

A SAS® Based Data Mining Approach to Find Database Solutions in the Banking Industry

Rama Jampani, Chevy Chase Bank, Frederick, MD

ABSTRACT

Faced with intense competition and rising loan delinquency rates, most banks are exploring ways to use their data assets to gain a competitive advantage. This paper demonstrates how the recently emerging data mining technology, if properly utilized, could turn gigabytes of data residing in huge data warehouses into useful information that offers database solutions. The data mining tools and techniques enable end users to develop predictive models, such as neural network models, by capturing the most significant relationships and patterns in a set of data, thereby diminishing the need for statistical training in the process. Neural network models could be developed to prevent fraud, identify credit risks, and provide effective service to customers. The data mining technology encompasses a wide variety of tools and techniques. There is currently no single software package that does it all. However, the SAS system provides a very broad variety of tools upon which customized data mining software can be developed.

INTRODUCTION

The data routinely collected in a bank's myriad master files (operational/production systems) in the normal course of business represent a valuable asset. Faced with intense competition and rising loan delinquency rates, most banks are looking for ways to effectively manage and leverage these data assets to achieve a competitive advantage. As a first step, several banks have built or are in the process of building large data warehouses that contain vast amounts of data. But the data truths or database solutions are often buried deep within these databases. The rapidly emerging data mining technology could be used to unlock the intelligence hidden in the databases. If banks really expect good paybacks from the large investments that are being made in creating data warehouses or data marts, they may need to turn to data mining technology for finding effective database solutions.

For banks to effectively use the data mining technology, it is important that the technology be integrated into the business decision-making process. Data mining does not solve business problems by itself, but it could be viewed as a base technology upon which database solutions could be built.

In the next section of this paper the concept of data mining is discussed, and some of the most widely used data mining tools and techniques are presented. The later section explores the capabilities of the SAS system to implement a database mining system that offers effective database solutions to business problems confronting banks. Specifically, the use of the SAS system to bring to light information regarding patterns and trends in a set of data, which may have gone unnoticed using traditional tools, are discussed. Wherever appropriate, several business applications from the banking industry are provided to demonstrate how the data mining technology could help banks gain a competitive advantage. In the last section, the conclusions are offered by evaluating the effectiveness of the tools provided by the SAS system to design a data mining system.

DATA MINING TECHNOLOGY

Data warehousing is all about making better business decisions through access to the information that managers and analysts need. Data warehouses collect data about an entire organization and integrate and summarize in such a way that it can support decision support systems (DSS) and executive information systems (EIS) queries. Data mining, on the other hand, using artificial intelligence tools and techniques, can help spot patterns present in the data that simple queries and reports cannot reveal. Also, data mining puts more of the burden for finding insights from data on the system

rather than the managers or analysts. With the data mining tools available today, end users do not necessarily need to know what they are looking for as the system itself provides insights which would be invaluable in making better business decisions.

Some of the recently emerging data mining tools that are being increasingly used in the banking industry to find database solutions are the neural network and scoring models. Neural networks are predictive models based on principles that are similar to those of the human brain. In a network of nodes, each node receives input and then sends output to subsequent nodes based on what it has received.

Neural Networks

A typical neural network is organized into layers. At the lowest level, there is the input layer that contains the nodes through which the data are input into the network. At the top there is the output layer that generates the output interpreted by the user. For example, a bankruptcy predictor model may have credit history and spending patterns of the consumer as the input, and the output node may indicate whether or not the consumer would declare bankruptcy within a specified time frame. Between the input layer and the output layer, there could be one or more layers, which are called hidden layers. The output of the input layer is fed to the first hidden layer in the network and forms the input signals to this hidden layer. The output from the first hidden layer is then fed to the next and so on, until the signals reach the output layer, which in turn generates a signal interpreted by the user. For example, sample credit histories could train a neural network to distinguish credit-worthy individuals from credit risks.

A neural network is thus a statistical technique that calculates weights (score points) for predictor characteristics (such as age and income) by self-learning from data examples (such as good and bad loans). A neural network could be trained to detect fraud by reviewing examples of good and fraudulent transactions on a bank's portfolio. Banks can set different thresholds on the transaction to determine the type and severity of the follow-up action they will take on the account. A key contributor to the neural network's accurate detection is its ability to factor in cardholder history in determining the probability that a given transaction is fraudulent. Adaptation is a strength of a neural network, as they are inherently learning systems that automatically adjust to changes in behavior patterns that define fraud activity. Different neural network models could be developed to help banks in acquiring, servicing, maintaining and managing the bank's credit portfolio. Credit card issuers can develop neural network models that help reduce the target segment to as small as possible, so that the issuer can focus on customers that are most likely to use a given product and thus avoid a wasteful mass marketing campaign. It would also help issuers to offer different products that are most appealing to a given target segment. Likewise, models could be developed to identify customers who are most likely to leave the bank and go to a competitor or accounts that are most likely to go bankrupt.

