The current issue and full text archive of this journal is available at
www.emeraldinsight.com/0263-5577.htm

Advanced analytics: opportunities and challenges

Ranjit Bose
Anderson School of Management, University of New Mexico, Albuquerque, New Mexico, USA

Received 24 July 2008
Revised 18 August 2008
Accepted 18 September 2008

Abstract

Purpose – Advanced analytics-driven data analyses allow enterprises to have a complete or "360 degrees" view of their operations and customers. The insight that they gain from such analyses is then used to direct, optimize, and automate their decision making to successfully achieve their organizational goals. Data, text, and web mining technologies are some of the key contributors to making advanced analytics possible. This paper aims to investigate these three mining technologies in terms of how they are used and the issues that are related to their effective implementation and management within the broader context of predictive or advanced analytics.

Design/methodology/approach – A range of recently published research literature on business intelligence (BI); predictive analytics; and data, text and web mining is reviewed to explore their current state, issues and challenges learned from their practice.

Findings – The findings are reported in two parts. The first part discusses a framework for BI using the data, text, and web mining technologies for advanced analytics; and the second part identifies and discusses the opportunities and challenges the business managers dealing with these technologies face for gaining competitive advantages for their businesses.

Originality/value – The study findings are intended to assist business managers to effectively understand the issues and emerging technologies behind advanced analytics implementation.

Keywords Data analysis, Competitive advantage

Paper type Research paper

1. Introduction

What differentiates companies in today's highly competitive markets is their ability to make accurate, timely, and effective decisions at all levels – operational, tactical, and strategic – to address their customers' preferences and priorities. Increasingly, companies in virtually every industry around the globe have started using advanced (also known as predictive) analytics to analyze their data (both structured and unstructured), combining information on past circumstances, present events, and projected future actions. By incorporating advanced analytics into their daily operations, these organizations gain control over the decisions they make every day, so that they can successfully meet their business goals (Apte *et al.*, 2003).

The advanced analytics driven data analyses allow enterprises to have a complete or "360 degrees" view of their operations and customers. The insight that they gain from such analyses is then used to direct, optimize, and automate their decision making. It results in successful achievement of a variety of specific organizational goals, whether they are associated with an increase in cross-sell revenue generation, a decrease in production or service cost, a reduction in fraudulent behavior, or an increase in promotional campaign response rates.

Industrial Management & Data Systems
Vol. 109 No. 2, 2009
pp. 155-172

0263-5577
DOI 10.1108/02635570910930073

Advanced analytics is a general term which simply means applying various advanced analytic techniques to data to answer questions or solve problems. It is not a technology in and of itself, but rather, a group of tools that are used in combination with one another to gain information, analyze that information, and predict outcomes of the problem solutions. Data integration and data mining are the basis for advanced analytics. The more information that is gathered and integrated allows for more pattern recognition and relationship identification. Statistical analysis is another very important component to see trends and patterns in the data. Some other techniques used to manipulate the data is fuzzy logic, to deal with incomplete or ambiguous data, and neural networks to anticipate decisions and assist in predictive analytics which helps predict likely outcomes (Wu *et al.*, 2006).

Data mining is a powerful emergent technology for the automatic extraction of patterns, associations, changes, anomalies and significant structures from data. These uncovered patterns from data play a critical role in decision making because they reveal areas for process improvement. Most of the value of data mining comes from using data mining technology to improve predictive modeling (Wang and Wang, 2008). For example, data mining can be used to generate predictive models automatically, which predict how much profit prospects and customers will provide and how much risk will entail from fraud, bankruptcy, charge-off and related problems.

Recent advances have led to the newest and hottest trends in data mining – text mining and web mining (Hearst, 2003; Fan *et al.*, 2006). These two data mining technologies open a rich vein of enterprise performance and customer data in the form of textual comments from survey research or e-mails and log files from Web servers, which were previously unusable. Applying text and web mining to these data adds a richness and depth to the patterns already uncovered through the company's data mining efforts. Text mining applies the same analytical functions of data mining to the domain of textual information, relying on sophisticated, text analysis techniques that distill information from free-text documents (Dorre *et al.*, 1999; Oliveira *et al.*, 2004). Web mining is a form of data mining used to discover patterns from the web. There are three categories of web mining: web content mining, web structure mining, and web usage mining.

