Intelligent Data Processing Approaches managers use for Business Decision Support

Nadia Baeshen
King Abdulaziz Univ, Jeddah, Saudi Arabia

Abstract

Data mining has been used successfully for a number of years in the private and public sectors in a broad range of applications. In the private sector, these applications include customer relationship management, market research, retail and supply chain analysis, medical analysis and diagnostics, financial analysis, and fraud detection. In the government, data mining was initially used to detect financial fraud and abuse. Decision making processes for managing and defining policies related to social and environmental impacts from industry are becoming more and more complex. Thus, there is a need to integrate the available information within a structured network and to design a decision support system suitable for helping decision makers to define sustainable policies. Data warehousing, data mining, online analytical processing (OLAP) – these terms dominate discussions of enterprise decision support systems (DSS). This paper discusses a comparative study, from business decision support point of view, intelligent data analysis approaches; namely online analytical processing (OLAP), Data warehousing and data mining.

Keywords: Decision Support Systems, Data Warehouses, online analytical processing (OLAP) and Data Mining.

1. Introduction

Today’s worldwide marketplace provides not only more customers, suppliers and competitors, but also increased complexity for the decision-making process. Necessary data often is unavailable to the decision-maker for analysis or the requisite analyses are infeasible. It becomes important to derive methods that facilities the search for the requested data in order to add the manager and analyst in the decision-making process. One of these methods is data warehouse that acts as a central repository to supply all significant parts of the data to decision-makers. Most of the time, data warehouses can provide so much information that users become lost in the possibilities. Data mining tools were proposed to simplify analyses by employing filters based upon specific user-defined as well as other qualifying criteria. Both data mining and data warehouse are the foundation for decision support systems (DSS).

On the other hand, Business managers have the dual responsibility of serving customers and investors. To serve customers, managers need to understand demand, as observed in the marketplace and as reflected by measured customer needs, behavior, and attitudes. To serve investors, managers must provide a positive return through interest payments, dividends, and capital appreciation. They need to offer the right products at the right times, at the right prices. Retailers, manufacturers, and service organization need to understand their customers and the business environment. All can benefit by learning from customer and market data. There is much data, much to learn about data, and much to learn about data mining.

This paper will be organized as follow. The second section will review Decision Support Systems. Data warehousing will be discussed after that as well as the OLAP tool. The fourth section will discuss data mining. The fifth section explores Business Data Bases Preparation Processes. The sixth section will show the relation between data warehouse, data mining and decision support system. The seventh section discusses The Role of Data Mining in Application Areas in Business. At the end of the report the relation between data mining and data warehouse are mentioned as well as how they both prepare the data to be available for managers to aid them in the decision-making process.

2. Decision Support Systems (DSS)

The term Decision Support (DS) is used often and in a variety of contexts related to decision making. Recently, for example, it is often mentioned in connection with Data Warehouses
and On-Line Analytical Processing (OLAP) [1]. Another recent trend is to associate DS with Data Mining [2], which attempts to exploit these two approaches in a complementary way in order to support difficult real-life problem solving.

A decision support system (DSS) is a computer-based system that provides information to the users in making decisions and doing their jobs more effectively. Decision support systems (DSSs) are software products that help users apply analytical and scientific methods to decision making. They work by using models and algorithms from disciplines such as decision analysis, mathematical programming and optimization, stochastic modeling, simulation, and logic modeling. DSS products can execute, interpret, visualize, and interactively analyze these models over multiple scenarios.

Decision support systems (DSS) disciplines deal with the use of information technology to support human decision-making processes. Michael Scott-Morton, who virtually invented the discipline in the early 1970s, offered this definition of DSS. Decision support systems couple the intellectual resources of individuals with the capabilities of the computer to improve the quality of decisions. It is a computer-based support system for management decision-makers who deal with semi-structured problems [3]

Even this founding definition – like all definitions – presents us with problems: who is the decision-maker? What kinds of data serve as inputs to the decision-making process? What does the decision-making process itself look like? What kinds of risks and constraints are associated with the decision-making process? How is the output of the decision-making process – a decision – evaluated, implemented and tracked?