SAS BASED DATA MINING APPROACH

In a traditional environment, the development of these models is done by a group of statisticians who are given a particular hypothesis to verify or disprove by using a sample population from the database. This approach puts an intermediary between the end user and the data, which often does not produce the desired results. This has led to the development of software that enable end users to build these models themselves. Some of the popular neural network tools include Marksman, from HNC Software Inc. in San Diego, CA, and the DataCruncher, from DataMind Inc. in Redwood City, CA. One of the major drawbacks of this kind of software is that they do not tell the users how the output is generated. Moreover, they are

not flexible to change the process of model development and have limited data manipulation capabilities.

Different data mining tools are better suited for different applications. There is no single best machine-learning or pattern recognition technique; different situations require the use of different data mining techniques such as associations, sequential patterns, classifiers, and clustering. For instance, neural network models work well when many of the predictors are partially redundant, while the tree-based models that are based on classification techniques are more appropriate at identifying a small number of significant explanatory variables from a large set of variables. There is currently no single high-power data mining software package that does it all.

The SAS system offers a very wide variety of data mining tools. The system includes a number of products such as SAS/STAT®, SAS/ETS®, SAS/INSIGHT®, SAS/GRAPH® and SAS macros, which provide several data mining tools that are invaluable in the development of models. The base SAS software provides the end users with effective data management capabilities such as massaging the data before being fed into a model. The SAS system also provides the SAS/AF® software upon which customized data mining software can be developed.

The SAS/STAT product offers a number of procedures that enable end users to use traditional statistical methods, such as cluster analysis, logistic regression, and discriminant analysis. The traditional methods help end users to understand the data and validate the output generated by other methods. Sometimes, simple data visualization may provide valuable insights. SAS/INSIGHT software has several features that facilitate graphical statistical visualization and the development of models based on the relationships discovered during the visualization process. End users can also use the SAS/SPECTRAVIEW® software to interactively view data in multiple dimensions. The SAS Institute's Professional Services Division offers a Data Mining Solution that includes use of developed SAS applications such as Neural Network Application, Research Application, and TREE Application. These applications provide a comprehensive set of data mining tools which, if effectively used, would enable users to understand complex relationships that exist in a given set of data.

CONCLUSIONS

The fast-emerging data mining technology can be used to identify patterns and spot trends in data buried deep inside the data warehouses or data marts. The data mining tools and techniques, if properly integrated into the decision-making process, could provide database solutions that give banks a sustained competitive advantage. Since some of the data mining tools have emerged fairly recently, a single high-power data mining software package that does it all is not yet available. The SAS system provides one of the most comprehensive set of tools upon which data mining software can be developed.

REFERENCES

- Chester M. (1993). *Neural Networks: A tutorial*. Englewood Cliffs NJ: Prentice-Hall
- Fayyad U.M., Piatetsky-Shapiro G., Smyth P and Uthurusamy R. (1996). *Advances in Knowledge Discovery and Data Mining*. Cambridge MA: AAAI Press/MIT Press
- Inmon, W. (1995). "Growth in the Data Warehouse." *Data Management Review*. Dec 1995.

ACKNOWLEDGMENTS

I would like to thank Laurie Liversidge for her encouragement, Richard Billig for his interest, and colleagues at Chevy Chase Bank for their support on this paper. Special thanks to Adrien Ndikumwami for providing valuable comments on a very short notice.

SAS, SAS/AF, SAS/ETS, SAS/GRAPH, SAS/INSIGHT, SAS/SPECTRAVIEW, and SAS/STAT are registered trademarks or trademarks of SAS Institute, Inc. in the USA and other countries. ® indicates USA registrations.

Other brand and product names are registered trademarks or trademarks of their respective companies.

Rama Jampani
5300 Spectrum Drive
Frederick, MD 21701
Voice(301) 620-8372
FAX(301) 620-8025
e-mail: rjampani@resi.towson.edu