

An Empirical Investigation on Kohonen Clustering in Indian Retail Industry

Ruchi Agarwal
Birla Institute of
Technology, India

Jayanthi Ranjan
Institute of Management
Technology, India

Tarun Pandeya
Indira Gandhi National
Open University
(IGNOU), India

S. L. Gupta
Birla Institute of
Technology, India

ABSTRACT

Kohonen clustering is one of the important functions of data mining. From the aspect of data mining, clustering research extracts valuable knowledge from large data sets intelligently and automatically. Kohonen clustering was proposed along with the development of databases and the emergence of data mining and Knowledge discovery technology. Kohonen clustering is applied in many areas, such as: pattern recognition, marketing, market segmentation and so on. In this paper, an empirical investigation was done using a data mining tool Clementine (a data mining tool of SPSS) and Kohonen neural network clustering algorithm to analyze the real sales database of the Indian retail organization, in order to find out the clusters of similar product categories.

Keywords: Kohonen Clustering, Data Mining

1. INTRODUCTION

Retail in India has emerged as one of the most dynamic and fast paced industries with several players entering the market. The data that retail business collect about their customers is one of the greatest assets of it. Buried within this vast amount of data is all sorts of valuable information that could make a significant difference to the way in which any business organization run their business, and interact with their current and prospective customers and gaining the competitive edge on their competitors. Retail industry is using Information Technology (IT) for generating, storing and analyzing mass produced data not only for operational purposes but also for enabling strategic decision making to survive in a competitive and dynamic environment. Data Mining helps in reducing information overload along with the improved decision-making by searching for relationships and patterns from the huge dataset collected by organizations. It enables a retail industry to focus on the most important information in the database and allows retailers to make more knowledgeable decisions by predicting future trends and behaviors. The Kohonen node generates a type of neural network that can be used to cluster the dataset into distinct groups. When the network is fully trained, records that are similar should appear close together on the output map, while records that are different will appear far apart. You can look at the number of observations captured by each unit in the generated model to identify the strong units. This may give you a sense of the appropriate number of clusters.

The arrival of retail boom caused the global technology vendors to quickly get into the marketplace with solutions that claim to make retailers' lives simpler. Retailers have to put in great efforts to really know their customers. Retail industry emphasized on quick delivery of customer focused services (offers, promos, etc) since adapting to customer needs in a very limited period of time is also very important. Retailers continuously get the advantage from information collected from

customers' transactions. Hence requirements of retail, technology wise would encompass business intelligence, data mining/warehousing, and other similar technologies since using these, retailers can constantly benefit from newly observed trends based on user purchases [29]. The changing consumption patterns trigger changes in shopping styles of consumers and also the factors that drive people into stores [15]. Hou and Tu [13] addressed that the managers in the contemporary marketing must importantly identify potential customer relationships to positively affect corporate performance. Ranjan and Bhatnagar [24] opined that the optimization of revenue can be accomplished by a better understanding of customers, based on their purchasing patterns and demographics, and better information empowerment at all customers touch points, whether with employees or other media interfaces. With the retail boom, companies are likely to deploy IT tools that help them enhance the end-customer's experience. Jones and Ranchhod [14] expressed that the strategic focus is required on the real complexity of the relationship that organizations are initially able to establish with customers. Sangle and Verma [28] opined that the customer relationship management unites the potential of marketing strategies and IT to create profitable, long-term relationships with customers and helps in enhancing the opportunities to use data and information to both understand customers and co-create value with them.

2. INDIAN RETAIL INDUSTRY

Retailers face a dynamic and competitive environment. With increased globalization, market saturation, and increased competitiveness through mergers and acquisitions, retailers are seeking competitive advantages by better managing customer relations through database management. This is not a new concept but seeking competitive advantage through improving relationships with customers has taken on new life. Rogers [26] addresses that the companies recognize that customer relationships are the underlying tool for building customer value, and they are finally realizing that growing customer value is the key to increasing enterprise value. The retail sector is a sector that is growing rapidly in the global, as well as in the Indian scenario.

