ISSN:0975-9646

Advanced Data Mining Tools in Web Based ERP, ASP Environment

Tiruveedula Gopi Krishna, Mohamed A.Abdelhadi, M.Madhusudana Subramanyam

Computer Science Department, Sirt University
Faculty of Arts and Science, Hoon,Al-Jufra,Libya

***Abstract*— The next few years will witness a true explosion of new information services as more and more files containing more and more data are opened up, which will repeatedly create new opportunities for business organizations real challenge. For companies that are just getting started with analytical enterprise resource planning and customer relationship management, integrating data mining can be a daunting task. In this paper we discussed how the proof of concept should create a solid business case for further integration of data mining into the company's marketing, sales, and customer-support operations using ERP. Furthermore, current trends indicate that more companies implementing Enterprise Resource Planning systems could further benefit in using data mining techniques.**

Keywords—ERP, hybrid systems, Extraction tool,ASP,CRM

I. INTRODUCTION

What kinds of business problems can data mining technology solve and what must users of these tools understand to apply these tools effectively? Questions such as: "Do sales of product X increase in the month of July?" or questions such as "Do sales of product X decrease when there is a promotion on Product Y?" are easily solved without the aid of data mining. Moreover existing tools such as OLAP and statistical techniques can be used in this situation to analyze those types of cases. In contrast, with data mining, a potential customer can ask questions such as: "What are the optimal factors in determining the sales of Product X?" However, not all data mining tools are necessarily optimal in solving certain kinds of problems. Some tools are better than others in specific types of problems. With traditional tools, an analyst trying to come up with an answer to questions such as those above will arduously try to generate a model through trial and error. He or she will first pose a set of assumptions with a hypothesis, then test it, and finally propose additional hypothesis and go through the repeated process of testing it and in this iterative way, build a model. With data mining, though certainly a set of assumptions along with a hypothesis needs to be postulated, along with testing it, and revising the procedures, the advantage in using a data mining tool is that a tool shifts much of the work of finding an appropriate model from the analyst to the computer. Thus, generating a model requires less manual effort alongside the added advantage that using a computer allows for a larger number of models to be evaluated increasing the odds of finding a proven and working model.

II. EXTRACTION TOOL

This stage is concerned with the extraction of patterns from the data. We discussed how these tools make it possible to specify source systems and mappings between different tables and files and their goal s where data comes from and what happens to it.

A. Filtering

Amount of information that is being generated each year is exceeding the ability of amounts of information, processing through these data are hampered by the volume of data. Algorithms are optimal to a certain level of processing. After a point, several every year, more and more data are accumulated into databases. Currently, the layers of processing need to be done to the data. This requires the use of filtering, where properly processing the data. Though databases can be constructed to hold plentiful the most types of information are extracted out to the user. In the case of web mining, web pages are filtered to a user to maximize the time a user spends in navigating through the site. The more filtering is done, the more beneficial it will be to the user in locating the information that is of interest to him or her. As such, three general techniques have been developed over the years to address the problem of information overflow. These techniques are information retrieval, information filtering, and collaborative filtering. Each of these technologies focuses on particular sets of tasks or problems. Information retrieval revolves around in fulfilling tasks such as fulfilling user interest queries. This essentially involves a query to a database for the extraction of information. Information filtering involves classifying streams of new content into categories such as finding any newly released soundtracks by Madonna or finding any newly released movies with the actor Sylvester Stallone. Lastly, collaborative filtering deals with focusing which items in a set should a user view based on the recommendations given by other users within the same community or group.

B. Information Retrieval

Information Retrieval consists of issuing a query against a database requesting specific information to the user. This may involve indexing a list of documents using a query to capture the sub list of matching documents pertaining to the query. In general, information retrieval techniques are less valuable in the actual recommendation process since they capture no new information about user preferences and add

no existing value to the existing information [1]. As such, for knowledge discovery processes, information retrieval is not considered part of data mining techniques as this technique merely involves issuing a query and does not involve digging through the layers of information trying to find hidden patterns in the data. One could indeed use a data mining tool to issue multiple queries against a database to find interesting results, but information retrieval does not yield in any new insight.