The key elements of a model of decision-making are fairly common, and include:

- **A decision-maker**: an individual or group charged with making a particular decision
- **A set of inputs to the decision-making process**: data, numerical or qualitative models for interpreting that data, historical experience with similar data sets or similar decision-making situations, and various kinds of cultural and psychological norms and constraints associated with decision-making
- **The decision-making process itself**: a set of steps, more or less well-understood, for transforming the inputs into outputs in the form of decisions,
- **A set of outputs from the decision-making process**, including the decisions themselves and (ideally) a set of criteria for evaluating decisions produced by the process against the set of needs, problems or objectives that occasioned the decision-making activity in the first place.

Their major characteristics are:

- DSS incorporate both data and models;
- They are designed to assist managers in semi-structured or unstructured decision-making processes;
- DSS support, rather than replace, managerial judgment;
- They are aimed at improving the effectiveness – rather than efficiency – of decisions.

3. Data Warehouse and OLAP

3.1. Data Warehouse

A Data Warehouse (DWH), as defined by Inmon [4], is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management’s decision-making process. The business success of an organization is highly dependable on the proactive use of information, which is stored in its operational systems. A DWH integrates the relevant information, originating from the diverse internal and external data sources. Data in a DWH is prepared for users with different analytical and software skills, and consequently also different types of requirements. Apart of responding to pure reporting and data analysis requests, a data warehouse also supports high-level users to track business trends, improve strategic decisions and enhance forecasting.

Fig 1: Data warehouse model [4]
Figure 1 shows the entire data warehouse process, from collecting source data to the ultimate data delivery to the decision makers. Source data is usually originated from the very diverse operational systems, where it is stored in source-system-specific formats. In the extraction phase, data is directly accessed from the legacy systems or, indirectly in case of proprietary systems. In the transformation phase, selected data is cleaned and converted into the format and structure compatible to the one existing inside the data warehouse. Syntactic and semantic distinctions between operational sources are adjusted and local logical models are mapped and integrated into the global data warehouse data model. In the data storage phase, new data is loaded into the data warehouse and merged with the existing and historically stored data. After being integrated into the data warehouse, data is ready for querying and analysing by OLAP and data mining tools. Decision makers and staff managers are supported by predefined reports or can retrieve desired information in an ad-hoc manner. Furthermore, there is a trend to incorporate alerting mechanisms into the data warehouse for necessary proactive near real-time actions.

A data warehouse has been defined as a collection of data in support of management decisions which is:

- Subject oriented
- Integrated
- Nonvolatile
- Time variant
- Metadata

Definitions are as follows:

**Subject oriented** Operational systems are organized around the applications that support the business functions of a company. ‘Subject’ refers to the queryable category of interest to the researcher. Data on a subject will likely come from multiple operational systems.

**Integrated** Data from all relevant database based applications is amalgamated so it can be accessed by the DSS professional. This is the most important aspect of a data warehouse. There is no point in bringing data over from the operational environment into the data warehouse without integrating it; then it could not be used to support a corporate view of data, which is the whole point of the decision support system in the first place.

**Nonvolatile** Operational data is volatile data. There is record-by-record manipulation of that data—it may be edited, deleted, and updated as well as accessed. In the data warehouse, there is a mass load of the information, and access is the only function performed on the data.

**Time variant** Operational data usually has a 60–90 day time horizon. Update of records is the primary task at hand. The key structure may or may not contain an element of time. In the analytical data-warehousing environment, the time horizon may be 5–10 years. It is a sophisticated snapshot of data at a point in time—which can be vital for future decision-making.

**Metadata** in a data warehouse plays the role of a card catalog in a library. It is metadata that allows an organization to track and understand where its data is. When an organization has an effective metadata structure, the DSS analysts can effectively find and analyze data.

### 3.2 OLAP (Online Analytical Processing)

On-line analytical processing is a data analysis technology that accomplishes the following:

- It presents a multidimensional, logical view of data to the end user with no requirements as to how the data is stored.
- It sorts, forecasts, tracks trends, and performs other complex analyses.
- It lets users move from one query to another and get results quickly and easily.

Why using OLAP with a data warehouse?