The Indian retail industry is ranked among the ten largest retail markets in the world. The changing lifestyle of the Indian consumer makes it imperative for the retailers to understand the patterns of consumption. The changing consumption patterns trigger changes in the shopping styles of consumers and also in the factors that drive people into stores [15]. The attitudinal shift of the Indian consumer in terms of 'choice preference', 'value for money' and the emergence of organized retail formats have transformed the face of retailing in India. Overall retail business efficiency in India is 2%, even though India is considered as one of the countries where IT is booming [22]. Rising incomes, increased advertising, and a jump in the number of women working in the country's urban centers have

made goods more attainable and enticing to a larger portion of the population. At the same time, trade liberalization and more sophisticated manufacturing techniques create goods that are less expensive and higher quality [12].

Retail is India's largest industry, accounting for over 10 per cent of the country's GDP and around eight per cent of the employment. The market size of Indian retail industry is about US \$312 billion. Organized retailing comprises only 2.8 per cent of the total retailing market and is estimated at around US\$ 8.7 billion. The organized retail sector is expected to grow to US \$ 70 billion by 2010 [30]. Estimates predict that the overall size of the retail sector in India is expected to touch US\$427 billion by 2010 and US\$637 billion by 2015 with the modern segment expected to account for 22 per cent by 2010, up from the present four per cent. Currently the size of the Indian retail sector is to the tune of US \$ 328 billion with unorganized retailing in a dominating mode. Organised retailing contributes to roughly around 4% of the whole market [7]. Retailers of multiple brands operate via a wholesale model 'cash-and-carry' or a franchise. The Indian retailing industry has seen phenomenal growth in the last five years (2001–2006). Organized retailing has finally emerged from the shadows of unorganized retailing and is contributing significantly to the growth of the Indian retail sector. There are two ways of making money in retailing: (i) making money on every transaction or (ii) selling to the customer several times. To make money on every transaction procurement and logistics processes and competencies are crucial. However, making money on every transaction is difficult. First time customers often buy promotional goods with low or negative margin [9]. Retail Profitability is now connected to customer insight. It is no secret that companies across different industries are adopting powerful new tools to get, keep, and grow relationships with their best customers. Years ago, the best retailers were those that could generate pedestrian traffic and had well merchandised stores offering fresh displays and good service. Now that we have entered the information age, leading retailers are leveraging a traditionally overlooked asset—their customers—and using this insight to focus their merchandising, marketing and customer service offerings into a powerful, integrated brand offering.

Retailers that successfully integrate their offerings into “one face” to the customer will be able to build two-way relationships with customers that create strong switching costs, thereby establishing a sustainable competitive advantage [23]. Customers know a great deal about the places they shop, and they certainly have a 360 –degree view of a retailer, whatever the interpretation of that may be. Retail data warehouse can be used to put yourself in your customer's shoes and get a 360-degree view of the business from their perspective [27].

A. Retail Sales in India

Retailing in India is witnessing a huge revamping exercise as can be seen in the graph shown below in Figure1. Most of the growth is in two sectors: Food and apparel (see Figure1).

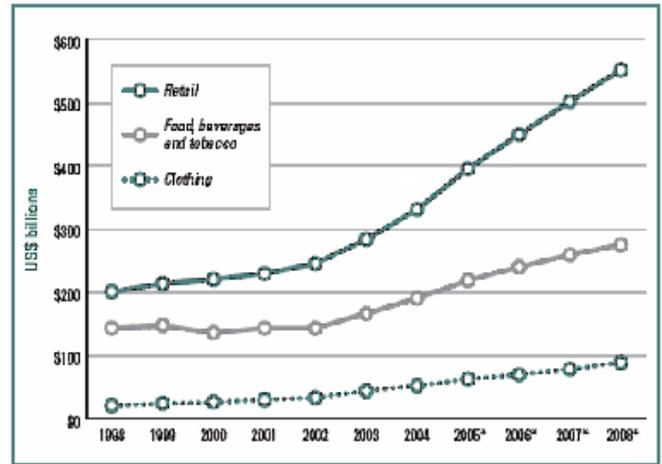


Figure1: Retail sales in India

Source: Economist Intelligence Unit and A.T. Kearney analysis

*Data for 2005-2008 is based on estimates

(Adopted from website http://www.atkearney.cz/shared_res/pdf/GRDI_2005_India.pdf)