This study aims to investigate these three mining technologies in terms of how they are used and the issues that are related to their effective implementation and management within the broader context of predictive or advanced analytics.

2. Current state of business intelligence

The managerial view of business intelligence (BI) is getting the right information to the right people at the right time so they can make decisions that ultimately improve enterprise performance. The technical view of BI usually centers on the process of, or applications and technologies for, gathering, storing, analyzing and providing access to data to help make better business decisions.

This section summarizes the evolution of BI over the years to its current state. Figure 1 diagrammatically captures the stages in the BI evolution. Each of these stages is briefly discussed below.

Rapid advances over the last several years in data capture, processing power, data transmission, data transformation, and storage capabilities have enabled organizations to integrate their various databases into data warehouses. Data warehousing is conceptualized as a process of centralized data management and retrieval. The core of a well developed BI program is the data warehouse.

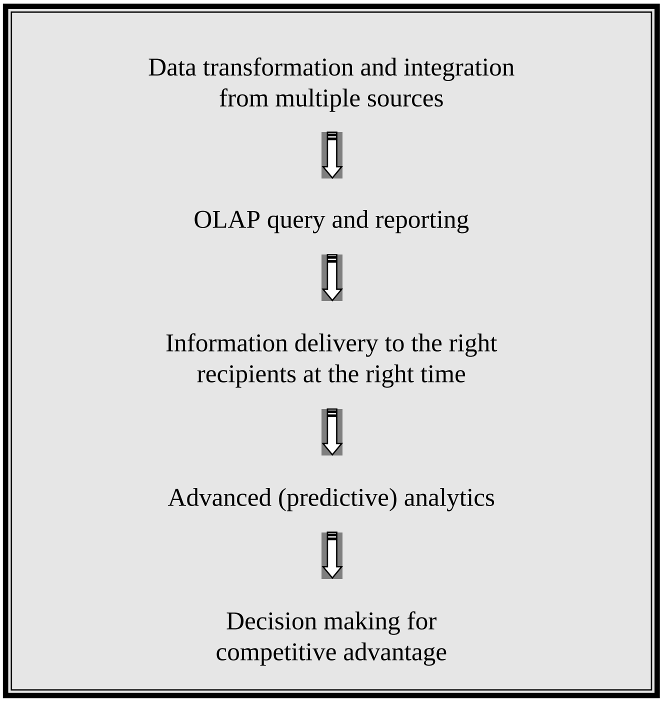

Figure 1.
BI evolution

While BI was analogous to OLAP (online analytical processing) query-and-reporting tools for many years, many organizations discovered that the effective use of information requires more than reports that show historical data. Therefore, in addition to the information-delivery component of BI, advanced analytics started garnering just as much attention in overall BI strategies. Effective decision making for competitive advantage is driving the need for such a more comprehensive approach to BI.

Recent forecasts by industry analysts suggest that advanced analytics will be the driving force in the BI market for some time to come. Organizations that are currently in the process of embracing each stage of the BI evolution in Figure 1, have traditionally focused on tasks that are inherently foundational to better enterprise performance – data capture, data architecture and information delivery. Increasingly these organizations are discovering that they can build a sustainable competitive advantage through adaptive and embedded analytics, and at the same time receive a better return on investment (ROI) on the BI infrastructure investments that they have made over the years.

The promise of embedded analytics or decision-oriented analytic applications is to allow for informed decision making. For example, it can be described in terms of a knowledge management initiative, in which the organization's best practices for each decision making process are pushed to the desktops of end-users as embedded logic within analytic applications. These applications are typically powered either by business rules engines (which apply logical conditions to determine how a certain case should be handled) or predictive models (which probabilistically identify the action most likely to achieve the desired results).

Another important observation in the BI evolution is that industry leaders are currently transitioning from an operational BI of the past to an analytical BI of the future that focuses on customers, resources, and abilities to drive new decisions everyday. They have implemented one or more forms of advanced analytics for meeting these business needs. Table I helps distinguish the operational and analytical BI. Advanced analytics provide the comprehensive insight necessary for pinpointing revenue opportunities, enhancing sales channels, and mitigating cost risks. By providing meaningful insight into data, as well as transactional predictions, advanced analytics based systems such as analytical BI enable businesses to ensure that rules and workflow are in step with customer demands.

