DATA MINING ACROSS THE BUSINESS CURRICULUM

Cuong Le
Arkansas State University
Department of Computer & Information Technology
Box 130
State University, AR 72467
(870) 450-5706 Viet.Le@smail.astate.edu

John Seydel
Arkansas State University
Department of Computer & Information Technology
Box 130
State University, AR 72467
(870) 680-8072 jseydel@.astate.edu

ABSTRACT

Effective business decision making represents the appropriate application of knowledge, which is the compilation of good information, which in turn requires access to reliable data. Consequently, the more data available, the better the decision making, given an efficient means of processing those data into knowledge. Data mining is a powerful tool that can make this happen. Since data exist to support decision making throughout the organization, data mining should therefore be considered an essential tool to be used by all functional areas of a business. A proposal is presented here for using data mining as a unifying theme that may greatly improve the value of MBA graduates. The proposal involves an extensive multifaceted case to be used in several core courses, a comprehensive data set to support that case, and a low cost yet powerful and easy to use platform for performing data mining.

INTRODUCTION

A business is an organization of interconnected and interacting components, and its behavior may not be obvious from the properties of the individual parts. Businesses are interactively influenced by both internal and external factors, and it is consequently true that neither top-down nor bottom-up approaches are effective and efficient models for business decision making. Managers today therefore are required not only to understand business-related concepts but also to utilize operational data to make rational decisions that affect the entire organization.

Of course, all areas of business involve decision making based upon available data. For many years, however, the data "available" were rather limited, but now there is an abundance of such data. In fact, there are far more data than can be exploited by conventional methods of analysis. Furthermore, much of what has been done in the past by means of subjective assessment can now be accomplished with much more sophisticated methods and in a manner that is far more

reliable. This is possible because of the development of various data mining techniques and the software that makes those techniques possible.

Given the unquestionable increase in amount of data collected, along with numerous data mining applications in business, MBA students can be aided greatly by the grasp of data mining techniques. Much of the coursework in a traditional MBA program nevertheless proceeds as if no such data nor data mining capability exists. That coursework generally consists of qualitative approaches to planning, organizing, and controlling and uses traditional tools (e.g., standard statistical analysis) to accomplish that which can be approached quantitatively. While various tools associated with business analytics are nevertheless taught, they are typically relegated to "silos" that often never dispense their contents.

This need not be the case, however, as applications can easily be found for data mining techniques in nearly every aspect of a business curriculum. Furthermore, access to the software required has opened up greatly so that the cost of incorporating data mining across the MBA curriculum can be essentially nil. Data mining can thus be used as a tool that may greatly improve the value of MBA graduates by, as W. Edwards Deming (1985) urged, "breaking down barriers" across disciplines.

PROPOSAL OVERVIEW

A unifying data mining experience (DME) involving three core MBA courses – Management, Marketing, and Finance – is therefore proposed. These classes have been identified as particularly appropriate, since most MBA programs include such courses or variations thereof. Nevertheless, other subject areas (e.g., Operations) could also be involved. This integrative experience would be based upon a comprehensive data set and a *collection of cases* addressing the organizational data model represented by that data set. The "collection" of cases would however be more correctly termed a single *multifaceted* case, whereas each facet involves investigating the data from the perspective of one or more of the business functional areas (e.g., Finance). In addition, an optional comprehensive case could be used for a capstone experience.

Note that this represents a variation of what the Kelly School of Business at Indiana University (IU) currently does with its I-Core program at the undergraduate level (Indiana University, 2011). At IU, upon completion of the Business foundation courses (Accounting, Economics, etc.) but prior to beginning coursework involved with the various majors, students are required to complete an I-Core semester. That semester includes standard three-credit classes in Marketing, Operations, Finance, and Strategic Management, as well a one-credit class intended to integrate the material from the other three courses. All of these courses are built around an I-Core case that is addressed from three different perspectives, each corresponding to one of the standard Business courses.

Data to be used for the proposed DME would result from building data warehouses and data marts out of a number of databases serving the various functional areas of an organization. Data mining would consequently be an integral part of each participating course and would provide

for continuity among courses in the MBA curriculum. The data could be mined one way to address questions associated with accounting and/or finance course topics and then mined in a different manner to deal with issues pertaining to marketing, human resources, information systems, etc.