B. Key Players in the Indian Retail Sector

The untapped scope of retailing has attracted superstores like Wal-Mart into India, leaving behind the kiranas that served us for years. Such companies are basically IT based. The other important participants in the Indian Retail sector are Bata, Big Bazaar, Pantaloons, Archies, Cafe Coffee Day, landmark, Khadims, Crossword, Shoppers' Stop Ltd., Big Bazaar, Westside (Tata Group), Future Group, Trent Ltd., RPG Retail, Vishal Retail Ltd., Bata India Ltd., Provogue India Ltd., Videocon Appliances Ltd., I.T.C. Ltd., Godrej Agrovert Ltd., DCM- Hariyali Kisaan Bazaar, Pantaloon Retail, Reliance Retail, Pizza Hut, etc. There is almost no discrimination done in terms of pricing or promotions, but there is need to differentiate amongst these firms. The use of IT technology gives that point of differentiation. In any organization, if the effective use of IT is prevalent, then it will definitely be offering something different from others.

A. T. Kearney analysis in 2005 addressed that supermarkets are the most popular format today still they are scarce, even in major urban centers (see Figure 2). In the coming years hypermarkets, which will offer lower prices and more choices, are expected to dominate.

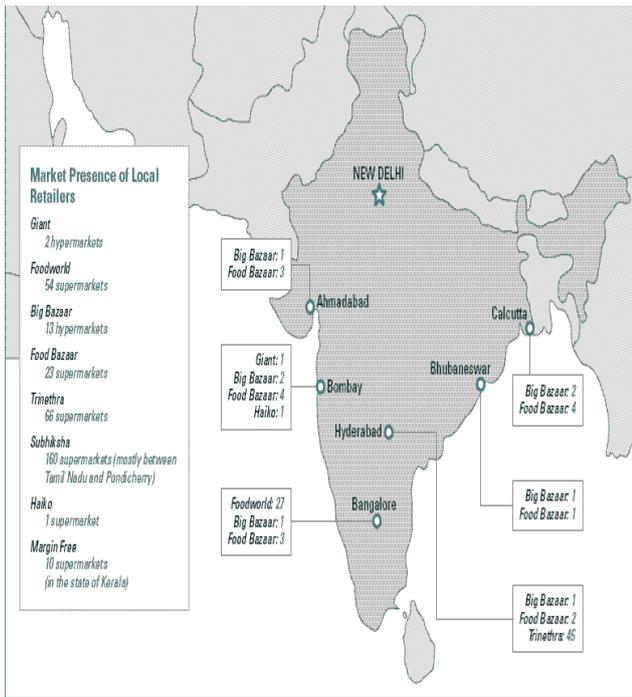


Figure 2: India's growing number of domestic grocery chains and early foreign entrants

Source: A. T. Kearney (Adopted from website http://www.atkearney.cz/shared_res/pdf/GRDI_2005_India.pdf)

C. The Growth Drivers

The Indian retailing sector is at an inflexion point where the growth of organized retail and growth in the consumption by Indians is going to adopt a higher growth trajectory. The Indian population is witnessing a significant change in its demographics. A large young working population with median age of 24 years, nuclear families in urban areas, along with increasing working-women population and emerging opportunities in the services sector are going to be the key growth drivers of the organized retail sector [6].

The Indian Retail growth can be attributed to the several factors including

- Demography Dynamics: Approximately 60 per cent of Indian population below 30 years of age.
- Double Incomes: Increasing instances of Double Incomes in most families coupled with the rise in spending power.
- Plastic Revolution: Increasing use of credit cards for categories relating to Apparel, Consumer Durable Goods, Food and Grocery etc.
- Urbanization: increased urbanization has led to higher customer density areas thus enabling retailers to use lesser number of stores to target the same number of customers. Aggregation of demand that occurs due to urbanization helps a retailer in reaping the economies of scale.
- Covering distances has become easier: with increased automobile penetration and an overall improvement in the transportation infrastructure, covering distances has become easier than before. Now a

customer can travel miles to reach a particular shop, if he or she sees value in shopping from a particular location.