3. Advanced analytics with data, text, and web mining

Advanced analytics is a suite or cluster of analytical applications that helps measure, predict, and optimize organizational performance and customer relationships. To execute a successful BI strategy, the IT infrastructure must be aligned with business needs in a way that the infrastructure supports the business in achieving goals and objectives. A successful BI infrastructure must be able to transform disparate data and systems into an efficient flow of information, analyze data with a forward-looking view, and deliver key information to decision makers on demand. Figure 2 depicts such an infrastructural framework for BI using advanced analytics.

The enterprise BI architecture consists of the following components:

- *Data integration* – capabilities for both structured and unstructured (such as text) data connectivity, data quality, ETL (extract, transform, and load), data migration, data synchronization, and data federation.
- *Advanced analytics* – data, text, and web mining capabilities, visualization systems, recommendation systems, predictive and descriptive modeling, forecasting, optimization, simulation, and more.

Characteristics	Operational BI	Analytical BI
Focus on	Profitable transactions to drive operational efficiency	Creation and use of intelligence to drive differentiation from the competitors
Decision types	Operational (day-to-day)	Tactical/strategic
Underlying technology	OLAP query and reporting	Advanced (predictive) analytics
Emphasize on	Customer acquisition	Customer retention
ROI	Moderate to high	High
Measures	Customer satisfaction	Customer value and loyalty
Customer management organized by	Function or product units	Customer segments
Rely on	Information about customer	Information from customers
Type of interactions	Proactive interactions with customers	Real-time, personalized interactions with customers
Scope	Internal, company focus toward improvement	External, customer focus toward improvement
Learning	Long-loop learning and implementation	Short-loop learning and implementation

Table I.
Operational BI vs analytical BI

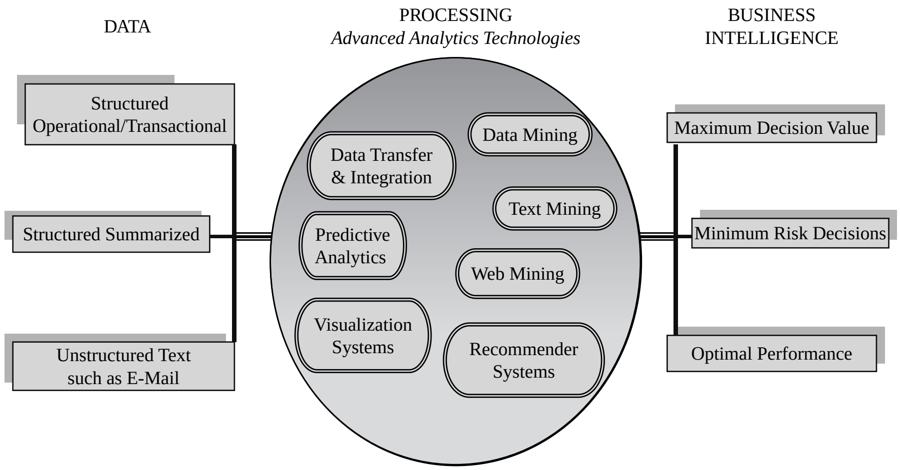

Figure 2.
Framework for BI using avanced analytics

- *Intelligence delivery* – capabilities to deliver the BI timely for effective decision making for competitive advantage across the enterprise.

Not only each of these components must be integrated within the BI architecture, but must also be with the existing investments in hardware and software. A comprehensive BI architecture will be able to access all enterprise data no matter where it resides and no matter what operating system is being used. It will transform the data into the required format, if necessary, store it properly, analyze it thoroughly, and disseminate intelligence to users through familiar interfaces. Advanced analytics must easily integrate into intelligence delivery component so that results can be shared with managers and decision makers. The implementation technologies should include a wide portfolio of algorithms, visualization techniques, flexible data exploration and manipulation capabilities that produce accurate models. Results should be delivered in a format that even non-technical users can understand – using everyday business terms to explain the results.

In the remainder of this section, summaries of data, text, and web mining backgrounds are provided to assist the discussions in the subsequent sections.

Data mining can be conceptualized as the automated extraction of hidden predictive information from databases. In other words, it is the process of analyzing large data sets in order to find patterns that can help to isolate key variables to build predictive models for management decision making.

There are many data mining methods for extracting patterns from data. These methods can have different goals, dependent on the intended outcome of the overall data mining process. Most data mining goals fall under the following categories: data processing; prediction; regression; classification; clustering; link analysis (associations); model visualization; and exploratory data analysis.