C. Information Filtering

Another use of a filtering protocol is that of information filtering. This type of filtering concerns itself with item content and the development of a personal user interest profile [1]. This differs from the focus of collaborative filtering where the goal is to identify users of similar tastes and the use of their opinions to predict the value of items for each user in the community. Computer users are connected via worldwide networks to an increasing number of data sources and other users. This interconnectivity provides users with previously unknown riches of available knowledge [2]. With today's increasing production of information coming from organizations, individuals, and society at large, even finding personal user interest profiles can be problematic. Imagine tailoring an Amazon.com site specific to every user. The amounts of information that would need to be maintained for each customer from the overwhelming amounts of data stored in a database would not be an optimal solution. An alternative approach can be using both collaborative and information filtering to address the needs of user and groups collectively. In that manner, content needs to be filtered with respects to specific user groups and not to individuals themselves. As computers, communication, and the Internet make it easier for anyone and everyone to speak to a large audience, well developed filtering techniques will need to be developed to meet reasonable performance standards [1].

D. Collaborative Filtering

Collaborative filtering or recommender systems help users make choices based on the opinions of other users [3]. Systems that use collaborative filters help people find articles they will like in the huge stream of available articles. Collaborative filtering is a technique mostly used in the context of web mining. Though web mining itself is a loosely defined term, the definition given to the kinds of applications web mining is used for is mostly in the context that has been discussed so far. Wherever an application can make use of web log files, ad files, previous customer purchases and other significant purchasing information to discover knowledge in the data, then the application is considered a web mining application. Another application such as querying a web server and obtaining information is a primitive protocol that does not reveal any hidden information or extracts out certain patterns in the data. Instead, it provides a fixed answer to a problem and does not try to do anything further. These applications can be called web mining, in the context of this paper, these types of applications are not classified under the discussion of knowledge discovery web applications. One of the best ways to find useful information is to find someone who has similar interests as another use and ask them for recommendations. Collaborative filtering is a way of mechanizing this form of information search [4]. Here are a few examples of collaborative filtering systems:

- Firefly (once known as Ringo) offers a "personal music recommendation agent". Thousands of people tell Firefly what music CDs they like; Firefly finds people with similar tastes and recommends music that other people in their similarity group like.
- Webhunter (once known as WebHound) does something similar for Web pages, finding pages for a current user from past users.
- Group Lens helps to filter Usenet News. After reading an item, each reader rates it according to how interesting it was. Subsequent readers see an "interest score" that is computed as a weighted average of previous readers. But the score is personalized: the weight a person's evaluation receives in this average depends on how often you have agreed with that person in the past [1].

  One of the chief problems facing providers of information services is how to filter information; collaborative filtering offers a solution to this problem and appears to be fertile area for research and development. Collaborative filtering is a unique approach to information filtering that does not rely on the content or shape of objects, as it is the case with content-based filtering. For example, content-based filtering would allow for recommendation based on a movie genre (fiction, horror, comedy, romance, etc.) and cast/credits (Mel Gibson, Arnold Schwarzenegger, Bruce Lee, etc.) [1]. On the other hand would tend to recommend a movie selection based on what group-minded individuals would tend to view. Collaborative filtering relies on meta-data (data containing information about other data) information pertaining to information about objects, such as CD's, movies, book, or web pages [5]. Data can either be collected automatically, by inferring from the user's interaction with the filtering system, or voluntarily collected where users supply the information. A specific type of collaborative filtering application is called active collaborative filtering [6]. This approach is based on encouraging people to share information with one another rather than collecting ratings and modeling user interests in order to compile recommendations as in traditional collaborative filtering techniques. Active collaborative filtering builds on the following premises:

- Every person says what items they like and dislike
- New items are recommended to a user based on the opinions of people with similar taste

Filtering can be applicable to music, movies, websites, news, TV programs, etc. On the other hand, passive collaborative filtering in the Usenet domain is based on providing users with data about the news readings of other Usenet users [5]. This approach is based on the observation that experienced users use not only the subject of discussions with a group but moreover use the occurrence of contributions by other users as indicators for potentially interesting discussions. Essentially then, collaborative

filtering uses a database about user preferences to predict additional topics or products a new user might like.

E. Hybrid Recommender Systems

Several systems have tried to combine the information and collaborative filtering techniques in an effort to overcome the limitations in each technique [1]. As noted above, Group Lens is a hybrid system that combines provided by users, data inferred from user behavior, e.g., the time spent reading articles as indicator for interest, and content based data extracted from the objects under investigation, such as the proportion of spelling errors and included text in documents [5]. The growth of Internet commerce has stimulated the use of collaborative filtering algorithms [7]. Such systems leverage knowledge about the known preferences of multiple users to recommend items of interest to other users. Microsoft Research Group has evaluated a new method called personality diagnosis [7]. Given a user's preferences for some items, they compute the probability that he or she is of the same "personality type" as other users, and, in turn, the probability that he or she will like new items. This new way of applying traditionally similarity weighting collaborative filtering approaches can be used in that all data are brought to bear on each prediction and new data can be added easily and incrementally. Another research application of collaborative filtering is currently being at developed at the MIT Collaborative Ontology Department. Ontologies are a means of categorizing objects, such as features of a product, a person's interests, or Web pages. This department has researched that in general ontologies developed by a single organization are necessarily sparse or coarse and can be slow to add important new categorizations. As such, this group is investigating innovative ways in which ontology can be developed by many distributed individuals and organizations. Their first prototype, tentatively called Mishmash, will evaluate one approach to how an easily extensible ontology can evolve collaboratively without a priori structure or centralized direction.

