A Framework For Incorporating Data Mining Into An Accounting Curriculum

Thomas G. Calderon, (Email: tcalderon@uakron.edu), University of Akron
John J. Cheh, (Email: cheh@uakron.edu), University of Akron

Abstract

As usage of data warehouses and data mining in corporate America increases, it is expected that accounting graduates will be called upon to contribute to data mining and knowledge discovery processes. Currently, the highly structured exercises and cases found in many accounting textbooks do not draw students' attention to the complex task of mining large databases for information that might be used in business decision-making. This paper builds a typology of data mining tasks that considers the complexity of a database as well as the complexity of data mining tools, and proposes a multi-tier approach for teaching data mining concepts and techniques to accounting students. Drawing from that typology, the authors propose that accounting students be first introduced to simple databases that use flat files and simple data mining tools. As they advance through the accounting curriculum and learn about enterprise scale database management systems such as Oracle, accounting instructors can progressively introduce students to more complex data mining tasks. Eventually, these students can progress to complex data mining tasks that require denormalization of data tables and utilization of comprehensive, fully functional data mining tools. The paper suggests that instructors may use pivot table reports (e.g., Excel's PivotTable reports) to provide accounting students with their first introduction to data mining. Product profitability analysis is used as an illustration. Although pivot tables would be at the low end in a typology of data mining tools, it is easy to prepare and use a pivot table to explore and summarize previously hidden patterns and relationships in a flat file. By using pivot tables, accounting students can summarize and analyze large data sets in a very familiar environment (e.g., Excel) and generate reports that allow them to apply filters and conveniently drill down to explore detailed patterns and relationships.

Introduction

New developments in information technology and the emergence of a new, knowledge-driven economy bring many challenges and opportunities to accounting practitioners and educators, particularly in relation to the use of large volumes of data to create business intelligence and foster competitive advantage. Normal business activities result in the creation of extensive databases. Many entities devote large amounts of resources to the design, development, and maintenance of large databases that cover several years of transactions, events,

Readers with comments or questions are encouraged to contact the authors via email.

and client or customer attributes. Though potentially rich in information, these databases are frequently underutilized or left untouched. Among the factors that contribute to underutilization of data are data complexity, difficulty in accessing data, misunderstanding of the value of data by managers and systems administrators, the necessity to reformat and preprocess data, the need to apply sophisticated statistical techniques to identify patterns and relationships, and training deficiencies. On the other hand, many business enterprises, including Johnson and Johnson (Business Week, 1996), General Electric Capital (Attar Software Limited, 1998), Fingerhut (Pearson, 1999), Procter and Gamble (Anthes, 1999), and Harrah's Casino (Binkley, 2000), have tapped into their corporate databases and transformed their raw data into knowledge that provide them with significant business intelligence and competitive advantage. This ability to create business intelligence from large volumes of data, known as data mining, is becoming a necessity in the new, knowledge-driven economy.

Data mining offers new business concepts and tools for helping accounting students effectively analyze data and uncover hidden patterns within the mountains of data that exist in most corporate databases. In this paper, we review fundamental data mining techniques that may be used to make many areas of accounting more dynamic and interesting, including cost analysis and product profitability analysis. The paper builds a typology of data mining tasks that considers the complexity of a database as well as the complexity of data mining tools, and proposes a multi-tier approach for teaching data mining concepts and techniques to accounting students. A simple example is employed to demonstrate how data mining concepts may be introduced to students in an accounting curriculum.

Background

As the need to create business intelligence from large databases expands, it is expected that management accountants will participate in that process. Accountants have traditionally been an important resource for processing data and in providing information to corporate managers for routine operational decisions as well as strategic initiatives. In order to retain this role and continue to offer value to their employers or clients, accountants will have to be trained to participate and contribute to knowledge discovery activities. However, the model employed in training accountants emphasizes decision-making in the context of a single product firm or in the context of a very structured, highly artificial decision situation. While there are opportunities for exploring data from multiple perspectives even in enterprises that make and sell a single product, most accounting textbooks tend to provide all the information needed for decision-making, without an opportunity for students to process data in a significant way. In the few cases where multi-product entities are addressed, most of the data needed for decision-making are provided in a highly processed form or sample sizes are unrealistically small. Thus, accounting students are not often exposed to the complexities and opportunities for mining corporate databases and extracting data and information that might be needed for decision-making. Yet the ability to mine complex data and extract coherent patterns and relationships is becoming critical in the new knowledge-driven economy.