Any data mining method that helps to get more information out of data is useful. Different methods serve different purposes, each offers its own advantages and disadvantages. The most commonly used methods for data mining can be classified into the following groups: statistical methods; case-based reasoning; neural networks;

decision trees; rule induction; Bayesian belief networks; genetic algorithms/evolutionary programming; fuzzy sets; and rough sets.

Different combinations of data mining goals and methods are used to ensure flexibility and the greatest accuracy possible in the process. The major benefit of data mining is that it is an aid to strategic, tactical and operational decision making in situations where numerous variables, affecting costs or benefits, impinge on the eventual outcome of the course of action that a company might decide to take. The modeling that accompanies data mining assimilates the information on costs and benefits of alternative courses of action as visualized in the form of a familiar method such as decision trees. Companies use such information to find new opportunities for growth, choose more effective means to achieve their business goals and streamline business processes to lower their costs.

Data mining and visualization tools are used in combination to improve the usability of advanced analytics systems. The purpose of data visualization system is to give the user an understanding of what is going on as far as data mining models and their outputs are concerned. Since data mining usually involves extracting "hidden" information from a database, this understanding process can get somewhat complicated. Because the user does not know beforehand what the data mining process has discovered, it is much bigger leap to take the output of the system and translate it into an actionable solution to a business problem.

While data mining successfully helps find the gold hidden in a company's data, it addresses only a very limited part of a company's total data assets: the structured information available in databases. Probably more than 90 percent of a company's data are never being tapped or looked at: letters from customers, email, correspondence, recording of phone calls with customers, contracts, technical documentation, patents, and so on (Weiss *et al.*, 2005). The emergent text mining tools help dig out the hidden gold from these unstructured information sources. Text mining technology has lately leaped from information search and retrieval to intelligence and knowledge discovery (Rao, 2003).

Text mining applies the same analytical functions of data mining to the domain of textual information, relying on sophisticated, text analysis techniques that distill information from free-text documents. Text mining software operates on the digitized form of organizational textual data to provide the capability of pattern identification, visualization support to aid pattern identification, modeling support to identify or confirm relationships, and drill-down query tools to enable analysts to focus on key problem areas (Emery, 2005). Report generation tools also aid the text mining process.

The text mining process typically includes the following steps:

(1) preprocessing of the data to the needed format for further analysis (data preprocessing);

(2) extraction of important concepts and terms through initial text analysis (concept extraction);

(3) writing a narrative analysis to identify patterns and co-occurrences of identified concepts (narrative analysis);

(4) developing an automated solution (automatic categorization); and

(5) building a taxonomy (taxonomy building).

Web mining involves the application of data mining techniques for the extraction of interesting and potentially useful patterns and implicit information from artifacts or activities related to the World Wide Web. Computers on the internet that host websites, the web servers, are capable of collecting information about websites usage. This information is a valuable repository for mining and discovering interesting patterns using statistical or artificial intelligence methods.

There are three knowledge discovery domains that pertain to web mining:

(1) Web content mining.

(2) Web structure mining.

(3) Web usage mining.

Web content mining is the process of extracting knowledge from the content, typically the text of documents or their descriptions. Web structure mining is the process of inferring knowledge from the World Wide Web organization and links between references and referents in the web. It investigates the site's hyperlink and document structure. Finally, web usage mining, also known as web log mining, analyzes user behavior on the web site by extracting interesting patterns from the web server logs. It is a method for understanding user behavior as it relates to the use of websites. The results of web mining can be used to provide metrics on the effectiveness of a company's web site or the success of a particular campaign (Li and Zaiane, 2004).

Recently, software vendors and researchers have been focusing on using the extracted patterns from web miming to predict the next user request during an online session with a web site, especially e-commerce (Mobasher *et al.*, 2000). Such systems are called recommender systems and are useful tools to predict user requests.

4. Opportunities with advanced analytics

This section identifies and discusses some of the opportunities the business managers dealing with the aforementioned mining technologies face for gaining competitive advantages for their businesses.

4.1 Data mining

Data mining is primarily used for competitive advantages by companies with a strong consumer focus. The focus of data mining applications amongst the business leaders has been steadily evolving from customer analytics to relationship analytics. Increasingly competitive business and consumer marketplaces make it imperative for companies not only to attract customers, but also to retain them especially that small percentage of highly profitable customers. Retention strategies for valued customers generally focus on financial and/or service-level incentives to promote loyalty. Since only few companies can enjoy the economies of scale (or investment capital) to sustain competitive differentiation on price alone, many businesses seek to maximize customer value by building loyalty through brand and service differentiation. This approach place a premium on the quality of every customer contact as each interaction serves to either build brand or destroy it.