III. ERP AND DATA MINING SYSTEMS

Enterprise Resource Planning Systems (ERP) comprises a commercial software package that promises the seamless integration of all the information flowing through the company - financial, accounting, human resources, supply chain, and customer information [8]. ERP systems are a collection of software programs that tie together all of an enterprise's various functions--HR, finance, manufacturing, sales, etc. This software also provides for the analysis of this data to plan production, forecast sales, and analyze quality. Application Service Provider companies seek to offer competing services to smaller organizations helping these companies leverage their data management. These two types of companies, ERP & ASP companies, could gain even greater footing in the marketplace with the addition of providing data mining services. Providing a software tool that can integrate a companies' existing data across many departments with the addition of providing a data-mining tool that is maximized to work most efficiently with that software package could yield significant benefit to client organizations. Examples of ERP packages are HRMS, Financials, Manufacturing, Distribution, and Sales. Each ERP Package may offer different functionality for different industries. Current targeted industries for ERP installations are Communications, Federal Government, Financial Services, Healthcare, Higher Education, Manufacturing, Public Sector, Retail, Service Industries, Transportation, and Utilities [9]. In these industries, only large companies have been targeted due to the length and cost of an implementation. Recently, ERP software manufactures are offering reduced software versions, with fewer features, to medium-sized companies [8]. ERP companies are now facing the pressures of providing smaller scaled services to these smaller companies. Application Service Provider companies (ASP) are competing for this market space in offering systems that can integrate and consolidate companies' information systems. Few mid-sized businesses have the capital to invest in, manage, and upgrade advanced e-commerce and e-marketing technologies in-house. In order to remain competitive, many companies are now turning to application service providers, organizations that serve the function of internal IT departments by developing and managing the technologies necessary to deliver key services over the Internet. The integration of the Internet is creating new dimensions of opportunities that will quickly and drastically affect the way that society functions [10]. One of the most innovative developments is occurring in the ASP market as these service organizations are allowing smaller companies to seek the advantages of ERP systems without actually having to implement anything in-house. ASP companies are now able to offer data mining services to smaller players in any industry. This is the advantage realized by corporations using an ASP model as they can provide leading services to companies not having enough capital to budget expensive ERP systems. Because ERP and ASP vendors offer significant advantages to corporations incorporating such solutions, integrating a data-mining tool alongside these services can provide an optimum solution in understanding a company's data. Enterprise Resource Planning [ERP] Systems: ERP systems try to build a single software program that serves the needs of people in finance as well as it does the people in human resources up to the managers making decisions. Each of those departments typically has its own computer system, each optimized for the particular ways that the department does its work. But ERP combines them all together into a single, integrated software program that runs off a single database so that the various departments can more easily share information and communicate with each other [11]. Enterprise Resource Planning software systems provide comprehensive management of financial, manufacturing, sales, distribution and human resources across the enterprise. The ability of ERP systems to support data drill down and to eliminate the need to reconcile across functions is designed to enable organizations to compete on the performances of the entire supply chain [12]. To utilize these capabilities managers have to learn how to manage processes in the ERP environment. That integrated approach can have a tremendous payback if companies install the software correctly. For example, in a common scenario experienced by frustrated customers consists of an order bouncing around different departments before an order has been finalized. When a customer places an order, that order begins a mostly paper-based journey from in-basket to in-basket around the company, often being keyed and re-keyed

