International Journal of
Science and Engineering Investigations vol. 1, issue 4, May 2012
ISSN: 2251-8843

Design of Knowledge Based System for Direct Marketing

Mehrdad Madhoushi[1], Abbas Namdar Aliabadi[2]
[1]Professor in Management, Department of Economy & Administrative Science, University of Mazandaran, Babolsar, Iran
[2]Master of Business Administration, Department of Economy & Administrative Science, University of Mazandaran, Babolsar
([1]madhoshi@umz.ac.ir, [2]ab.namdar@gmail.com)

Abstract- Since recommendation systems have been increasing gradually, it is difficult for decision makers to find the customers which interest them as well as representative lists. How to utilize meaningful information effectively to improve the service quality of recommendation system appears to be very important. The purpose of this paper is to provide recommendation system architecture to promote direct marketing services in electronic commerce. In the proposed architecture, a two-phase data mining process used by association rule and clustering methods is designed to generate a recommendation system. The process considers not only the relationship of a cluster of customers but also the associations among the information accessed. The recommendation supported by the proposed system architecture would be closely served to meet customers' needs. This paper not only constructs a recommendation system for decision makers to search customers but takes the initiative in finding the most suitable customers for them as well. Furthermore, managers are expected to contact with core customers from a limited budget to maintain and satisfy the requirements along with promoting direct marketing.

Keywords- *Direct Marketing; Data Mining; Framework.*

I. INTRODUCTION

The number of people using the internet has increased dramatically because of the widely accepted web environment. The internet has also rapidly accumulated a huge mass of data and has grown to be one of the most powerful means of information storage. In such a web environment, the concept of the direct marketing is interesting as it includes information technology which could produce plenty of complex data. The appearance of direct marketing system storing digitize data makes it possible to search more easily and conveniently. Traditionally, the direct marketing system used to play a passive role in that it merely provided customers. It is a crucial subject, however, for a marketing manager to think about how to guide customers to find what they want in an aggressive way and promote the selling rate at the same time.

This paper specifies how direct marketing systems can benefit from huge digital resources to enhance the quality of various services, and an approach is presented to identify valuable customers. In past research, most researchers have analyzed the content of direct marketing system. Then, they tried to discover the best techniques, as well. However, there are more and more tasks for direct marketing systems such as Post, EMAIL, SMS, etc. Under these situations, it is hard to analyze the content in them so as to identify customers' recommendation information.

New direct marketing system architecture is established in this paper to enable customized services and management. The association rules and clustering along with the data mining methods have been applied to discover the most adaptive customers [1]. First, records are clustered according to some characteristics of customers. The proposed approach utilizes the automatic clustering feature of the Kmeans Clustering Algorithm to form a user group with similar properties. Second, based on minimal support and confidence, the apriori Algorithm is proposed to exhibit the ability of locating the associated rules between subjects to generate recommending rules. The association rules will judge which customers in the same cluster are used as the basis of recommendation. Finally, an automatic online recommendation system is proposed. This paper not only constructs a real-time recommendation service, but also takes the initiative in finding the most suitable customers. Furthermore, managers are expected to contact with core customers from a limited budget to maintain and satisfy the requirements of customers along with promoting direct marketing systems. Direct marketing system could provide better services via the seamless integration of diverse approaches towards collecting, organizing, storing, accessing, and applying knowledge.

II. LITERATURE REVIEW

This section provides a general definition of data mining, which is the main component of the proposed methods. The definition of knowledge discovery in databases (KDD), given by Fayyad et al [5], is defined as the "nontrivial process of identifying valid, novel, potentially useful, and ultimately understandable patterns in data". In their view, the term "knowledge discovery in database" is used to denote the entire process of turning chaotic data into valuable knowledge. Also illustrated that the whole KDD process covered several key steps: data cleaning, data reduction and transformation (data integration), data mining, pattern evaluation and then knowledge discovered [9]. The overall process is outlined in

Figure 1. It is without doubt that data mining is considered as a central step in the process that involves extracting patterns from data. Additional steps are also essential to make certain that what we extract from data is useful knowledge. This paper also follows additional steps to demonstrate the proposed method in section 3.

The illiberal definition of data mining is the application of specific algorithms to uncover useful information from a large degree of data, and its purpose is to explore interesting knowledge from a database, data warehouse, or some other large information storage unit [9]. From a technical viewpoint, it combines a method of gathering and cataloguing information then proceeds to generate rule-like knowledge from a large amount of data. The more common model functions in the current data mining algorithms include classification, clustering, association rules, rule generation, summarization, dependency modeling, and sequence analysis [12].

