Credit Card Users’ Data Mining

André de Carvalho
USP São Carlos, Brazil

Antonio P. Braga
UFMG, Brazil

Teresa Ludermir
UFPE, Brazil

MINING CREDIT RISK ANALYSIS DATA

The widespread use of databases and the fast increase in the volume of data they store are creating problems and new opportunities for credit companies. These companies are realizing the necessity of making efficient use of the information stored in their databases, extracting useful knowledge to support their decision-making processes.

Nowadays, knowledge is the most valuable asset a company or nation may have. Several companies are investing large sums of money in the development of new computational tools able to extract meaningful knowledge from large volumes of data collected over many years. Among such companies, companies working with credit risk analysis have invested heavily in sophisticated computational tools to perform efficient data mining in their databases.

Credit risk analysis is concerned with the evaluation of the profit and guaranty of a credit application. A typical credit risk analysis database is composed of several thousands of credit applications. These credit applications can be related to either companies or people. Examples of personal credit applications include for student loans, personal loans, credit card concessions, and home mortgages. Examples of company credits are loans, stocks, and bonds (Ross, Westerfield, & Jaffe, 1993).

 Usually, the higher the value of the credit asked, the more rigorous is the credit risk analysis. Some large companies have whole departments dedicated to credit risk analysis.

The traditional approach employed by bank managers largely depends on their previous experience and does not follow the procedures defined by their institutions. Besides, several deficiencies in the data set available for credit risk assessment, together with the high volume of data currently available, make the manual analysis almost impossible. The treatment of these large databases overcomes the human capability of understanding and efficiently dealing with them, creating the need for a new generation of computational tools and techniques to perform automatic and intelligent analysis of large databases.

Credit analysis is essentially a classification task that involves the evaluation of the reliability and profitability of a credit application. The application of data mining techniques for credit risk analysis may provide important information that can improve the understanding of the current credit market and support the work of credit analysts (Carvalho, Braga, Rezende, Ludermir, & Martineli, 2002; Eberlein, Breckling, & Kokic, 2000; He, Hawkins, Graco, & Yao, 2000; Horst, Padilha, Rocha, Rezende, & Carvalho, 1998).

The extraction of useful knowledge from large databases is called knowledge discovery in databases (KDD). KDD is a demanding task and requires the use of sophisticated computing techniques (Brachman & Anand, 1996; Fayyad, Piatetsky-Shapiro, Amith, & Smyth, 1996). The recent advances in hardware and software make possible the development of new computing tools to support such tasks. According to Fayyad, Piatetsky-Shapiro, Amith, and Smyth (1996), KDD comprises a sequence of stages:

1. Selection
2. Preprocessing
3. Transformation
4. Data mining
5. Interpretation/evaluation

It is then important to stress the difference between KDD and data mining (DM). While KDD denotes the whole process of knowledge discovery, DM is a component of this process. The DM stage is used as the extraction of patterns or models from observed data. KDD can be understood as a process that contains, at least, the steps of application domain understanding, data selection and preprocessing, DM, knowledge evaluation and consolidation, and use of the knowledge. At the core of the knowledge discovery process, the DM step usually
Related Content

Academic Virtual Meetings Management with a Synthesis Builder
[www.igi-global.com/article/academic-virtual-meetings-management-with-a-synthesis-builder/178573?camid=4v1a](www.igi-global.com/article/academic-virtual-meetings-management-with-a-synthesis-builder/178573?camid=4v1a)

Census Data for Health Preparedness and Response
[www.igi-global.com/chapter/census-data-health-preparedness-response/14265?camid=4v1a](www.igi-global.com/chapter/census-data-health-preparedness-response/14265?camid=4v1a)

Using RFID to Track and Trace High Value Products: The Case of City Healthcare
[www.igi-global.com/article/using-rfid-track-trace-high/3214?camid=4v1a](www.igi-global.com/article/using-rfid-track-trace-high/3214?camid=4v1a)

A New Technique for Estimating the Distribution of a Stochastic Project Makespan
Yuval Cohen and Ofer Zwikael (2012). *Project Management Techniques and Innovations in Information Technology* (pp. 33-47).
[www.igi-global.com/chapter/new-technique-estimating-distribution-stochastic/64953?camid=4v1a](www.igi-global.com/chapter/new-technique-estimating-distribution-stochastic/64953?camid=4v1a)