One challenge associated with the proposed DME is that, at many (if not most) universities, MBA courses are taken in no particular sequence, other than the capstone course required during the student's final term. Similarly, for various reasons (e.g., part-time students, lack of critical mass, etc.) it is generally not feasible to require MBA students to take a block of more than two courses simultaneously as is done at IU with their I-Core. The proposed DME therefore does not rely upon any particular sequencing of courses in the MBA program. By the same reasoning, no assumption of prior exposure to data mining can be expected. Ideally data mining would be incorporated into the quantitative methods course that most MBA programs include as part of their cores and/or prerequisites. It would also be helpful for such a course to be required at the beginning of a student's curriculum. However, these aspects may not be feasible and are not essential. Students would nevertheless be expected to have developed competency and familiarity with various computer applications, such as spreadsheets and possibly data management tools and concepts. In addition, a first course in Applied Statistics should also have been completed by students prior to their taking courses that are a part of the DME.

Finally, a number of options exist for data mining software, and several of those options (e.g., Orange and RapidMiner) are available as free / open source. In addition, powerful commercial software is available from Microsoft, SAS, SPSS, STATISTICA, and Oracle. Microsoft's SQL Server 2008 in particular provides, through its Analysis Services, a very attractive alternative that uses the familiar Excel interface. Because of the broad familiarity most MBA students are likely to have with Excel and the zero cost of incorporating the SQL Server Analysis Services add-in, the DME proposal recommends this approach, which is discussed further below.

LOW-COST ACADEMIC PLATFORM FOR DATA MINING

Three components comprise a low-cost, highly effective platform for supporting the proposed DME: Microsoft SQL Server 2008, Microsoft Office (in particular, Excel, and optionally Visio), and the SQL Server data mining add-ins. In addition, someone with database administration experience will be needed for installing/configuring/maintaining the databases and developing the data warehouses and data marts to be mined for the courses involved.

SQL Server, an enterprise level database management system, is available for around $1000. This software would need to be installed on a networked computer that can act as a repository for the data that will be accessed by students using Excel for data mining. Nearly all MBA students have both access to and somewhat extensive experience working with spreadsheet software, especially Excel. SQL Server's data mining add-ins are a free download and consist of: the Table Analysis Tools for Excel, the Data Mining Client for Excel, and the Data Mining Templates for Visio, a highly capable and widely used charting application. Visio can be freely available to faculty and students in departments that are members of the MSDN Academic

Alliance, for which a subscription costs qualifying departments around $500 annually.

Using Microsoft SQL Server Analysis Services

The Analysis Services feature is best summarized by McLennan et al (2009): "Microsoft SQL Server 2008 is the third version of SQL Server that ships with included data mining technology. Since it was introduced in SQL Server 2000, data mining has become a key feature of the larger product. Data mining has grown from an isolated part of SQL Server Analysis Services with two algorithms, to an intrinsic part of the SQL Server Business Intelligence (BI) platform that is fully integrated with OLAP, Integration Services, and Reporting Services. Other Microsoft applications (such as Microsoft Dynamix CRM and Microsoft Performance Point Server) seamlessly integrate SQL Server Data Mining to accentuate their functionality with predictive power. SQL Server Data Mining has become the most widely deployed data mining server in the industry, with many third-party software and consulting companies building on, specializing, and extending the platform. Enterprise, small and medium business, and even academic and scientific users have all adopted or switched to SQL Server Data Mining because of its scalability, availability, extensive functionality, and ease of use."

Data mining add-ins for Excel

In its white paper, "Predictive Analysis with SQL Server 2008", Microsoft (2008) writes that SQL Server 2008 Analysis Services provides a complete data mining platform that organizations can use to infuse insight and predictive capabilities into everyday business decisions. Pervasive delivery through the data mining add-ins therefore delivers data mining capabilities such as predictive analysis capabilities with intuitive tools and clear results that are available throughout the enterprise at the desktop. Traditionally, these types of analysis are available only to managers having extensive expertise in statistics. Microsoft SQL Server 2008 Data Mining Add-Ins for Office 2007 (compatible also with Office 2010) therefore enables a wider group of users to exploit highly sophisticated data mining technology within a familiar spreadsheet environment. Consequently, it is for the following reasons that this platform is attractive for use in the proposed DME:

