CHAPTER 1

INTRODUCTION

Customer is the person or the organization who buys goods produced by a business. They are the primary concern for any business because they are the main source for profits. The competitive environment in which companies operate is forcing companies to adopt customer-centered strategies. Thus, organizations are, in essence, moving away from product or brand-centric marketing toward a customer-centric approach. Marketing managers can develop long-term and pleasant relationships with customers if they can detect and predict changes in the customer consuming behavior (Chopoorian et al 2001). Hence, analyzing the customer purchasing behavior is an important and critical research topic. The technologic development observed in recent years enabled companies to keep databases with customer related data. This allows the use of data mining techniques to extract knowledge from these databases in order to gain competitive advantage and remain at the leading edge (Berry & Linoff 2000). In this dissertation, this problem is investigated and algorithms are proposed and experimented with real data sets. This thesis develops a methodology to analyze customer purchasing behavior by applying data mining techniques.

1.1 RESEARCH MOTIVATION

In the 21st century, organizations are faced with more customers, more competitors and less time to react with the changes in the business environment. This means that understanding customers is becoming more
difficult and being able to keep up with changing customer behavior is more vital for organization survival. Thus, customer behavior analysis is an important task in Customer Relationship Management (CRM) because of the dynamic nature of the customers. It requires deep investigation and proper description of the customer characteristics. It helps business leaders to gain insight about the behavior of customers so that they can increase their profit by acting according to the customer characteristics (Berson et al 2000). The customer data contains various attributes like customer id, customer name, address, phone number, e-mail, date of purchase, amount of purchase and payment date. The attributes are to be chosen carefully so that the customer characteristics are portrayed correctly. The foremost purpose of the research depends on the proper collection of the main attributes of the customer. The task of analyzing the customer behavior based on the chosen attributes is subdivided into two subtasks namely customer segmentation and target customer analysis (Woo et al 2005).

Customer segmentation divides the customers into groups according to their purchasing behavior. It allows the business leaders to design and establish different strategies for each group of customers and thus maximize the value of customers to the business (Ling & Yen 2001). A key role of marketing is to identify the customers or segments with the greatest value-creating potential and then to target them successfully with corresponding marketing strategies. This helps in reducing the risk of losing these high lifetime value customers to the competitors (Banasiewicz 2004). Clustering algorithm, a data mining technique is widely used in customer segmentation (Liang 2010). Target customer analysis involves analyzing customer behavior or characteristics in each customer segment. It helps the business leaders to predict new customers’ behavior and to differentiate their discount rates or schemes for existing customers (Woo et al 2005). Rule induction algorithm, a
data mining technique is used for target customer analysis. It generates rules or patterns to describe the characteristics of the customers in each segment (Grzymala-Busse 1997). There are various clustering and rule induction algorithms each serving different purposes. The challenge involved in applying these data mining techniques to customer related data is that these techniques should handle uncertainty in customer behavior.

Whereas, Rough Set Theory (RST) by Pawlak (1982) has received a great deal of attention due to its stable results, ability to handle uncertainty, no requirement of domain expertise and no training time. It has the ability to find hidden patterns in data, evaluates significance of data, generates decision rules and interprets obtained result (Grzymala-Busse 2005). Therefore, this research was motivated by these advantages of RST.

1.2 OBJECTIVES AND SCOPE

As stated in the motivation, the objective of this research is to analyze the behavior of customers by:

- Designing an improved clustering algorithm based on RST for grouping the customers according to their purchasing behavior.
- Designing an improved rule induction algorithm based on RST for describing the customer characteristics in each group.
- Providing a consolidated solution for analyzing the customer behavior.

The scope of the work carried out is limited to showing how the proposed ideas can be implemented and tested on the collected customer data.
1.3 SUMMARY OF RESEARCH

First of all, an extensive literature survey of clustering algorithms is carried out. The study reveals that the rough set based clustering algorithm is advantageous. Therefore, the focus is given to the existing rough set based clustering algorithms. The existing rough set based clustering algorithms are mainly based on roughness calculation for selecting the clustering attribute. The roughness calculation may produce same value and hence results in difficulty of choosing the correct attribute for splitting the cluster. This major issue is considered in the proposed RST based clustering algorithm. Next, the clustering algorithm is used to segment the customers according to their purchasing behavior.

Further, the characteristics of segmented customers are described using the rule induction algorithm. Therefore, the rule induction algorithm using various techniques is studied. Based on this study, importance is given to the rough set based rule induction algorithms. As our main focus is to analyze the customer behavior, the proposed rough set based rule induction is designed in such a way that both the proposed clustering and rule induction algorithm together analyzes the customer behavior more accurately.

1.4 THESIS OUTLINE

The thesis is structured as follows:

The first Chapter contains the introduction. Chapter 2 discusses the basics of RST and its relevance in data mining. Chapter 3 describes the importance of CRM and then analytical CRM in detail. It also focuses on the relevance of data mining and its challenges in CRM.
Chapter 4 provides the literature survey of clustering algorithms based on its types. The categorical clustering algorithms are concentrated in this Chapter because the attributes used to describe the customer characteristics are categorical in nature. Further, the study reveals that RST is suitable and thus rough set based categorical clustering algorithms are analyzed in this Chapter.

Based on the above study, Chapter 5 proposes improved clustering algorithm based on RST. Real customer data is collected and customer segmentation is performed using clustering algorithms. The performance of the clustering algorithms is compared in this Chapter.

Chapter 6 studies the various ways of extracting the rules from the database and in particular, the rule induction algorithms based on RST are dealt in detail. Chapter 7 proposes rule induction algorithm based on RST by considering the demerits of the existing algorithms. The results of the clustering algorithm are used by the rule induction algorithm to analyze the behavior of customers in each cluster. The rule induction algorithms are compared using performance metrics. Chapter 8 summarizes the work presented in this thesis, highlights the significance of the contributions made and discusses directions for future work.