3. DATA MINING

Data Mining, a synonym to “knowledge discovery in databases” is a process of analyzing data from different perspective and summarizing it into useful information. It reveals patterns and trends that are hidden among the data. It is often viewed as a process of extracting valid, previously unknown, non-trivial and useful information from large databases [25]. Noonan [21] addresses that data mining is a process for sifting through lots of data to find information useful for decision making. Data mining helps in predicting the future of your business. Today, data mining can make the difference in every industry and organization throughout the world. The data can be mined and the results can be used to determine not only what the customers want, but to also predict what they will do. West [33] opinions that by relying on the power of data mining, retailers can maintain the consistency and accuracy of their underwriting decisions; they can significantly reduce the impact of fraudulent claims; and can have a better understanding of their customer's wants and needs. Data mining can be used to control costs as well as contribute to revenue increases [30].

4. SEGMENTATION

Market segmentation is one of the most fundamental strategic marketing concepts. The better the segment(s) chosen for targeting by a particular organisation, the more successful the organization is assumed to be in the marketplace. The basis for selecting the optimal market segment to target is a (number of) segmentation solution(s) resulting from partitioning empirical data. Therefore the quality of groupings management chooses from is crucial to organisational success and requires professional use of techniques to determine potentially useful market segments. Thus, the methodology applied when constructing ([19],[32],[5]) or revealing([10],[8],[20] ,[2]) clusters from empirical survey data becomes a discriminating success factor and potential source of competitive advantage.

Segmenting a marketing target makes it easier to adapt the offer (price, service) and the communication to the economic value of the segment and to their sensitivities. There are many useable variables to separate the groups and they can be classified into different categories: attitudes, socio-demographic characteristics, geographical, frequentation of the media and distribution channels, relationship with the company and purchasing behavior [2]. This last group of variables has become the object of a growing interest because the information allies availability, precision, and certainty with a real behavior on behalf of the target [18]. The interest for buying behavior is further strengthened when the decision-maker no longer has to manage a few products, as a producer situation at the heart of the traditional marketing approach, but a large number of products, even a full distributor's assortment [4].

The Segmentation module is recommended in cases where the specific result is unknown (for example, when identifying new patterns of fraud, or when identifying groups of interest in your customer base). Clustering models focus on identifying groups of similar records and labeling the records according to the group to which they belong. This is done without the benefit of prior knowledge about the groups and their characteristics, and it distinguishes clustering models from the other machine-learning techniques available in Clementine- there is no

predefined output or target field for the model to predict. There is no right or wrong answers for these models. Their value is determined by their ability to capture interesting groupings in the data and provide useful descriptions of those groupings. Clustering models are often used to create clusters or segments that are then used as inputs in subsequent analyses (for example, by segmenting potential customers into homogeneous subgroups).

5. KOHONEN CLUSTERING

Kohonen neural network was proposed in 1981. It is also called self-organizing feature maps neural network clustering method. It gets the clusters of data set by using iterative algorithm to optimize the objective function ([16], [35]). Clustering technique named Kohonen network (also called self-organizing map) is very different from statistical clustering methods such as K-Means. This technique is based on neural network concepts. They cluster or segment data on the basis of patterns of the input variables so that similar patterns (i.e., observations) are grouped together. In other words, a kohonen network can be deemed to be a system of nodes where each node gathers observations that are similar. That is, each node forms a cluster. An interesting feature of a Kohonen network is that the clustering results can be represented on a map (or grid) of clusters, where clusters that have more similar profiles are closer together. Kohonen neural network is one of the most popular neural network methods for cluster analysis. Its goal is to represent all points in a high-dimensional source space by points in a low-dimensional (usually 2-D or 3-D) target space, such that the distance and proximity relationships are preserved as much as possible. The method is particularly useful when a nonlinear mapping is inherent in the problem itself [11].

