IDEA GROUP PUBLISHING
701 E Chocolate Avenue, Hershey PA 17033, USA
Tel: 717/533-8845; Fax: 717/533-8661; URL: www.idea-group.com
ITB7263

Chapter XIV

Understanding Credit Card Users' Behaviour: A Data Mining Approach

A. de Carvalho
University of Guelph, Canada and
University of São Paulo, São Carlos,Brazil

A. P. Braga
Federal University of Minas Gerais, Pampulha, Brazil

S.O. Rezende and E. Martineli
University of São Paulo, São Carlos,Brazil

T. Ludermir
Federal University of Pernambuco, Brazil

In the last few years, a large number of companies are starting to realize the value of their databases. These databases, which usually cover transactions performed over several years, may lead to a better understanding of the customer's profile, thus supporting the offer of new products or services. The treatment of these large databases surpasses the human ability to understand and efficiently deal with these data, creating the need for a new generation of tools and techniques to perform automatic and intelligent analyses of large databases. The extraction of useful knowledge from large databases is named knowledge discovery. Knowledge discovery is a very demanding task and requires the use of sophisticated techniques. The recent advances in hardware and software make possible the development of new computing tools to support such

This chapter appears in the book, *Heuristic and Optimization for Knowledge Discovery*, edited by Ruhul Sarker, Hussein Abbass and Charles Newton.

tasks. Knowledge discovery in databases comprises a sequence of stages. One of its main stages, the data mining process, provides efficient methods and tools to extract meaningful information from large databases. In this chapter, data mining methods are used to predict the behavior of credit card users. These methods are employed to extract meaningful knowledge from a credit card database using machine learning techniques. The performance of these techniques are compared by analyzing both their correct classification rates and the knowledge extracted in a linguistic representation (rule sets or decision trees). The use of a linguistic representation for expressing knowledge acquired by learning systems aims to improve the user understanding. Under this assumption, and to make sure that these systems will be accepted, several techniques have been developed by the artificial intelligence community, using both the symbolic and the connectionist approaches.

INTRODUCTION

The widespread use of databases and the fast increase in volume of data in these databases are creating a problem and a new opportunity for a large number of companies. These companies are realizing the necessity of making an efficient use of their stored databases.

Moreover, as a result of the "information technology revolution," storage and processing capabilities have faced an explosive increase in the last decades. Today, commercial and scientific applications easily produce gigabytes or terabytes of data in a few hours. These data hold variable information, e.g., trends and patterns, which can be used to improve business decisions and to optimize performance.

However, today's databases contain so much data that it has become almost impossible to manually analyze them for valuable decision-making information. In many cases, hundreds of independent attributes need to be simultaneously considered in order to accurately model systems behavior. Nowadays, this need for automatic extraction of useful knowledge from large amounts of data is widely recognized.

Data mining (DM) techniques are employed to discover strategic information hidden in large databases. Before they are explored, these databases are cleaned. Next, a representative set of samples is selected. Machine learning techniques are then applied to these selected samples.

This chapter investigates three different machine learning techniques. Two of these techniques are the symbolic learning algorithms C4.5 (Quinlan, 1993) and CN2 (Clark & Boswell, 1991). The other technique is a multilayer perceptron neural network (Rumelhart & McClelland, 1986) with a knowledge extraction technique, the TREPAN algorithm (Craven, 1996).

Despite their successful use on a large number of tasks, artificial neural networks (ANNs) have been much criticized for not presenting a clear indication of