These business leaders use advanced analytics with data mining to optimize their customer relationships (Marsella *et al.*, 2005). Examples include: improving the effectiveness of marketing campaigns and attracting new customers, maximizing the value of sales to existing customers (cross-selling and up-selling), minimizing customer loss (churn), credit risk scoring, and lifetime value modeling and analysis. Data mining techniques are also used to analyze and monitor levels of customer satisfaction and loyalty and diagnose the causes of changes in these levels.

An area of use of advanced and predictive analytics that has been growing exponentially, since September 11, 2001 is in fighting crime and terrorism (McCue, 2005; Jonas and Harper, 2006). Many law enforcement agencies have turned to advanced analytics in an attempt to get a leg up on crime fighting. So far, it has been most helpful in the areas of fraud detection, identity theft, and tax evasion.

4.2 Text mining

As is the case with data mining technology, one of the primary application areas of text mining is collecting and condensing facts as a basis for decision support. There are three main advantages the text mining technology offers over a traditional "information broker" business:

(1) The ability to quickly process large amounts of textual data compared to the speed of human readers.

(2) The "objectivity" and customizability of the process –, i.e. the results solely depend on the outcome of the linguistic processing algorithms and statistical calculations provided by the text mining technology.

(3) The possibility to automate labor-intensive routine tasks and leave the more demanding tasks to human readers.

Utilizing these advantages, text mining applications today are typically used to:

- extract relevant information from a document (summarization, feature extraction, etc.);
- gain insights about trends, relations between people/places/organizations, etc. by automatically aggregating and comparing information extracted from documents of a certain type (e.g. incoming mail, customer letters, news-wires, etc.);
- classify and organize documents according to their content, i.e. automatically pre-select groups of documents with a specific topic and assign them to the appropriate person; and
- organize repositories of document-related meta-information for search and retrieval; and retrieve documents based on various sorts of information about the document content.

This list of activities shows that the text mining technology covers two main application areas:

(1) knowledge discovery (mining proper); and

(2) information "distillation" (mining on the basis of some pre-established structure) (Froelich *et al.*, 2005).

The technologies used in the text mining process include: information extraction, topic tracking, summarization, categorization (identifying the main theme of a document), clustering, concept linkage (tools connect related documents by identifying their shared concepts, helping users find information they perhaps would not have found through traditional search methods), information visualization (puts large textual sources in a visual hierarchy or map and provides browsing capabilities, in addition to simple searching), and question answering.

A new trend involves integration of data mining and text mining into a single system, a combination known as duo-mining (Feldman and Sanger, 2007). This combination has proven especially useful in banking and credit card customer relationship management. Instead of being able to analyze only the structured data they collect from transactions, they can add call logs associated with customer service and further analyze customer spending patterns from the text-mining technology – beyond simple search methods – are the key to information discovery and promise support in all areas. Companies with vast document collections sitting idle should consider investing in text mining applications that would help them analyze their documents and provide payback with the information they provide.

Another application of text mining, which is gaining in popularity, is named electronic discovery. Electronic discovery has its roots in the field of civil litigation support and deals with organizing electronic files using their attached metadata (Volonino, 2003). Because of the large volume encountered, these files are usually incorporated into a litigation retrieval system to allow review and production in an easy methodology. Legal data management principles are used, including redaction rules and production methodologies. Electronic discovery tasks usually begin after the files are captured. File metadata is used to organize and cull the collections. Documents can be examined in their native file format or converted to TIF or PDF images to allow for redaction and easy production.

4.3 Web mining

There are several kinds of web mining software tools available today. A web usage mining and analysis tool tracks user browsing patterns, generates reports to help webmasters refine web site structure and navigation (Mobasher, 2005). There are also tools that provide online behavioral analysis of the users. Yet other tools allow a comprehensive access log analysis. They allow one to keep track of activity on their site by month, week, day, and hour, to monitor total hits, bytes transferred and page views, and to keep track of the site's most popular pages.