When the main core of management accounting was written in 1960s, the market place was very different from today. Time compression in decision-making and quick response times were not as integrated into the business culture as they are in the New Economy. Furthermore, many organizations at that time had a command-and-control type organizational structure, so that companies were able to compete even with slow response times and a top-down oriented business environment. Since then, there have been many changes in corporate America. One major change has been an unprecedented increase in the volume of business data collected and processed everyday. The Internet, World

Wide Web, advances in information technology, relatively low cost of hardware, and new modes of doing business have contributed significantly to the increase in business data. Moreover, both internal and external users and consumers of information have increasingly demanded quick response times. Recently, there has been a shift toward decentralization of information in organizations with a corresponding increase in the end-user orientation of information. This shift has contributed to the democratization of corporate databases and has enhanced the ability of business enterprises to respond to customer demands. At this juncture, incorporating data mining into the accounting curriculum will help accounting students become more sophisticated, with the ability to address both conceptual and implementation issues related to data mining and the decentralization of corporate data.

Data Mining in Accounting

The concept of mining data has been around for a long time, even though it did not become practical until data infrastructure and appropriate information technologies were in place. Some activities in accounting may be considered as a step toward data mining to the extent that corporate accountants try to extract valuable information from available financial data by querying the database for the information desired. However, accountants and other users of financial data tend to pigeonhole their queries. Because their queries are so narrowly defined, they omit many possibilities and potential answers to the business situation they are attempting to address. A pigeonhole approach to knowledge discovery is logical and sensible when the business model is well defined and processes are well known. This is not the case in the New Economy.

The traditional approach to knowledge discovery made more sense in the Old Economy, because the availability of the data was confined to a small subset of possible databases and patterns were relatively easy to identify and predict. However, the emergence of the New Economy added new dimensions to the traditional approach to doing business. The new economy is more customer-focused and more market driven; and there tends to be more of what may be referred to as *managed variability*. With *managed variability*, companies use mass customization in the way products and services are produced, sold, and marketed. Another major characteristic of the New Economy is the availability of large-scale databases and highly sophisticated information demands to support the *managed variability* strategy employed by corporations. Currently, accountants are not trained to tap into large databases, such as data marts and data warehouses; however, these databases provide significant opportunities for accountants to find meaningful patterns and relationships that may be leveraged to enhance management and organizational effectiveness.

The traditional performance metrics (e.g., unit production cost, cost variances, gross profit, and operating expenses) that accountants frequently work with are important, because they provide valuable information. However, accountants and managers often work with simple averages and aggregated data that tend to mask the detailed patterns and relationships needed to support a *managed variability* strategy. The information culled from the detailed data that reside in large contemporary databases provides new insight and opportunities to create incremental value that traditional performance metrics often miss. To fill this void, data mining brings new business concepts that have not been well practiced in the past because of the limitation of the traditional information technology infrastructure.

Data mining adds additional dimensions to managers and accountants because it goes beyond traditional statistical analysis. Statistics are used to provide reliable evidence for either verifying or rejecting a particular hypothesis. In most statistical analysis, the researchers form a hypothesis, using various methodologies such as model building, empiricism, or other theoretical processes. Statistics are then used to either support such claims or reject them. However, in data mining, the degree of importance of hypothesis building and testing varies. The emphasis is on finding a business solution, and not necessarily testing a hypothesis.

The importance of hypothesis building depends on the type of data mining. Dyche (2000) offers insight, summarized in Table 1, into the various types of data mining activities and the degree of hypothesis building that might be employed in those activities. Dyche observes that only the most complex decision support activities do not require a strong degree of hypothesis building and concludes that activities that require a high degree of hypothesis building, such as multidimensional analysis and standard queries, are not true data mining activities.

Although Dyche (2000) views only high-end activities such as modeling/segmentation and knowledge discovery as data mining, many software vendors consider all four types of activities in Table 1 as data mining activities. In this paper, we define data mining as intelligent query activities that help business managers find hidden patterns and relationships in corporate data by employing four commonly used processes (Westphal and Blaxton, 1998)--data modeling for data mining, defining business problems, data preparation, and applying data mining methods. A description of each process follows.

Data modeling

Data modeling for data mining is one of the most critical processes in a data mining project as it determines the variables that may be accessible and the types of patterns and relationships that may be discovered. In the data modeling stage, the systems architects who build data marts or data warehouses presume that the data will be used for data mining in the future. Thus, the architecture of data marts or data warehouses is designed to support data mining processes. For instance, data mining requires very detailed data structure (Berry and Linoff, 1997; Dyche, 2000; Westphal and Blaxton, 1998). If the data are descriptive, the analyses will be confined to the use of descriptive models, whereas transactional data may be represented in either descriptive or transactional models (Westphal and Blaxton, 1998). In general, the data structure used in a data mart or warehouse can affect the effectiveness of data mining processes and is, therefore, a critical issue. This issue must be resolved within the context of the goal of the data mining project and the nature of the business environment in which the project takes place. The goal of the data mining project and the environment in which the project takes place will determine the object classes to be modeled and the attributes that will be used to describe or categorize their behavior.