A Kohonen map, a neural network, allows the projection and clustering of data for which the proximity presents a meaning or interest. Kohonen neural network is a two layers feed-forward neural network, it has input layer and output layer. The number of neurons of input layer is M , which amounts to the dimension of input sample vector. The neurons of output layer are competitive output neurons, whose values are in $\{0, 1\}$. After self-organizing learning, Kohonen neural network makes the density of the connection weight vectors consistent with the probability distribution of input model, which means the density of the connection weight vectors could reflect the statistical features of input model. The layout of neurons in output layer has many forms; the typical one is two-dimension plane matrix ([34], [17]).

6. EMPIRICAL INVESTIGATION

The Indian retail organization under study is the fastest growing retail organizations in India. The retail organization data were collected and analyzed by applying various Data mining techniques and tools to extract hidden and useful information. We use a data mining tool Clementine (a data mining tool of SPSS) and Kohonen neural network clustering algorithm to analyze the sales data of the retail organization, in order to find out the clusters of similar product categories. The sales data contained division, major category, major category description, bill quantity, price per unit, gross amount, discount, net value. It consisted of 6,598 records. After applying Kohonen Algorithm six different groups were found. In each group the characteristics of product with similar characteristic is found. Out of six groups, one group of FMCG ($X=3, Y=0$) having maximum number of observations (2135 records) is selected to find the behavior of this group (Figure 3).

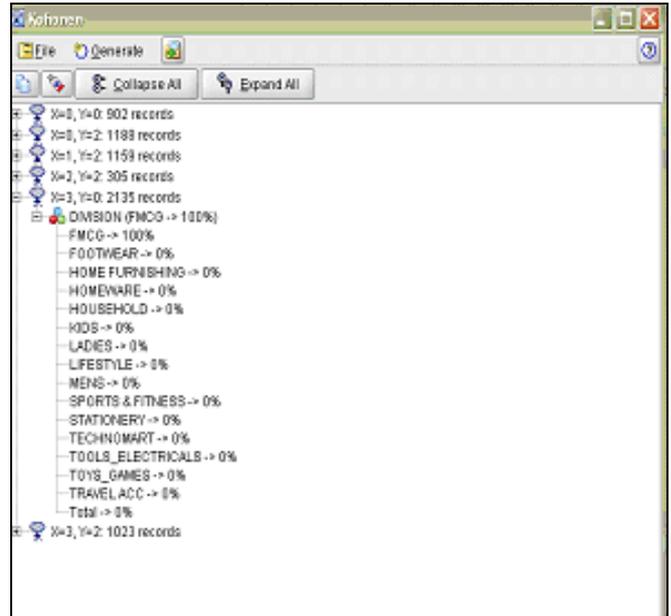


Figure 3: Kohonen clustering

The classification of FMCG product by using Clustering in terms of count & percentage are shown in the following Figure 4.