- The comprehensive set of tools allows users to make a wide range of decisions in a few simple steps by providing prompt analyses and actionable recommendations.
- The Table Analysis Tools feature hides complex data mining algorithms behind intuitive tasks so as to allow users a consistent experience of exploring data and discovering patterns.
- The Data Mining Client feature for Excel allows a wide range of users to work through the entire data mining life cycle using spreadsheet and other data, which may or may not be accessible through SQL Server 2008 Analysis Services.
- The Data Mining Templates feature for Visio allows users to render annotatable graphical visualizations of the data mining models so that findings may be presented in a professional and understandable manner.

In this manner, the data mining technology of SQL Server 2008 through its integration with

Office 2010 delivers a powerful and familiar tool for a wide range of enterprise data. In other words, this add-in enables users of Excel to gain access to features provided by predictive and other types of analysis:

- Such solutions are most effective when they retrieve and incorporate data from all business divisions to support decisions at the enterprise level.
- These solutions bear practical meaning as they require MBA students to incorporate qualitative and quantitative approaches in their business decision making.
- Self-service data mining for supporting prompt decision making is provided for managers/students who do not have much training in information technology or statistics.

DATA MINING ACROSS THE MBA CURRICULUM

An MBA curriculum is rich with opportunities to exploit data mining capabilities and thus provide instruction in real-world decision making. Table 1 below summarizes the most common data mining tasks, along with the algorithms used to accomplish those tasks. Clearly these tasks reach well beyond the material addressed in most MBA classes in Marketing, Management, and Finance, but the table can serve well as a guide to identifying opportunities to exploit data mining, both in coursework and on the job.

Table 1: Data mining tasks (Microsoft, 2007)

Task	Description	Algorithms
Market Basket Analysis	Discover items sold together to create recommendations on-the-fly and to determine how product placement can directly contribute to your bottom line.	Association Decision Trees
Churn Analysis	Anticipate customers who may be considering canceling their service and identify the benefits that will keep them from leaving.	Decision Trees Linear Regression Logistic Regression
Market Analysis	Define market segments by automatically grouping similar customers together. Use these segments to seek profitable customers.	Clustering Sequence Clustering
Forecasting	Predict sales and inventory amounts and learn how they are interrelated to foresee bottlenecks and improve performance.	Decision Trees Time Series
Data Exploration	Analyze profitability across customers, or compare customers that prefer different brands of the same product to discover new opportunities.	Neural Network
Unsupervised Learning	Identify previously unknown relationships between various elements of your business to inform your decisions.	Neural Network
Website Analysis	Understand how people use your Web site and group similar usage patterns to offer a better experience.	Sequence Clustering
Campaign Analysis	Spend marketing funds more effectively by targeting the customers most likely to respond to a promotion.	Decision Trees Naïve Bayes Clustering
Information Quality	Identify and handle anomalies during data entry or data loading	Linear Regression

	to improve the quality of information.	Logistic Regression
Text Analysis	Analyze feedback to find common themes and trends that concern your customers or employees, informing decisions with unstructured input.	Text Mining

Certainly it is not the intention of this DME proposal for the focus of MBA coursework to be data mining. Students do indeed need to learn the general concepts of primary subjects such as Marketing, Management, and Finance, and there is not sufficient time during a semester to delve into data mining. However, incorporating data mining into each of these classes enables students to do more than learn *about* various concepts and techniques associated with the primary subjects. Students can also gain actual *experience* with those concepts and techniques.

Because there can be no assumption of prior exposure to data mining, each course would nevertheless need to devote a class meeting to the fundamentals of data mining and demonstrating the basics. In addition, the data mining case for that course would be introduced. Students would be expected to be prepared through outside readings and spreadsheet exercises and then would be referred to additional readings and exercises following that class meeting. Beyond that, further exercises associated with the case, along with related readings, would be assigned as topics from the primary subject dictated. Some of those topics, as well as corresponding analyses, are addressed below.