Problem definition

Data mining may be used in finding solutions to many different problems. Among the problems that have been addressed through data mining are new purchasing trends, the mix of products consumers purchase, direct marketing and marketing with coupons, fraud and unauthorized expenditures, and crime detection and prevention. Accountants may use data mining techniques to discover new patterns and trends in, for example, cost variances, product profitability analysis, cost driver analysis,

Table 1
Comparison of Data Mining Activities with Hypothesis Building

Type of Data Mining Activities	Degree of Hypothesis Building
Knowledge Discovery	No Hypothesis
Modeling/Segmentation	Mild Hypothesis
Multidimensional	Moderate Hypothesis
Standard Queries	Strong Hypothesis

quality costs, cost of various activities in the value chain, and life cycle costs. Because data mining involves using and processing data from large databases, it is important to identify the business problem to be addressed. Otherwise, resources may be consumed on activities that will produce results that are not used, thereby undermining the value of data mining activities.

In knowledge discovery, where sophisticated modeling techniques such as Kohonen's neural network architecture are used (Westphal and Blaxton, 1998), an explicit hypothesis or rigid problem definition may not be necessary. Nevertheless, defining and understanding the scope of the business problems to be solved will help end-users narrow down their search domain to a considerable degree and make data mining more efficient.

Data preparation

Once the business problems are defined, data need to be accessed, extracted, converted, integrated, and preprocessed into the proper format for data mining processes. In the past, data were normalized before they were stored for reduction of redundancies. In that way, corporations saved resources. However, in the process of normalizing data, managers may lose sight of the transactional details that could be mined. To avoid missing transactional details, depending on the software used, many corporations must denormalize data from their relational corporate databases and pipe the data into a flat file in preparation for data mining. Some applications, such as IBM's Intelligent Miner, may be used to mine both normalized tables in a relational databases and flat files, thereby reducing the number preprocessing activities in a data mining task (Biggs, 1999). Nonetheless, preparing data for mining, or preprocessing of data, is a crucial step in creating business intelligence from corporate data.

Data mining methods

Data preparation allows business managers to evaluate various data mining models identified as possible candidates for solving a particular problem. There are many competing techniques that business managers can use for data mining. They include artificial neural networks, clustering, decision trees, genetic algorithms, link analysis, and visualization (Berry and Linoff, 1997). In general, these techniques are used to find hidden patterns in data. However, many of these techniques may need a solid understanding in artificial intelligence and statistics.

A Typology of Data Mining

Because most accounting students may not have a strong background in the artificial intelligence and statistics, introducing data mining to those students may be most effective if it is done by using an incremental, multi-tier approach. Initially, for sophomore and junior students, instructors may introduce simple databases (e.g., flat files in Excel or Access) and low-end, relatively simple data mining tools. As they progress through the accounting curriculum and develop skills in using complex database management systems (DBMS) such as Oracle and DB2, accounting instructors can progressively introduce students to more complex data mining tasks. Eventually, these students can progress to complex data mining tasks that require denormalization of data tables and utilization of comprehensive, fully functional data mining tools. In all cases, albeit with varying degrees of rigor and complexity, students would be provided with an opportunity to address issues related to the four primary tasks in data mining--- data modeling for data mining, defining business problems, data preparation, and applying data mining methods. Table 2 summarizes the typology of data mining, identifies the types of students for which each level of data mining might be appropriate, and presents examples of data mining tools and DBMS.

Table 2
Incremental, Multi-tier Approach to Teaching Data Mining

Degree of sophistication of DBMS	Degree of Sophistication of Data Mining Tools: Low End Data Mining Tools	Degree of Sophistication of Data Mining Tools: High End Data Mining Tools
Simple DBMS	Simple data mining tasks appropriate for undergraduate sophomores and juniors. **Example data mining tools** Pivot tables, Graf-FX **Example databases** Flat files in MS Excel or Access	Moderately complex data mining tasks appropriate for undergraduate seniors. **Example data mining tools** Clementine, Enterprise Miner, CART, and Oracle Darwin **Example databases** Flat files in MS Excel or Access
Complex DBMS	Fairly complex data mining task appropriate for undergraduate seniors with good database background **Example data mining tools** Pivot tables, Graf-FX **Example databases** Normalized relational databases in enterprise scale DBMS such as Oracle or DB2	Complex data mining tasks appropriate for technically oriented undergraduate seniors and graduate students. **Example data mining tools** Clementine, Enterprise Miner, CART, and Oracle Darwin **Example databases** Normalized relational databases in enterprise scale DBMS such as Oracle or DB2

Data mining tools are classified in the typology presented in Table 1 as either low-end data mining tools or high-end data mining tools. Among the most significant distinguishing features between high-end and low-end tools are scalable data mining capability, functionalities of data mining, and automation of data mining processes.

Scalability

A significant difference between low-end and high-end data mining tools is the scalability. For instance, a low end data mining tool such as Graf-FX, a third party vendor product built on Microsoft Access, has a very limited capability in the number of variables and records that it can process (Westphal and Blaxton, 1998). On the other hand, high-end data mining tools such as Clementine, Enterprise Miner, and Oracle Darwin (Tamayo, Berlin, Dayanand, Drescher, Mani and Wang, 1997) provide ability to process a far larger database. For instance, Oracle Darwin uses parallel processing, and the size of the processable database is limited only by the capacity of the machine.