As a data warehouse is a “subject-oriented, integrated, time-varying, non-volatile collection of data that is used primarily in organizational decision making”, the data warehouse is maintained separately from the organization’s operational databases. The data warehouse supports on-line analytical processing (OLAP), the functional and performance requirements of which are quite different from those of the on-line transaction processing (OLTP) applications traditionally supported by the operational databases (see fig. 2).

OLTP applications typically automate clerical data processing tasks such as order entry and banking transactions that are the bread-and-butter day-to-day operations of an organization. These tasks are structured and repetitive, and consist of short, atomic, isolated transactions. The transactions require detailed, up-to-date data, and read or update a few (tens of) records accessed typically on their primary keys. Operational databases tend to be hundreds of megabytes to gigabytes in size. Consistency and recoverability of the database are critical, and maximizing transaction throughput is the key performance metric.

Data warehouses, in contrast, are targeted for decision support. Historical, summarized and
Data mining is becoming increasingly common in both the private and public sectors. Industries such as banking, insurance, medicine, and retailing commonly use data mining to reduce costs, enhance research, and increase sales.

Data mining involves the use of sophisticated data analysis tools to discover previously unknown, valid patterns and relationships in large data sets [5]. These tools can include statistical models, mathematical algorithms, and machine learning methods (algorithms that improve their performance automatically through experience, such as neural networks or decision trees). Consequently, data mining consists of more than collecting and managing data, it also includes analysis and prediction.

Data mining can be performed on data represented in quantitative, textual, or multimedia forms. Data mining applications can use a variety of parameters to examine the data. They include association (patterns where one event is connected to another event, such as purchasing a pen and purchasing paper), sequence or path analysis (patterns where one event leads to another event, such as the birth of a child and purchasing diapers), classification (identification of new patterns, such as coincidences between duct tape purchases and plastic sheeting purchases), clustering (finding and visually documenting groups of previously unknown facts, such as geographic location and brand preferences), and forecasting (discovering patterns from which one can make reasonable predictions regarding future activities, such as the prediction that people who join an athletic club may take exercise classes) [6].

Data mining does not require a human to ask specific questions. This makes data mining a fast and inexpensive way of summarizing, exploring, understanding, and analyzing data. Data mining integrates concepts from modern statistics, intelligent information systems, machine learning, pattern recognition, decision theory, data engineering and database management, and provides powerful tools that can reveal complex and hidden relationships in large amounts of data. The approaches include neural networks, genetic programming, and tree-based methods. Data mining already has a major impact on business and finance.

5. Business Data Bases Preparation Processes

Data are often unstructured, and we must impose structure. Data may be improperly structured, and we must restructure them prior to any data analysis. Data preparation, as described below, involves processes important to ensuring data integrity as well as proper structure.

1. **Recognition.** A first step in data preparation is to identify relevant data sources and to recognize which data items are relevant to the research problem at hand. Some use the term “data mart” to refer to a database developed to serve a specific research objective. To build a data mart, which is a collection of data extracted from other databases (both internal and external to the firm), we must recognize relevant data for inclusion in the data mart.

2. **Parsing.** The verb “to parse” has a grammatical meaning: to resolve a sentence into its component parts of speech. Statements in a computer language must also be parsed. Statements follow grammatical rules and must be broken down into their component parts in order to be executed. Likewise, data parsing involves structuring and restructuring data for further analysis. We take data in their original form and reformat them, so they may serve as input to data analysis programs. A common parsed structure has individual cases (units of
information that users become lost in the

Sometimes, data warehouses can provide so much

warehouse.

of applications that can make use of the data

The relation between data warehouse, data mining and decision support system could be viewed in the following way, both data mining and a data warehouse are considered as the foundation architecture of the decision support system. But at the same time, data mining and DSS are two kinds of applications that can make use of the data warehouse.

Sometimes, data warehouses can provide so much information that users become lost in the

possibilities. Data mining tools simplify analyses by employing filters based upon specific user-defined, qualifying criteria and rank filtering. Users can specify information to be found regarding a particular business unit and compare it to that of multiple business units or to the company as a whole. Sometimes scanning all relevant data can help decision-makers extract similarities among events and hence inspire hypotheses.