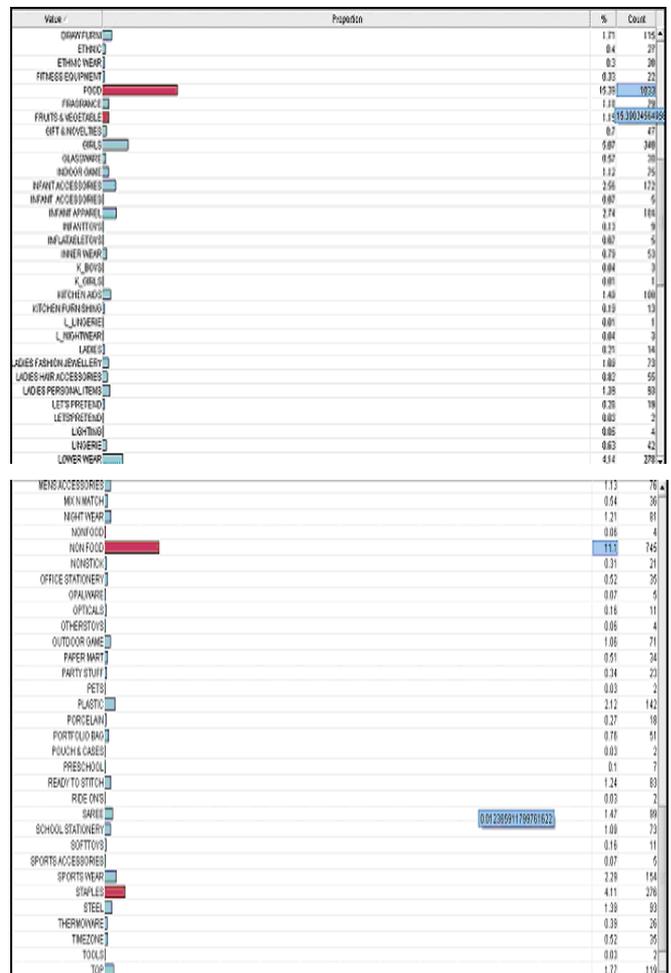


Figure 4: Distribution of FMCG cluster

This clearly shows that the fifth cluster of FMCG contains 15.39 percent 'Food' products with count of 1033 shows that in FMCG sector the Food is the one of the dominant category or in 5th group all the individuals buy food items frequently, 11.1 percent of 'Non Food' products with count of 745, 4.11 percent staples with count of 276 and 1.15 percent of fruits and vegetables with count of 75. The major portion is covered by 'Food' Products. So, in this group the people are generally buying the kitchen related items.

7. CONCLUSION

Data Mining based Kohonen Clustering in Indian Retail Industry helps in extracting important information from existing data and enables better decision-making throughout retail industry. In this paper, an empirical investigation was done using a data mining tool Clementine and Kohonen neural network clustering algorithm to analyze the real sales database of the Indian retail organization, in order to find out the clusters of similar product categories. The leading corporate in India have recognized that the business world is knowledge intensive with innovative or die approaches. All corporate utilize the technology for storing and managing enterprise related data. Those retailers that have realized the utility of data mining and are in the process of building a data mining environment for their decision-making process will reap immense benefit and derive considerable competitive advantage to withstand competition in future.