Functionalities

In essence, many data mining tools use any one or all of the following three data mining techniques: (1) cluster detection, (2) decision trees, and (3) neural networks (Berry and Linoff, 2000). However, some of data mining tool users may not have the background or training to use those sophisticated techniques. Thus, many vendors add an interface layer above the data mining engines to simplify the data mining tasks of end-users. In addition to cluster detection, decision trees, or neural networks, high-end data mining tools include visualization as well as non-visualization techniques as functionalities,

Exhibit 1
A Pivot Table Report

	M	N	O	P	Q	R
5	PIVOT TABLE					
6		Product				
7	Data	1	2	3	4	Grand Total
8	Average of Actual Price	$ 965.43	$ 1,863.64	$ 2,704.03	$ 3,642.32	$ 2,224.31
9	Average of Unit Var. Cost	472.99	1,070.70	1,765.47	2,400.22	1,377.21
10	Sum of Quantity	24,765	25,737	27,342	20,462	98,306
11	Sum of Net Revenues	$23,688,116	$46,855,784	$72,105,203	$72,079,872	$ 214,728,975
12	Sum of Total Var. Cost	11,724,890	27,544,720	48,337,050	49,030,800	136,637,460
13	Sum of Contribution Margin	$11,963,226	$19,311,064	$23,768,153	$23,049,072	$ 78,091,515

thereby providing users with a complete solution. Low-end data mining tools are generally weak in terms of functionality and do not provide as complete a solution as more sophisticated, high-end tools. For instance, some low-end data mining tools such as Microsoft Excel PivotTable Wizard do not offer comprehensive visualization possibilities, and users can execute only a simple data mining technique. On the other hand, Graf-FX offers visualization, but does not offer a comprehensive data mining tool set. Many data mining tools appear to be no more than a graphical tool, which is only a rudimentary feature in data mining. Those tools may be considered as low-end data mining tools.

Automation

Besides visualization, automation is usually the area that differentiates high-end data mining tools from low-end data mining tools or other statistical data analysis tools. High-end data mining tools often shield ordinary users from the complex data mining processes by providing an intuitive interface and hidden, back-end processing capabilities that make the task of data mining appear seamless. Despite the benefits that automation can offer to younger students, high-end data mining tools also offer many functionalities that sophomore or junior students may find overwhelming. Furthermore, many high-end tools are designed to handle complex data mining tasks that employ very large databases. Sophomore and junior students may lack the background and/or degree of business sophistication to analyze and fully appreciate business implications of the results provided by a complex data mining exercise. Thus, for the younger students, a simple approach to introducing data mining may be effective in the earlier years of their learning.

Microsoft Office products and similar productivity enhancement tools are now widely used in undergraduate accounting programs. It is presumed that most sophomore and junior students are quite comfortable with using desktop productivity tools such as Microsoft Excel and Access. Considering the students' comfort level with using such products, accounting instructors may introduce data mining to their sophomore or junior students, taking advantage of the students' familiarity with those products. For instance, as discussed in a subsequent section, instructors may use some of the features in Microsoft Excel to introduce data mining concepts. Alternatively, they may use a low-end data mining tool, such as Graf-FX, along with a simple database. Graf-FX is a relatively easy-to-use data mining tool that works with Microsoft Access (Westphal and Blaxton, 1998).

As students learn about and work with a sophisticated, enterprise scale DBMS such as Oracle 8i, they are better prepared for using more complex data mining tools such as Clementine, Enterprise Miner (Westphal and Blaxton, 1998), and Oracle Darwin (Oracle, 2000). Seniors and graduate students with knowledge and skills in working with enterprise scale DBMS may be assigned data mining tasks that emphasize all four processes in data mining---data modeling for data mining, defining business problems, data preparation, and applying data mining methods. Undergraduate seniors without knowledge and skills in working with enterprise scale DBMS would be assigned tasks that emphasize defining business problems and applying data mining methods, but not data modeling for data mining and data preparation. These students would work with denormalized tables in Access or Excel and focus more on finding answers to specific business questions and issues.