Recent studies have shown that content and structure of a web site can play an important role in the quality of recommendations provided by a recommender system (Adomavicius and Tuzhilin, 2005). These systems have traditionally focused on applications in e-commerce, where users are recommended products such as books, movies, or other commercial products from a large catalog of available items. The goal of such systems has been to tap the user's interest by utilizing information from various data sources to make relevant recommendations, keeping the user engaged and, hopefully, resulting in increased sales.

Recommender systems is currently an active area of research and development, primarily due to its applicability in a variety of practical situations requiring a user to select from a set of options best suiting his/her interests (Srivastava *et al.*, 2004). Recommender systems are categorized based on their approach, into four categories:

(1) *Content-based approach* – these recommendations are derived from the similarity of a product the user is interested in with other products that the user has bought/referred to in the past and preference level to those products as expressed by the user.

(2) *Collaborative approach* – similarity is measured with respect to the end products purchased by other users. For instance, a shared subset of previously-purchased items between the user and other customers would be a driving factor for recommendations.
(3) *Usage-behavior recommendations* – similarity based on usage behavior of users is used as a criterion for recommendations. These recommendations are mapped to correlations in browsing behavior and mined from logs such as those generated by web servers.
(4) *Hybrid approach* – the approach can be any combination of the above three approaches. Please note that the content-based and collaborative filtering approaches require the availability of user profiles and their explicit solicited opinion on the targeted products.

4.4 Popular advanced analytics tools

Table II provides a list of some popular advanced analytics products, their company's contact, and a brief overview of their capabilities.

5. Challenges with advanced analytics

While the promise of advanced analytics, both for cost reduction and revenue increase is significant, this cannot be achieved unless there is successful adoption and use of it within an organization. Some of the key challenges encountered in advanced analytics adoption and use are identified and summarized below.

5.1 Organizational "buy in"

While data mining and data warehousing are very powerful technologies, with a proven track record, there are also enough examples of failures when technology is deployed without sufficient "buy in." It is crucial to have a "buy in" from all the business units to ensure that the results will be used appropriately. A number of steps must be taken to ensure this happens:

(1) there needs to be a cross functional team involved in implementing an advanced analytics pilot project in the organization. While the technical members on the team play an important role, an executive on the business side, who should also be the project owner and sponsor, should head the team;
(2) processes need to be adopted, with an appropriate set of measureable metrics, to ensure that all steps for project success are being taken; and
(3) incentives for performing well on the project should be included as part of the reward structure to ensure motivation.

5.2 Implementation of advanced analytics

Introduction of advanced analytics in an organization must be managed carefully. Given its high initial cost, and significant change on the organizational processes, it is quite possible that insufficient care in its introduction may lead to high expense, seemingly small early benefits, which can lead to low morale and excessive finger-pointing.

It is better to have an incremental approach rather than a big bang approach. The benefits accrued from the first stage become evident, and act as a catalyst for accelerating the subsequent stages. This makes the choice of the first project and its

Product category/name	Company's website	Comment
Comprehensive BI suite		
Business Objects XI 3.0	www.businessobjects.com	Offers a broad family of BI tools and applications for different sizes of businesses
IBM Cognos 8 BI	www.cognos.com	Offers a wide range of BI capabilities on a single, service-oriented –architecture (SOA)
Data Mining Tool		
Clementine from SPSS	www.spss.com	Leading visual rapid modeling environment for data mining, available in several major languages
DBMiner 2.0 (Enterprise)	www.dbminer.com	Powerful and affordable tool to mine large databases; uses Microsoft SQL Server 7.0 Plato
Oracle Data Mining (ODM)	www.oracle.com	Uses data mining functionality embedded in Oracle Database 10 g Enterprise Edition
SAS Enterprise Miner	www.sas.com	Integrated suite, provides an user-friendly GUI frontend to the SEMMA (Sample, Explore, Modify, Model, Assess) process
Text Mining Tool		
Inxight	www.businessobjects.com	Enterprise software solutions to understand and analyze information contained in unstructured text, in all major languages
SAS Text Miner	www.sas.com	Provides a rich suite of tools for discovering and extracting knowledge from text documents
SPSS LexiQuest	www.spss.com	To mine the "unstructured" information contained in text documents, integrated with SPSS Clementine data mining tool
Web Mining Tool		
MicroStrategy Web Traffic Analysis Module	www.microstrategy.com	Built on MicroStrategy 7 platform, provides traffic highlights, contents analysis, and web visitor analysis reports
SAS Webhound	www.sas.com	Analyzes website traffic to answer questions like: Who is visiting? How long do they stay? What are they looking at?
SPSS Web Mining for Clementine	www.spss.com	Enables extraction of Web events, including online campaign results, uses this online behavior in predictive modeling
WebTrends	www.webtrends.com	A suite for data mining of web traffic information to learn and use the customer behavior to take targeted action

Table II.
A few popular advanced analytics tools

team very critical. Ideally, the project must be in a potentially high-impact area, where the current process is very ineffective.