Actually, data mining has been applied to various domains, such as customer service support, decision support, web intelligence, etc. [6]. The most well-known example is "bear and diapers", where the giant supermarket chain Wal-Mart wanted to know that which items were sold together from their huge sales records. They analyzed billions of transaction records and finally found that bears and diapers punched together could stimulate a purchase. In this paper, we applied association rules and clustering algorithms to extract similar interests customers and recommend to them. These are briefly explained below.

A. *Association rules*

Depending on the above-mentioned discussion of data mining applications, in direct marketing systems, one of the widely used examples of data mining is the discovery of association rules, especially market basket analysis. Huge amounts of customer purchase data are collected daily at the checkout counters of shopping malls and retailers are interested in purchasing the behavior of their customers. This technique, association rules, frequently found co-purchase items. Moreover, the uncovered relationships can be represented in the form of association rules. This provided retailers an opportunity for cross-selling their products to customers.

A general definition of association rule has been suggested: Let $L = I_1, I_2, \ldots . I_m$ be a set of object features. Let T be a set of records. Each record t is represented by a binary vector t [K] = 1 if t contains the feature I_k and t [k] = 0 if t does not contain the feature I_k $(k = \overline{1,m})$. Let X be a subset including some features from L, i.e. $X \subseteq L$. We say that the record t satisfies X if $\forall I_k \subseteq X, t[k] = 1$. and association rule is an expression in the form $X \to Y$, where $X \subseteq L$, $Y \subseteq L$, at that $X \cap Y = \emptyset$ [2].

The Apriori Algorithm was proposed by Agrawal and Srikant and is a famous algorithm in the mining association rule area [1]. Those things that appear simultaneously in certain events or data are called associations. Association rules mining aims to discover interesting associations or correlation relationships from large data sets [9]. In the research of Berry and Linoff, they were applied to analyze market baskets and would indicate which items should be bought at the same time [3]. Support and confidence are the two important parameters required to generate effective association rules. Support is the number of transactions with all the items in the rule, and confidence is the ratio of the number of transactions with all the items in the rule to the number of transactions with just the items in the condition [3].

B. *Clustering*

Clustering techniques work by identifying groups of users who appear to have similar preferences and dividing groups who have very different preferences. Unlike classification, the class label of each group is unknown. This is the way to naturally segment data into undefined groups, called clustering. In contrast, classification is assigning data into defined groups [4]. Briefly, a good clustering method produces high quality clusters with high intra-class but low inter-class similarity. However, how good a cluster is ultimately depends on the opinion of the user.

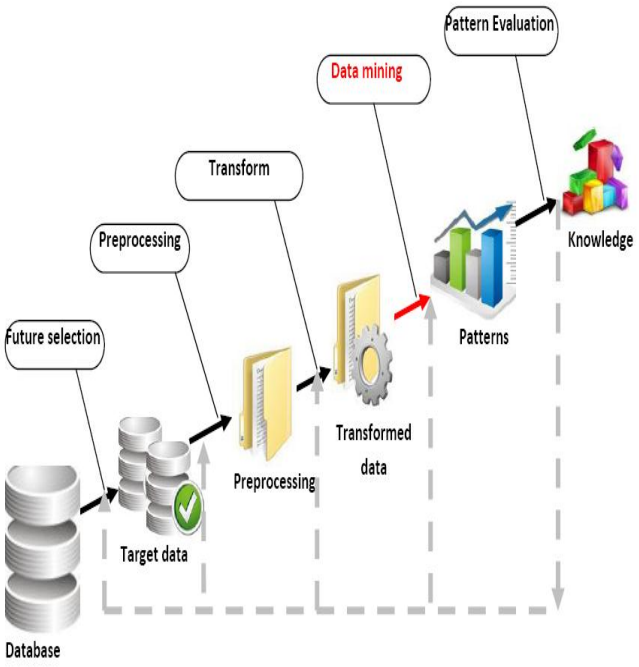

Figure 1: Knowledge discovery in database

III. PROPOSED FRAMEWORK

In the digital era, people can get information easily because of the development of information technology and the internet [14]. They can discover interesting information and digital content via surfing on the internet. When decision maker access the recommendation system, they use the function to discover the information that they want [10]. However, the

results do not always satisfy them [8]. In the past, there has been some research on direct marketing systems. However, last systems could not necessarily reflect the semantic expectations of users [11]. Therefore, further research tried to support some recommendation for users. In 1999, Luis led a project named the Active Recommendation Project at the Los Alamos National Laboratory. It was developing research on recommendation systems for large databases and the worldwide web, which adapt to the expectations of users [13]. Heylighen and Bollen proposed recommendation system based on Hebbian algorithms [7].