Finance

In finance classes, students are taught the importance of, as well as a variety of methods for, evaluating capital assets and predicting revenues and stock prices using both cross-sectional and time series data. In addition, determining and understanding predictors of creditworthiness, as well as a variety of classification issues must be addressed by financial managers. However, conventional methods have significant limitations when complex relationships, such as nonlinear and/or seasonal ones, are involved. Such problems, which are more common than not, are much harder to address, given that students typically are not equipped with effective quantitative tools that can take advantage of the massive amounts of available data.

Very commonly conventional methods involving the use of spreadsheet software, such as Excel, serve as the mainstay of the student financial analyst. Excel provides a large number of both scalar and array functions that can be used for basic forecasting purposes. For linear forecasting, these include LINEST, TREND, FORECAST, SLOPE, and STEYX, while, for exponential forecasts, LOGEST and GROWTH are some of the functions that are often used. Although many of the available functions can deal with multiple independent variables, there are many situations that are much too complex. In these cases, analysts must generally accept predictions provided by outside sources or rely on subjective interpretations of data, charts, and the other unsophisticated sources. This can be very problematic, as predictions for sales and other revenues drive financial planning decisions, such as how much debt to take on or how much stock to issue in order to achieve pro-forma balances for upcoming time periods.

This is where data mining, in particular, using Excel enhanced by the SQL Server data mining add-ins, can be quit helpful. For example, there is a Forecast function that employs two separate time series algorithms – ARTXP (autoregressive tree with cross prediction) and ARIMA (autoregressive integrated moving average) – for making predictions. The former was introduced in SQL Server 2005 and is optimized for predicting the next likely value in a time series. The latter was added to SQL Server 2008 to improve accuracy of long-term forecasts as it does well in predictions involving data with periodic regularity (Kovalerchuk & Vityaev, 2000). While ARIMA, which requires user involvement at several points, is available through most statistical software packages (e.g., SAS and SPSS), the ARTXP algorithm was developed by Microsoft specifically as a means of implementing decision tree methodology within Analysis Services. By default, the Forecast function uses both methods, ARTXP and ARIMA, and blends the results to improve prediction accuracy. As might be expected, the algorithm parameters can be set to use only ARTXP or ARIMA, as well as to control how the results of the algorithms are combined. Cross-prediction, however, is available only when either ARTXP by itself or some blend of algorithms is selected for the model.

Marketing

Many areas within marketing have benefited from data mining. For example, customer relationship management, market basket analysis, and sales prediction are a few of the successful applications of data mining. Much data exists and continues to increase, especially with the ever-growing application of automatic identification and data capture technologies such as radio frequency (RFID) and light frequency (bar code) identification. In addition, the growth of online sales provides considerably more information about customers, potential customers, and their buying habits.

Whereas marketers have previously needed to rely on survey data, which can be biased and is also quite limited and expensive, we are now able to generate highly reliable analyses because of the large data sets that exist. Furthermore, data aggregators such as Acxiom have been able to compile considerable amounts of data, both focused and general, for purchased by interested parties (Kroenke, 2012). This has enabled sales campaigns to be much more precise and efficient, leading to substantially improved effectiveness of organizations' marketing efforts.

It is nevertheless not uncommon for data mining not even to be mentioned in the MBA course in Marketing, let alone actually employed. While there is much discussion of cross-selling, as well as building, improving, and maintaining customer loyalty, for example, students gain little if any experience with practical applications of these concepts. Classes such as Marketing Research do teach traditional methods of data analysis (e.g., regression analysis, ANOVA, and contingency analysis). This is good and provides an important foundation for data mining applications in marketing, but it stops short of providing students with modern approaches to dealing with modern data.

As an example of how easily a typical analysis might be done using Excel with the SQL Server data mining add-in, consider a large data mart consisting of demographic, lifestyle, and purchase

behaviors of people in a given region. Such data might be publicly accessible, acquired through surveys and/or loyalty programs, and/or purchased from a data aggregator. (The "and/or" indicates the possibility of combining data from a variety of sources.) Using the Key Influencer tool, a very simple and quick process, can develop a report on what factors have greater and lesser influence on, for example, the decision by a person to buy a bicycle.