A Simple Approach to Introducing Data Mining

Microsoft Excel PivotTable reports provide a simple solution for mining corporate data without burdening students with a heavy dose of prerequisites in artificial intelligence, enterprise scale DBMS, and complex statistical methods. Although pivot tables would be at the low end in a typology of data mining tools, it is easy to prepare a pivot table and to use it to explore and summarize previously hidden patterns and relationships. Using a pivot table is a process similar to online analytical processing (OLAP). As indicated by Berry and Linoff (1997), OLAP and data mining are complimentary. OLAP tools are used for reporting on data, and data mining tools focus on finding pattern hidden in the data. In Microsoft Excel, there are some differences between PivotTable reports that are based on OLAP source data and reports based on non-OLAP source data. For instance, OLAP databases provide some of the analysis features on the server, including precalculating summary values and organizing the data hierarchically. They are also organized to facilitate the retrieval and analysis of large amounts of data. When working on OLAP source data, an OLAP server performs calculations to summarize the data and then Microsoft Excel displays summarized data in a PivotTable report. Only the summarized data are returned to Excel, on an as-needed basis. With non-OLAP external databases, all the individual source records are returned, and then Excel does the summarizing and filtering. Although OLAP databases can provide Excel with the ability to analyze very large volumes of external data, the simplest and most efficient approach to introducing data mining tools is to work directly with flat files within Excel. Accounting students and business managers are often familiar with flat file structures and can conveniently employ Excel's PivotTable reports to discover simple hidden patterns and relationships within such a file.

The remainder of this section explains what a pivot table report does, and demonstrates how Excel's PivotTable reports may be used to find simple hidden patterns in a flat data file.

What is a pivot table report?

A pivot table report is an interactive table that one can use to quickly summarize large amounts of data. The layout and content of the report can be changed interactively by using simple filters to provide different views of the data and to highlight specific details of interest to an analyst. Interactive filters are embedded in the pivot table report. Exhibit 1 shows an example of a simple pivot table report. The exhibit summarizes the annual sales, volume, cost, and contribution margin data for four different products sold by a company. The labels for the source data appears on the left, and the products--labeled 1, 2, 3, and 4--appear as column headings.

When to use a pivot table report?

Accounting students should use a pivot table report when they want to summarize and analyze a large data set that contains many rows and columns and it is not easy to identify patterns and relationships by merely reading the raw data. For example, the pivot table report in Exhibit 1 was created from a data set containing 904 rows and more than six columns. The data set contained 24 months of sales for four different products. Prices and variable expenses for each product varied from month to month and there were approximately 10 different sales transactions per product each month. Management needed to know the average selling prices, variable expenses, operating results for each product, the mix of products sold, and the effect that various combinations of product would have on the firm's average contribution. Once the basic data set was compiled, Exhibit 1 was created in under two minutes. Management used Table 1 to analyze the performance of each product, assess product mix issues, and review the contribution that each product and product combinations made to the company's unallocated common fixed costs and operating profits.

Because of its interactive nature, a pivot table report is ideal for drilling down or further aggregating summarized data. For example, a manager who uses the pivot table report in Exhibit 1 can obtain detailed insight into selling prices and performance data for specific months by adding a month column in the pivot table report and selecting the months he/she wishes to analyze.

How to create a pivot table report?

Pivot table reports can be created very easily by using Excel's PivotTable and PivotChart wizard. This wizard uses an intuitive interface to guide users through the process of building a pivot table report. Users access the Wizard by clicking the Data menu | PivotTable and PivotChart and then use the following steps to create a pivot table report from an Excel data set:

