerging Technology and Advanced Engineering*.
- [6] Jeetesh Kumar Jain, Nirupama Tiwari, Manoj Ramaiya. (2013). A Survey: On Association Rule Mining. *International Journal of Engineering Research and Applications (IJERA)*.
- [7] Surbhi K. Solanki, Jalpa T. Patel. (2015). A Survey on Association Rule Mining. *International Conference on Advanced Computing & Communication Technologies*. IEEE.
- [8] Yunkang Yue, Tiebao Zhang, Xingbing Xu. (2010). Analysis on the Effect of Data-mining to Supermarket. *International Conference on Computer Science & Education Hefei, China*. IEEE.

- [9] Zhang Aiguo, Jiang Lanling, Song Ping. (2011). Application of Data Mining in Supermarket. *International Conference on Consumer Electronics, Communications and Networks (CECNet)*. IEEE.
- [10] Smita, Priti Sharma. (2014). Use of Data Mining in Various Field: A Survey Paper. *IOSR-Journal of Computer Engineering (IOSR-JCE)*.
- [11] (n.d.). Retrieved from <https://www.quora.com/What-is-the-difference-between-clustering-and-association-rule-mining>
- [12] Jiawei Han and Micheline Kamber. (2006). *Data Mining: Concepts and Techniques*. Diane Cerra.
- [13] Dunham, M.H. (2003). *Data Mining: Introductory and Advanced Topics*. Pearson Education.