In some other times, through a DSS alone, a manager can make informed decisions. But, by interacting with a data warehouse a DSS can make better use of quality information and interpreting the information in a more efficient manner. Essentially, a data warehouse is a central repository for all or significant parts of the data that an enterprise or organization’s business systems collect.

7. The Role of Data Mining in Application Areas in Business

We can use data mining methods to explore data or search for structure. We can use them to complement or supplement traditional methods of analysis. Data mining methods may be used for explanation and predication. Let’s review common application areas in business[20].

1. Credit scoring and fraud detection.

Businesses want customers who buy in volume or at high margin, and businesses want customers who pay their bills. Credit scoring helps firms to find and keep customers who don’t. Past financial behavior and current financial circumstance can be good predictors of further financial behavior. Data mining models may be used for credit scoring, identifying those individuals who are credit risks.

2. Customer relationship management.

Customer relationship management is a natural extension of direct or database marketing. This is the one-to-one marketing approach, as described by Peppers and Rogers (1993). Firms build relationships with customers over time. A record of business transactions serves as a guide for fostering further transactions. Systems for customer relationship management can integrate business transactions data with marketing research data (Miller and James 2004). Firms can use what they know about customers to develop individualized or customized marketing plans.

3. Information management.

Last but not least in our list is the most general of data and text mining applications: information management.
Managers complain of information overload. There are too many documents to read, too many statistics to understand. Some define the field of statistics as the science of “data reduction”. Data and text mining can be of great assistance to the modern manager in reducing the amount of information that she must review.

4. **Pricing.** Pricing decisions can be informed by models for market response. They can also be built upon information about competitor pricing policies and consumer response to product features and product availability, as we learn from consumer choice studies. Data mining models can guide decision about pricing.

5. **Product positioning.** How do products within a category compare with one another? What is the nature of the product space? Which products serve as close substitutes for one another? How important is brand name as a differentiator of products?

6. **Brand Loyalty and Buyer Behavior.** Some customers are product- and brand- loyal; they purchase the same products and brands repeatedly. Other customers switch from one product or brand to the next, influenced by lower prices, promotions, or a desire for variety. Switching costs can be high in product categories that require consumer learning, such as computer software. In other categories, such as cellular phones, products and services may be viewed as interchangeable, and switching costs are low. Models of buyer behavior help us to understand buying processes, including trial purchase, repeat purchase, brand loyalty, and switching. Methods of analysis draw upon applied probability (stochastic) models, recurrent events data (reliability) analysis, and survival data analysis.

7. **Competitive Intelligence.** Learning about the competition is critical to business success. The Internet and World Wide provide access to a vast information store for research about the competition and competitive products. We can use data mining tools to organize and reduce data and to search for relevant information. The methods of text mining are especially useful in this area.

8. **Consumer choice modeling.** Choice is fundamental to business, and models that predict consumer choice are of great value to management. Some people buy a pay their bills. Some shop online; others shop in person. Adding up the choices of individual consumers across products within a category, we get estimates of market shares. Finding out what makes people choose one product over another is a key to successful marketing. Logistic regression is a traditional method of choice modeling. Classification trees and neural networks are examples of modern data-adaptive methods for choice modeling.

9. **Direct and database marketing.** Direct and database marketing are common applications for data mining. Direct marketing firms contact prospects and customers directly, sending promotional brochures. These firms can increase profits by mailing only to those prospects and customers most likely to buy. Data mining models may be used to identify these customers. Anand and Buchner (2002), for example, discuss applications of data mining techniques in cross-selling, finding prospects for additional products from an existing customers list.

10. **Market basket analysis.** Data to the things we buy and the prices we pay are important to retail management. These are the empirical ramifications of demand. Walking through the aisles of supermarkets, we see hundreds of thousands of items from which to choose. The market basket of products we take to the checkout counter reveals our product preferences in a way that no survey could fully capture. Individual items, identified by stock keeping units (SKUs) and universal product codes (UPCs), might be grouped by product or category type. Sometimes called affinity or association analysis, market basket analysis asks, “What goes with what?” There are obvious things, like hot dogs and hot sometimes bogus, of less obvious things going together, like diapers and beer. We can use market basket analysis to guide product placement in stores and to develop co-marketing and product bundling plans. More generally, we can apply market basket analysis in the study of consumer lifestyles and behavior.