Incorporating data mining into the Marketing course should therefore provide both an increased depth of understanding and a practical experience that will improve the attractiveness of MBA graduates in the job market. Beyond the standard tools of statistical analysis, students can now have a working knowledge of decision trees for predicting customer behavior, market segmentation through cluster analysis, and sales forecasting via artificial neural networks, to mention a few of the major applications. In addition, students can develop powerful descriptive analyses relatively easily by developing OLAP cubes that organize demographic, market, and other data.

Management

While data mining has been (and continues to be) successfully applied in practice for marketing and finance applications, much potential is yet to be realized for data mining in management. The core course in Management at the MBA level typically addresses individual (essentially human resource) issues, group processes, and organizational processes. Although the last item tends to be strategic in nature and may not lend itself to data mining, both individual and group aspects, to a greater or lesser extent, do stand to benefit from decision making based upon data mining. Consequently, training MBA students to recognize and exploit opportunities for employing data mining in these areas provides the potential for important contributions our students might make once they are on the job.

In the area of human resources management in particular, data mining can provide an excellent tool for helping researchers understand important issues. Both internal data and external data can be used to identify characteristics of successful employees, to determine which employees are likely to leave an organization, predict, as well as to identify determinants of, job satisfaction, turnover, and related concerns. Classifying workers so as to determine which types of training might work most efficiently and effectively could be another application of data mining, based on research that could identify worker types in much the same way market segmentation is done. Another example might be using data available in large repositories to model turnover as a function of engagement, which in turn might be modeled as a function of autonomy, self-efficacy, supervisor support, coworker characteristics, idealism, pride, etc. Finally, recruiting efforts could be optimized by the appropriate focus upon universities, degrees, experience, etc.

For the most part, assigning employees to teams has been addressed via subjective means. However, there are a number of critical success factors that might be addressed quantitatively. For example, much work in modern organizations involves teams of employees, often geographically, if not temporally, dispersed. Matching employee qualifications, personalities, and other characteristics might perhaps be done through some sort of classification and/or

prediction algorithm. Teams could then be determined based upon classification scores and success (i.e., cohesiveness, compatibility, etc.) predicted without the problems associated with actually deploying the teams. Very possibly mistakes could be avoided by experimenting with data mining models rather than with the actual organization. While no research to this has been found in the literature, MBA students could be assigned projects of this nature. Whether this would be an appropriate application of data mining or not, students would be introduced to the idea of using data mining (as well as other tools) to address previously unexplored situations.

CONCLUDING REMARKS AND FUTURE DIRECTIONS

A number of recommendations have been provided here for *how* data mining might be incorporated into MBA curricula. Reasons have also been suggested for *why* the development of an integrative data mining experience would be beneficial. In general, the *value-added* can be substantial, both for MBA graduates and for the organizations that hire them. The addition, strengthening, and reinforcement of *data management skills* resulting from the DME will prepare graduates to become much better decision makers in today's world of ubiquitous data and powerful technology capable of handling those data. Furthermore, graduates will have been involved in an experience that *transcends boundaries* among functional areas and will thus have a better understanding of the needs for those functional areas to interact. Organizations hiring these graduates will consequently be better able to capture and exploit data that will enable such organizations to operate more effectively.

To summarize this DME proposal, note that there are three fundamental components involved. First of these is a comprehensive data set comprised of databases from various functional areas within an enterprise. A second component is a multifaceted case study based upon those data and capable of being addressed from several perspectives, each corresponding to a functional area. Finally, a data mining tool that is powerful, yet easy to use is needed. It may initially be surprising that Excel is recommended for this third component. Using SQL Server for maintaining a central online repository for the data provides much of the power needed for data mining. The remainder of that power is achieved through the use of the data mining add-ins for Excel. Ultimately this platform is an effective and low cost client-server based solution, one that takes advantage of the widespread familiarity among business people with Excel.

What remains now for this proposal to be completed is a proof-of-concept, which is currently being developed. First, course content for MBA programs at various universities is being examined to identify opportunities for data mining in those curricula. Based upon the findings, as well as consultation with various faculty involved across the MBA curriculum, a draft of the comprehensive case must be developed. In the meantime, work is being done to identify and/or develop a single data set that will support various aspects of data mining in conjunction with multiple functional areas.

REFERENCES

Available upon request from John Seydel