In this paper, we proposed a two-phase data mining recommendation system through analyzing the access behavior of customers. In the first phase, we proposed the Kmeans Clustering Algorithm as the data mining method and separated users into several clusters depending on access records. Customers who have similar interests and behavior are collected in the same cluster. In the second phase, we further analyzed the user records in the same cluster. We proposed association rules as the data mining method and discovered the associations among users' interests and access behavior. Then, the rules for the recommendation service were built. The framework is shown as Figure 2.

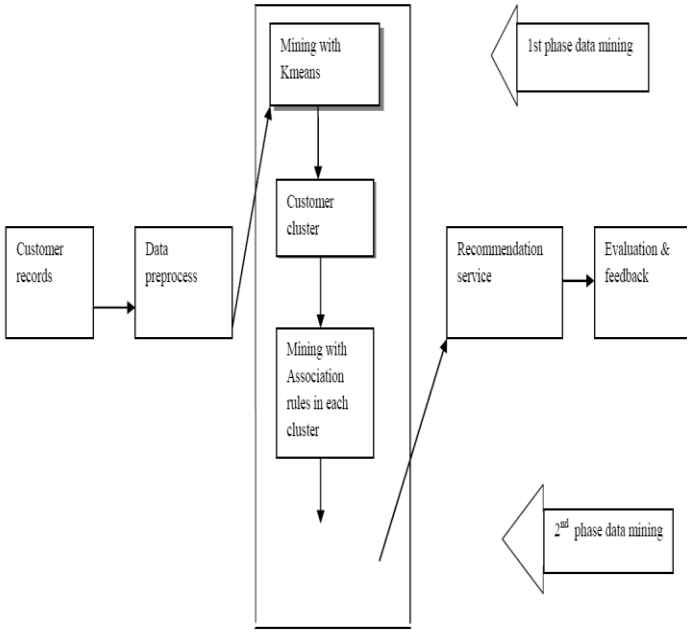

Figure 2: Proposed framework

In this paper, we propose a recommendation service model which combines the Kmeans Clustering Algorithm with association rules to discover customers with the same interests.

The first phase for the data mining method is to cluster data in each group. It is necessary to input the customer's features like as sex, age, job and etc. After iterating several rounds by the Kmeans Clustering Algorithm, the groups can be output.

The second phase for the data mining method is to find out the patterns of relationships in each cluster by association rules. Before processing the method, the data must be integrated. As mentioned, we can take the item set in a customer's basket for a transaction record. Likewise, shopping records are regarded as transaction records in a digital commerce database. A shopping record is defined as a set of records that a customer shops continuously in a period of time. For example, there are shopping records which record a customer's shopping list, Each attribute contains a serial number, customer's id, object id, object type, shopping date and cost. Association rules whose support and confidence exceed user-supplied thresholds are output by the Apriori Algorithm. Then, we can recommend products and services depending on association rules. The recommendation system architecture is shown in Figure 3.