into different departments' computer systems along the way. Meanwhile, no one in the company may not truly know what the status of the order is at any given point because there is no way for the finance department, for example, to get into the warehouse's computer system to see whether the item has been shipped. This is usually attributed to firewalls existing in the system. If a customer is interested in finding out the status of an order, he or she may have to go through several departments asking each one what the status of the order. Without an ERP system in place, this is a very inefficient system. There are three major reasons why companies undertake ERP. These include integrating financial data, standardizing manufacturing and HR information [12]. Integrating financial data becomes a major reason to undertake the implementation for a system such as ERP as it simplifies the decision making for a top executive in a corporation. For example, as the CEO tries to understand the company's overall performance, he or she may find many different versions of the truth. The finance department has its own set of revenue numbers, the department of sales has another version, and other business units may each have their own versions of how much they contributed to revenues. ERP creates a single version of the truth that cannot be questioned because everyone is using the same system. The second major reason to undertake ERP involved standardizing manufacturing processes. Manufacturing companies often find that multiple business units across the company produce the same product using different methods and computer systems. Standardizing those processes and using a single, integrated computer system can save time, increase productivity and reduce headcount. Finally standardizing HR information is a final reason to implement an ERP system. In companies with multiple business units, HR may not have a unified, simple method for tracking employee time and communicating with them about benefits and services. ERP can serve to provide a more integrated approach to facilitate easier communication. ERP systems have emerged because the past decade the business environment has changed dramatically. Today, organizations are confronting new markets, new competition and increasing customer expectations. This has put a tremendous demand on manufacturers to deal with current day problems such as:

- Lower total costs in the complete supply chain
- Shorten throughput times
- Reduce stock to a minimum
- Enlarge product assortment
- Improve Product quality
- Provide more reliable delivery dates and higher service to the customer
- Efficiently coordinate global demand, supply and production.

Trying to grasp all these challenging goals becomes increasingly difficult if the necessary facilities are not available to attain those goals. Today's organizations have to constantly re-engineer their business practices and procedures to be more responsive to customers and competition [13].

IV. CONCLUSIONS

Data mining is a technology that has emerged to provide organizations whether large or small the opportunity to discovery-hidden trends and patterns in their data. This realization has come about as a result of the increasing loads of data being stored in organization’s databases. To take advantage of this storage data mining can use a data warehouse to manage the data before applying a data mining application. The reasons data mining has caught the attention of so many companies is that data mining has proven itself as a satisfactory tool. With the advent of ERP companies making progress in providing leading products and services, consolidation of data mining services alongside these services is a challenging path that can lead to very promising results. The future is still very uncertain. Because of the value that ERP companies can provide to organizations through their respective tools, an even greater benefit to companies is providing a data-mining tool that further analyses the data. In future every organization will have to find its way through this jungle of information, and data mining will play an active and crucially important role.

REFERENCES

[1]. Good, Nathaniel Schafer, Ben J. Konstan Joseph A. Borchers, Al Sarwar, Badru Herlocker, Jon, and John Riedl. Combining Collaborative Filtering with Personal Agents for Better Recommendations. Group lens Research Project. American Association for Artificial Intelligence. 1999.

[2]. Messinger, Eli Shoens, Kurt Thomas, John and Luniewski, Allen. Rufus: The Information Sponge. IBM Research. 1991.

[3]. Resnick, Paul Iacovou, Neophytos, Suchak, Mitesh Bergstrom,Peter and Riedl, John. GroupLens: An Open Architecture for Collaborative Filtering of Netnews.Association for Computing Machinery. 1999.

[4]. Breese, Jack Heckerman, David Kadie, Carl. Empirical Analysis of Predictive Algorithms for Collaborative Filtering. Microsoft Research. 1999.

[5]. Lueg, Christopher. Issues in Understanding Collaborative Filtering. AI-Lab, Department of Computer Science. University of Zurich. 1999.

[6]. Maltz, David and Ehrlich, Kate. Pointing the Way: Active Collaborative Filtering. Proceedings of the Annual ACM Sigch, Conference on Human Factors in Computing Systems. ACM Press. May 1995.

[7]. Pennock, David M. Horvitz, Eric.Collaborative Filtering by Personality Diagnosis: A Hybrid Memory- and Model- Based Approach. Microsoft Research. 1999.Directions Through ERP. ERP World 1999Proceedings. 1999.

[8]. Davenport, T.J. Enterprise Resource Planning Systems. Harvard Business Review. July-Aug 1998.

[9]. Ripma, Mark. Vendor Care and Feeding. ERP News. 1999.

[10]. Pallesen, Soren. How to Use the Internet to Reduce Sales Costs and Sell More. ERP World Proceedings. 1999.

[11]. Maude, Dan. Balancing Time, Scope, and Cost in ERP Implementations. ER World 1999 Proceedings. 1999.

[12]. Shtub, Avraham. Enterprise Resource Planning: The Dynamics of Operation Management. Kluwer Academic Publishing. 1999.

[13]. Shankarnarayanan S. ERP Systems – Using IT to gain a competitive advantage. Baan Infosystems India Pvt.Ltd. 1998.