Exhibit 2
Pivot Table Wizard, Step 1

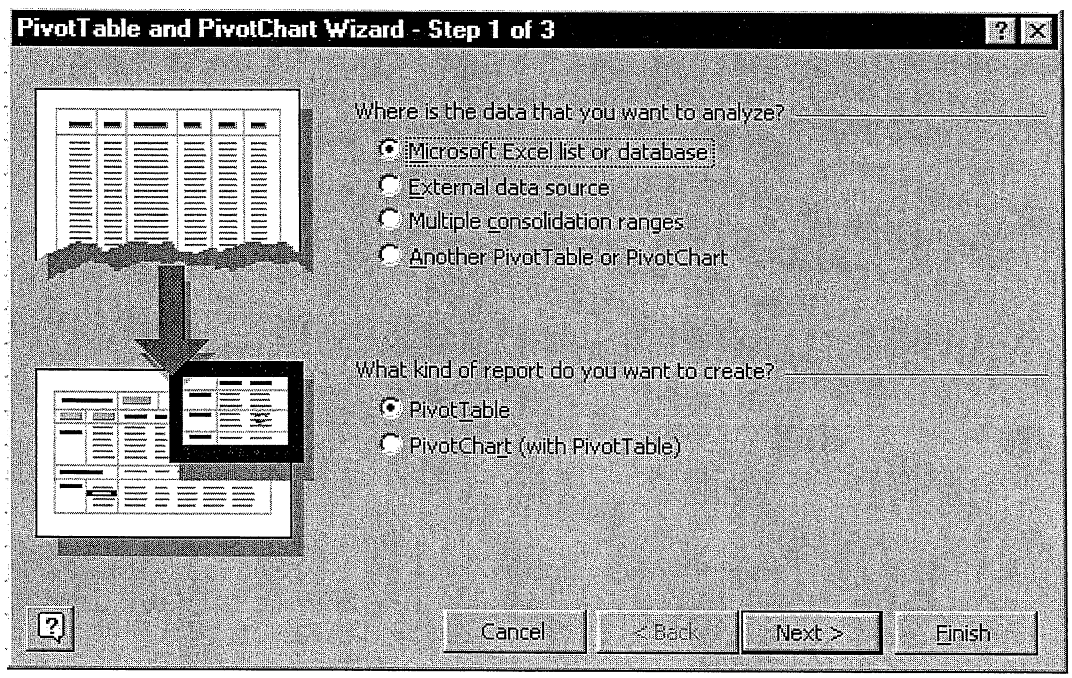


Exhibit 3
Pivot Table Wizard, Step 2

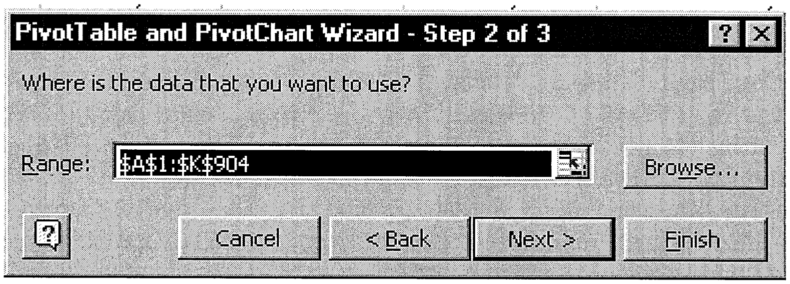
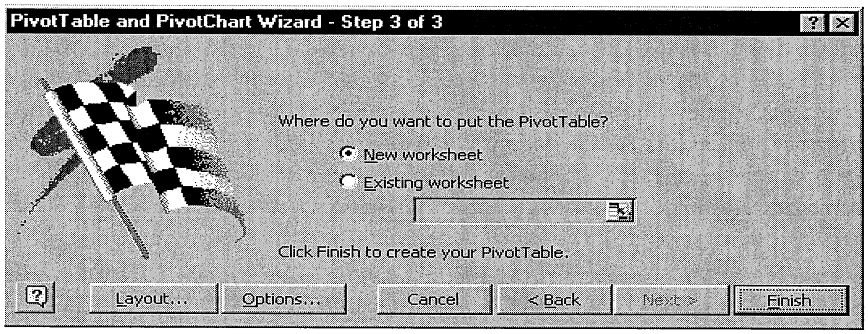


Exhibit 4
Pivot Table Wizard, Step 3



Open the workbook where the pivot table report will be created. Usually, this is the workbook that contains the data set the user wants to mine.

Click any cell in the data set. The data set must be comprised of a contiguous set of rows and columns.

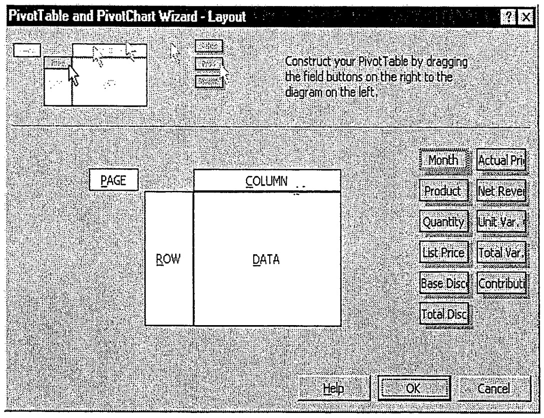
On the Data menu, click PivotTable and PivotChart report.

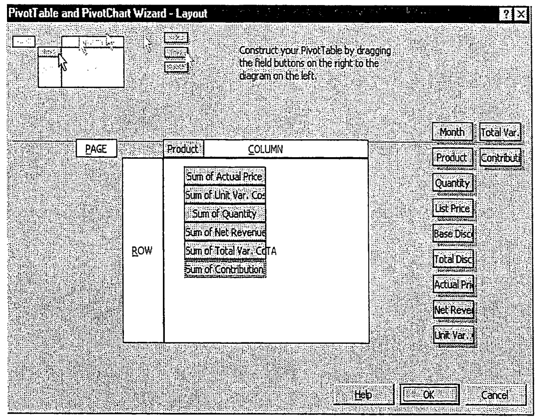
In step 1 of the PivotTable and PivotChart wizard, follow the instructions, and click PivotTable under "What kind of report do you want to create?" Users should also click "Microsoft Excel list or database" under "Where is the data that you want to analyze?" Exhibit 2 shows the dialogue box for step 1.

Click Next and follow the instructions in step 2 of the wizard. If you started off by clicking a cell in the data set containing the data from which you want to create the pivot table, Excel will automatically identify the dataset range, and the dialogue box for step 2 will appear as shown in Exhibit 3.

In step 3 of the wizard, determine whether the user needs to click Layout. Layout must be clicked for all, but the simplest tables. The Wizard also prompts users for the location of the pivot table. The user should click the choice (in the existing worksheet or a new worksheet) that as appropriate. Exhibit 4 shows the step 3 dialogue box for building the pivot table report shown in Exhibit 1. The report was placed in a separate worksheet and it was necessary to specify the layout of the report.