8. REFERENCES

- [1] Aldenderfer, M.S. and Blashfield, R.K. (1984). Cluster Analysis. Sage Series on quantitative applications in the social sciences. Beverly Hills: Sage Publications.
- [2] Blattberg R.C, Sen S.K. (1974) Market segmentation using models of multidimensional purchasing behavior, *Journal of Marketing* 38, pp.17-28.
- [3] Chen, M. (1996). Neural Network Model. Dalian University of Technology Press. Dalian
- [4] Desmet, P. (2002), Buying behavior study with basket analysis: pre-clustering with a Kohonen map, *European Journal of Economic and Social Systems* 15 N° 2 (2001) 17-30.
- [5] Dolni-ar, S. and Leisch, F. (2001). Knowing What You Get - a Conceptual Clustering Framework for Increased Transparency of Market Segmentation Studies. Paper presented at the Marketing Science, Edmonton, Canada.
- [6] Dominic K (2007), "Indian Retail: An Overview", *Network magazine*, March.
- [7] FICCI Retail Report 2007, www.ficci.com (accessed on 25 th June 2008).
- [8] Frank, R. E., Massy, W. F. and Wind, Y. (1972). Market Segmentation. Englewood Cliffs: Prentice-Hall.
- [9] Gunnesson, T. and Soderlund, K. (2001), "Creating Competitive Advantage in Mature e-Retail Markets", *Information Management, Handelshogskolan I Stockholm, Stockholm School of Economics*.
- [10] Haley, R. J. (1968). Benefit Segmentation: A Decision-Oriented Research Tool. *Journal of Marketing*, 32, 30-35.
- [11] Han, J., Kamber, M. (2001). *Data Mining Concepts and Techniques*. Academic Press, New York.
- [12] Hanna, J. (2004), "Ground-Floor Opportunities for Retail in India", *Harvard Business School Newsletter*.
- [13] Hou, J.J. and Tu, H.H.J. (2008) 'Customer relationship management strategy and firm performance: an empirical study', *Int. J. Electronic Customer Relationship Management*, Vol. 2, No. 4, pp.364–375.
- [14] Jones, S. and Ranchhod, A. (2007) 'Marketing strategies through customer attention: beyond technology enabled customer relationship management', *International Journal of Electronic Customer Relationship Management*, Vol. 1, No. 3, pp.279–286.
- [15] Kaur, P. and Singh, R. (2007), "Uncovering retail shopping motives of Indian youth", *Young Consumers: Insight and Ideas for Responsible Marketers*, Vol. 8, No. 2, pp.128–138.
- [16] Li, Z., Deng, Q. and Li, H. (2004). Kohonen SOFM Neural Network Evolution and Research. *Computer Engineering and Design*. 1729-1830.
- [17] Malone, J., etc.. (2005). Data mining Using Rule Extraction from Kohonen Self organising Maps. *Neural Comput & Applic* 15: 9-17.
- [18] Manchanda, P., Ansari A., Gupta S. (1999) The "Shopping Basket": A Model for multicategory purchase incidence decisions, *Marketing Science* 18, pp. 95-114.
- [19] Mazanec, J. and Strasser, H. (2000). *A Nonparametric Approach Market Segmentation: Foundations*. Berlin: Springer.
- [20] Myers, J.H. and Tauber, E. (1977). *Market structure analysis*. Chicago: American Marketing Association.
- [21] Noonan, J. (2000), "Data Mining Strategies", *DM Review*.
- [22] Pande, S. and Collins, T. (2007), "Strategic implementation of information technology to improve retail supply chain in India", *International Journal of Logistics Systems and Management*, Vol. 3, No. 1, pp.85–100.
- [23] Puleo, P. (2002), "How Retailers are using Customer Insight to Build Competitive Advantage", *Peppes & Rogers Group*.
- [24] Ranjan, J. and Bhatnagar, V. (2008), 'Data Mining tools: a CRM perspective', *International Journal Electronic Customer Relationship Management*, Vol. 2, No. 4, pp.315-331.
- [25] Rao, I. K. R. (2003), "Data Mining and Clustering Techniques", *DRTC Workshop on Semantic Web*, December.
- [26] Rogers, M. (2005), "Customer strategy: observations from the trenches", *Journal of Marketing*, Vol. 69 No.4, pp.262.
- [27] Ross, D. (2006), "Retail Data Warehouse, Analyzing your customers' 360 degree view of you", *Business Intelligence Network Newsletter*, May 16.
- [28] Sangle, P.S. and Verma, S. (2008) 'Analysing the adoption of customer relationship management in Indian service sector: an empirical study', *Int. J. Electronic Customer Relationship Management*, Vol. 2, No. 1, pp.85–99.

- [29] Sohoni, A. (2007), "Indian Retailers - Ready for Take Off?" available from <http://www.tech2.com/biz/india/features/retail/indian-retailers-ready-for-take-off/1313/0> (accessed on 02-august-2008).
- [30] Two Crows corporation, "Introduction to Data Mining and Knowledge Discovery", available at <http://www.twocrows.com/> (Accessed on 25/july/2008).
- [31] Vector, D. (2007), "Indian Retail Industry: Strategies, Trends and Opportunities 2007", available at <http://www.Marketresearch.com/product/display.asp?productid=1497236> (accessed on 4th July 2008).
- [32] Wedel, M. and Kamakura, W. (1998). *Market Segmentation - Conceptual and Methodological Foundations*. Boston: Kluwer Academic Publishers.
- [33] West, D. (2005) 'Enhancing value through data mining: Insurers can use data mining technology to improve their competitive position', *Insurance Networking News: Executive Strategies for Technology Management*, October.
- [34] Yan, P., Zhang, Ch. (2000). *Artificial Neuron Network and Simulated Evolution Computing*. Tsinghua University Press. Beijing.
- [35] Yuan, C. (2000). *Artificial Neuron Network and Application*. Tsinghua University Press. Beijing.