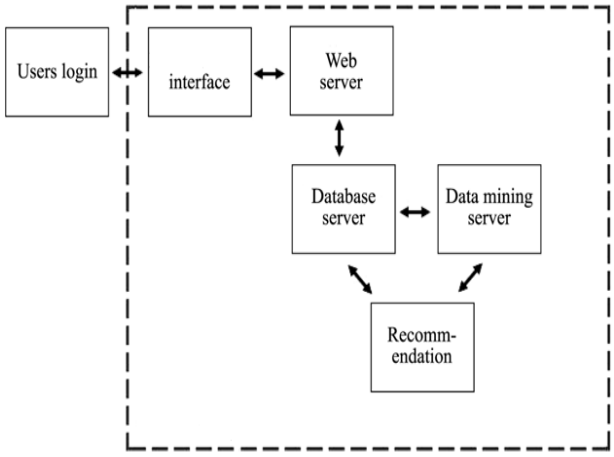

Figure 3: Direct marketing system architecture

IV. CONCLOUSIONS

As noted, in the digital era users can get information easily and conveniently via information technology tools. The powerful development of information technology makes the function of personal services more important than before. As per customers 'needs, it is worth providing valuable and proper information actively. This paper has proposed personalized direct marketing system architecture to enable personalized services and management in the digital commerce. The architecture applied data mining technology to support recommendation services for customers based on customers' interests. We proposed the Kmeans Clustering Algorithm and association rules to design a two-phase data mining process to generate recommendations. The process considered not only the relationships in the customers' cluster but also the associations among the information accessed. With the advanced filter, the recommendation supported by the proposed system architecture would closely meet customers' needs. This paper has not only constructed a recommendation mechanism for managers in searching customers from the web but has also taken the initiative in

finding the most adaptive customers for products and services. Business managers are expected to contact core customers from a limited budget to maintain and satisfy the requirements of customers along with promoting digital commerce services. Furthermore, more and more emphasis will be put on personal service by users in the future. The proposed system architecture could be also applied to real digital databases to support a personal recommendation service.

REFERENCES

[1] Agrawal, R. and Srikant, R., 1994. Fast algorithms for mining association rules, Proceedings of the 20th VLDB Conference, Santiago, 487-99.

[2] Agrawal, R., Imielinski, T. and Swami, A., 1993. Mining association rules between sets of items in large databases, Proceedings of the ACM SIGMOD Conference on Management of Data, Washington, DC, 207-16.

[3] Berry, M. and Linoff, G., 1997. *Data Mining Techniques for Marketing, Sales, and Customer Support*, Wiley, New York, NY.

[4] Edelstein, H., 2000. *Building Profitable Customer Relationships with Data Mining*, white paper, SPSS, Inc., Chicago, IL.

[5] Fayyad, U., Piatetsky-Shapiro, G. and Smyth, P. 1996, From Data Mining to Knowledge Discovery in Databases, American Association for Artificial Intelligence, Menlo Park, CA, 37-54.

[6] Fong, A.C.M., Hui, S.C. and Jha, G., 2002. *Data mining for decision support, IEEE IT Professional*, 4. 2, 9-17.

[7] Heylighen, F. and Bollen, J.2002. Hebbian algorithms for a digital library recommendation system, Proceedings of International Conference on Parallel Processing Workshops, Vancouver, 439-44.

[8] Han, J. and Chang, K.C.-C. 2002, *Data mining for web intelligence, IEEE Computer*, 35. 11. 64-70.

[9] Han, J. and Kamber, M. 2006. *Data Mining: Concepts and Techniques*, San Mateo, CA.

[10] Hui, S.C. and Jha, G. 2000. Data mining for customer service support, Information and Management, 38. 1, 1-13.

[11] Leguizamon, G. and Michalewicz, Z. 1999, A new version of ant system for subset problems, Proceedings of the Congress on Evolutionary Computation, Washington, DC, 1459-64.

[12] Mitra, S., Pal, S.K. and Mitra, P. 2002. Data mining in soft computing framework: a survey, IEEE Transactions on Neural Networks, 13. 1, 3-14.

[13] Rocha, L.M. 1999. TalkMine and the Adaptive Recommendation Project, Proceedings of the Association for Computing Machinery (ACM) – Digital Libraries, University Of California, Berkeley, CA, 242-3.

[14] Segel, L. and Cohen, I. 2001. *Design Principles for the Immune System and other Distributed Autonomous Systems*, Santa Fe Institute Series in the Sciences of Complexity, Oxford University Press, Oxford, 305-34.

Name: Mehrdad Madhoushi
Place and Date of Birth :Iran/ 11th may 1956
Education:
· Ph.D in Management Information System
· Business Strategic Management

Awards:
As highly qualified Researcher of the faculty in 2003.
As highly qualified Researcher of the university in 2005.
As highly qualified Researcher of the university in Mazandaran province 2008.
Current Position:
Associate professor of department of business management.

Name: Abbas Namdar Aliabadi
Place and Date of Birth :Iran/ 20 December 1980
Education:
· Master of Business Administration at University of Mazandaran
· Bachelor of Applied Mathematics at University of Tabriz

Awards:
Ranking 2nd in the MA program, 2008-2011
Winning a Best Paper award in 4th Iranian Data mining conference
Ranking 1st in the Computational Intelligence based Forecasting Energy. CIFE 2012.
Current Position:
Market Researcher at QS