11. **Market response modeling and sales forecasting.** Explanatory variables associated with the marketing mix (product characteristics, price, promotions, and advertising) affect market response, consumer demand, or sales. Traditional regression and modern data-adaptive regression are appropriate tools for many response modeling and sales forecasting problems.

12. **Market segmentation.** Mass marketing involves using a common method to sell to
everyone. One-to-one marketing involves customizing a firm’s marketing plan to fit the needs of each individual consumer. Targeted marketing falls between mass marketing and one-to-one marketing, and market segmentation is a tool for targeted marketing. We identify groups or segments of customers who are more similar to one another than to customers outside their respective groups for segments. Segments are often based upon geo-demographic data. “Location, location, location,” long the mantra of business, has relevance today despite our migration toward a networked nation. Many products and services are location-dependent. Grocery and convenience store, restaurants, Laundromats, and health and fitness clubs know that there are advantages in being close to where people live and work.

8. Data mining and Data Warehouse; The foundation of Decision Support System

There is a symbolic relationship between the activity of data mining and the data warehouse—the architectural foundation of decision support systems. The data warehouse sets the stage for effective data mining. Data mining can be done where there is no data warehouse, but the data warehouse greatly improves the chances of success in data mining.

How does the data warehouse set the stage?
Consider the nature of a data warehouse, which includes:
- Integrated data
- Detailed and summarized data
- Historical data
- Metadata

Each of these elements enhances the data mining process and the prospects for success:

Integrated data allows the miner to easily and quickly look across vistas of data. Without integrated data, the miner would spend inordinate amounts of time cleansing and conditioning the data before the process of data mining could commence in an effective manner. Keys would have to be reconstituted, encoded values reconciled, structures of data standardized, and so forth before the miner could actually get to the task of doing data mining with raw data. The data warehouse is integrated and has all these tasks (and many more) done so the miner can concentrate on mining data rather than cleansing and integrating data.

Detailed and summarized data are both included in the data warehouse. Detailed data is necessary when the miner wishes to examine data in its most granular form. Very low levels of detail hide important patterns that can be discerned no other way than by carefully scrutinizing the detail. By the same token, summarized data ensures that if a previous analysis is already made, the miner does not have to repeat the work someone else has already done before beginning the process of exploration. Summarized data ensures the miner can build on the work of others rather than build everything from the ground up. This capability, afforded through easy availability of summarized data, saves huge amounts of unnecessary work for the miner.

Historical data is important to the miner because important nuggets of information are hidden there. A miner who has to work with only very current information can never detect trends and long-term patterns of behavior. Historical information is crucial to understanding the seasonality of business and the larger cycles of business to which every corporation is subject.

Metadata serves as a road map to the miner, who uses metadata to describe not the content but the context of information. When information is being examined over time, context becomes as relevant as content. Stated another way, raw content of data becomes very difficult for the data miner to work with when there is no explanation for the meaning of the data. Therefore, the data warehouse sets the stage for successful and efficient exploration of the world of data, and the miner lucky enough to work on a foundation of a data warehouse enjoys success that comes from exploitation of this data as a resource.

9. Conclusion

Both data mining and data warehouse are considered the architectural foundation of decision support systems. These two models could work together in order to produce subject-oriented, integrated, time-varying, non-volatile collection of data that is used primarily in organizational decision making.

Reference

3. 1 P.G. Keen and M.S. Scott-Morton. Decision Support Systems: An


7. An Overview of Data Warehousing and OLAP Technology, Surajit Chaudhuri Microsoft Research, Redmond, Umeshwar Dayal Hewlett-Packard Labs, Palo Alto

8. Intuitive Decision-Making, Vicki L. Sauter

9. An Introduction to Data Warehousing: What Are the Implications for the Network? Katherine Jones,

10. Data Warehousing and OLAP for Decision Support, Surajit Chaudhuri Microsoft Research, Redmond, Umeshwar Dayal Hewlett-Packard Labs, Palo Alto

11. The Data Warehouse and Data Mining, W. H. Inmon

12. Data Mining: Concepts and Techniques, Jiawei Han and Micheline Kamber.