Exhibit 5
Pivot Table Report Layout, Incomplete and Complete Dialogue Boxes





Click Layout to specify the layout of the report. Exhibit 5 shows the layout used for the pivot table report in Exhibit 1. The dialogue box on the left in Exhibit 5 is blank. To lay out the report, drag and drop the buttons on the right to the appropriate location within the dialogue box. The dialogue box on the right shows the completed layout, after dragging and dropping the relevant buttons.

After completing the Layout, click OK to return to the dialogue box in Step 3. Click Finish.

The pivot table report is now complete. However, the user will need to refine the report by changing the number format and the field settings, as appropriate. The default for each row in the pivot table is a summation of each field listed in the Layout you design (Exhibit 5). For example, the report created from the Layout in Exhibit 5 is shown in Exhibit 6. Exhibit 6 includes fields such as Sum of Actual Price and Sum of Unit Variable Cost, which are not very meaningful. Average actual price and average unit variable costs are more meaningful. Thus, the field settings for actual price and unit variable cost had to be changed in order to end up with a report that looks like the report in Exhibit 1. The field setting for actual price was changed from "Sum of Actual Price" to "Average of Actual Price" as follows:

- Go to the cell labeled "Sum of Actual Price"
- Right click on your mouse.
- Click Field Settings to bring up the PivotTable Field dialogue box.
- Click Average
- Click OK

Sum of Unit Variable Cost was changed to Average of Unit Variable Cost in a similar manner.

Are there different types of pivot table reports?

A default pivot table report looks like the example in Exhibit 6. This report may also be displayed in different formats. For example, users may drag and drop the quantity field from its present location to the top of the report. Similarly, a manager may be interested in the average number of units of products sold per transaction, as well as the total number of items sold. This can be done by dragging the quantity field from the PivotTable toolbar and placing a second quantity field ("Sum of Quantity2") in the pivot table report. One would then change the second

Exhibit 6
Default Pivot Table Report

	Product				
Data	**1**	**2**	**3**	**4**	**Grand Total**
Sum of Actual Price	223013.28	452864.66	662488.14	670186.1	2008552.18
Sum of Unit Var. Cost	109260	260180	432540	441640	1243620
Sum of Quantity	24765	25737	27342	20462	98306
Sum of Net Revenues	23688116.1	46855783.89	72105202.99	72079871.57	214728974.6
Sum of Total Var. Cost	11724890	27544720	48337050	49030800	136637460
Sum of Contribution Margin	11963226.1	19311063.89	23768152.99	23049071.57	78091514.55

quantity field from "Sum of Quantity2" to "Average Quantity."

A user can also create a pivot chart report to view the data graphically, and can make a pivot table report available on the Web by using a PivotTable list on a Web page. When one publishes an Excel PivotTable report to a PivotTable list, users can view and interact with the data from within their Web browsers. For example, Microsoft uses pivot tables on its investor relations Web site to allow viewers to analyze the company's financial history and revenue composition (Microsoft, 2000).

Conclusion

As usage of data warehouses and data mining in corporate America increases, it is expected that accounting graduates will be called upon to contribute to data mining and knowledge discovery processes. Currently, the highly structured exercises and cases found in many accounting textbooks do not draw students' attention to the complex task of mining large databases for information that might be used in business decision-making. The paper builds a typology of data mining tasks that considers the complexity of a database as well as the complexity of data mining tools, and proposes that accounting instructors use a multi-tier approach for introducing their students to data mining. Drawing from typology of data mining tasks, the authors suggest that accounting students be first introduced to simple databases that use flat files and simple data mining tools. As they advance through the accounting curriculum and learn about enterprise scale database management systems such as Oracle and DB2, accounting instructors can progressively introduce students to more complex data mining tasks.

Pivot tables could be used as the first step in the proposed multi-tier approach to integrate data mining techniques into the accounting curriculum. The paper illustrates this first step by using a simple example that focuses on product profitability analysis. However, there are several other applications that are possible, including analysis of activities in activity-based costing systems, analysis of product cost in cost accumulation systems, expense report analysis, general ledger account analysis, revenue composition and product mix analysis, analysis of financial history, and customer relationship analysis. Pivot table reports may be published over a corporate intranet to provide many different individuals with access to multiple views of important data.

By gradually integrating data mining into the accounting curriculum, accounting educators will have an intuitive and powerful tool for exposing students to complex business scenarios and to the intricacies of knowledge discovery in the New Economy. Introduction to simple data mining tasks that leverage low-end tools and simple database structures build the foundation for analyzing and using complex databases for responding to issues and business problems that come from multiple databases. This responds to a vital recommendation of the Accounting Education Change Commission (Accounting Education Change Commission, 1990). Eventually, seniors and graduate students would be exposed to complex data mining tasks that emphasize key data mining processes such as data modeling, problem definition, data preparation, and application of comprehensive, fully functional data mining tools and techniques.

Suggestions for Future Research

There are a number of possibilities and opportunities for accounting researchers in the area of data mining. Future research may focus on both pedagogical issues related to data mining as well as empirical issues related to the implementation of data mining in business. Pedagogical research might focus on actual experiences and outcomes associated with the teaching of data mining in the accounting curriculum. Reports that describe teaching experiences and outcomes associated with the classroom use of high-end data mining tools and enterprise scale databases would be particularly interesting. With respect to empirical research issues, future researchers may make valuable contributions to the literature by providing evidence on how extensively data mining is used in the business community, and the extent of user satisfaction with the tools that are used for business intelligence. Though there is some documented evidence in this regard, as noted in Attar Software Limited (2000) and similar Web sites, most provide only anecdotal evidence. In general, with few exceptions (Burger, 1998; Ernst & Young, 1999), very little has been done to document the extent of use of data mining and other knowledge management tools in business. Data mining and knowledge management technology is relatively new. Therefore, a study of the factors that drive the diffusion of innovation in this area would make a valuable contribution to the literature. This could be accomplished through comprehensive case studies, field research, and/or surveys. In-depth case studies would facilitate an understanding of the factors that drive the diffusion of innovation, provide detailed insight into the data mining technologies in use by various companies, and document the benefits and costs of knowledge management. Though there is some anecdotal evidence about the benefits of data mining, there is little systematic research that coherently documents the costs and benefits of data mining activities. In this regard, readers can find valuable data mining articles written in a case study format on the Web (Pearson, 1998). However, these articles are often written with the business executive in mind, and generally skip detailed issues that might benefit students and future practitioners of data mining. In-depth cases and research on data mining will bring many benefits to the business community, accounting educators, and students. 🕮

References

1. Accounting Education Change Commission, Position Statement Number One. *Objectives of Education For Accountants*. September 1990. Available online: http://206.170.119.32/pubs/position/position1.htm (Accessed June 27, 2000).
2. Anthes, Gary, H., "P&G Uses Data Mining To Cut Animal Testing," Computer World, December 6, 1999. pp. 44-45.
3. Attar Software Limited, "Striking commercial gold with data mining at GE Capital," Available online: *http://www.attar.com/* (Accessed June 27, 2000).
4. Berry, Michael J. A. and Gordon Linoff, *Mastering Data Mining: The Art and Science of Customer Relationship Management*, John Wiley & Sons, Inc., New York, New York, 2000.
5. Berry, Michael J. A. and Gordon Linoff, *Data Mining Techniques: For Marketing, Sales, and Customer Support*, New York, New York: John Wiley & Sons, Inc., 1997.

6. Biggs, Maggie, "Intelligent Miner Reveals Corporate Gems," *Infoworld*, December 6, 1999. pp. 94, 98.
7. Binkley, Christina, "Casino Chain Mines Data on Gamblers, And Strikes Pay Dirt with Low-Rollers." *Wall Street Journal*, May 4, 2000. p. A1.
8. Burger, Dale, "Popularity of data mining a myth: Study," *Computing Canada*, Vol. 24, No. 31, pp. 1-2, 1998.
9. Business Week, "Special Report, Finance's Info-Tech Revolution," Business Week, October 28, 1996. pp. 129-150.
10. Chapman, Pete, Randy Kerber, Julian Clinton, Thomas Khabaza, Thomas Reinartz, and Rudiger Wirth,"The CRISP-DM Process Model," *CRISP-DM Consortium*, March 1999.
11. Dyche, Jill, *e-Data: Turning Data into Information with Data Warehousing*, Addision-Wesley, Reading, Massachusetts, 2000.
12. Ernst & Young, 18th Annual Survey of Retail Information Technology. Available online: http://www.ey.com/global/vault.nsf/US/1999_RETAIL_it_Study/$file/1999retail-it.pdf (Accessed July 5, 2000).
13. Microsoft Inc., Investor Relations Analysis Tools. Available online: http://www.microsoft.com/msft/tools.htm (Accessed June 28, 2000).
14. Oracle, Oracle Darwin. Available online: http://www.oracle.com/datawarehouse/products/datamining/index.html (Accessed June 27, 2000).
15. Pearson, David, "Marketing for Survival." *CIO Magazine*, April 15, 1998. http://cio.com/archive.
16. Tamayo, P, J. Berlin, N. Dayanand, G. Drescher, D. R. Mani and C. Wang, "Darwin: A Scalable Integrated System for Data Mining," Thinking Machines Corp. Bedford, Massachusetts, May 14, 1997. Available online: http://www.oracle.com/datawarehouse/products/datamining/documents/darwin_arch.pdf (Accessed June 27, 2000).
17. Westphal, Christopher, and Teresa Blaxton, *Data Mining Solutions*, John Wiley & Sons, Inc., New York, New York, 1998.