5.3 Regulations and privacy information

Privacy is often associated with protection of personal information such as names, addresses, social security numbers, credit card numbers, medical records, and financial records. In the context of business analytics, individual privacy disclosure occurs when the identity of an individual or the specific value of sensitive information of an individual is revealed during the analysis. Organizational privacy disclosure occurs when information about the business operations (e.g. customer profiles, product ranges) and strategies are disclosed to unauthorized parties (Lui and Qiu, 2007).

The need to slice and dice data more carefully has been sparked by regulations designed to protect personal medical and financial information, in addition to other types of data. State and national do-not-call registries have set restrictions, and pending anti-spam legislation seeks to do the same on an electronic level. With identity-theft threats and social security issues, consumers are wary of marketing messages that seem to draw on their personal preferences. This means marketers are limited in how they can create customer lists, contact customers, and craft their messages.

5.4 Challenges inherent in the advanced analytics technologies

Another challenge is that the science of advanced analytics is not an easy concept or technology for users to understand or know how to use. Since it is being used among business enterprises, either computer scientists/specialists need to be hired and trained or business managers need to be trained to be able to understand and utilize these systems. Due to this limitation, enterprises or agencies that can afford the technology may not have the expertise to utilize it effectively or to its fullest potential.

Yet another challenge is the rate with which the technology changes, evolves, and advances in this relatively new area of specialization. It can be a difficult line between innovative technology and technology that is stable and works. Enterprises have to be able to balance progresses in technology with stable systems that are known to work but if they stay with it too long, it will be outdated.

The form of output is yet another challenge. The inputs to advanced analytics include immense amounts of data but the output needs to be simple, concise, readable, and usable. Finding or designing a system that is able to analyze the data and return output in a way that is valuable to the end-users is extremely important. By using support tools like dashboards, reports, and visualization systems the information can be relayed in usable and understandable formats.

5.5 Availability and sharing data among organizations

Data are the key building blocks for systems focused on modeling and analysis. Internal data – sometimes called "household data" – are core differentiators in the analytical BI process. The most sophisticated analytical tool can be rendered ineffective if the appropriate data is not available or the data quality is poor. To truly excel at advanced analytics, an organization needs detailed information about the needs, values, and wants of its customers. Leading organizations gather data from many customer touch points and external sources, and bring it together in a common, centralized repository, in a form that is available and ready to be analyzed when needed.

Another challenge is sharing data across organizations because the data has to be secured, maintain privacy and confidentiality for some of the more sensitive data, and handling shared data across different platforms containing different vocabularies.

6. Using advanced analytics

Advanced analytics technologies such as data-, text-, and web-mining are primarily used by companies with a strong consumer focus. They form the basis of rapidly growing tools in management decision making. Advanced analytics support executives at all levels – strategic, tactical, and operational – for their decision making needs. Strategic managers use them for competitive intelligence, identifying market opportunities, product launch decisions and product positioning. Tactical managers use them to make decisions in the areas of sales forecasting, direct marketing, customer acquisition, retention and extension purposes, and marketing campaign analysis. Operational managers use them for decisions involving better utilization of facilities or supply chain management.

The term advanced analytics denotes the overall process of turning low-level data – database, textual, and Web – into high-level knowledge by extracting patterns or models from observed data. The mining of data in these three forms uncovers patterns in them using predictive techniques. These patterns play a critical role in decision making because they reveal areas for process improvement. Using mining of data, organizations can increase the profitability of their interactions with customers, detect fraud, and improve risk management. The patterns uncovered in the mining process help organizations make better and timelier decisions.

Profiling or segmentation of the customer base is frequently used in many organizations to demonstrate the significance of using mining of data. The task of profiling consists of identifying homogeneous groups of customers who exhibit similar patterns of behavior. With this information, companies learn to target specific customers rather than randomly promote products to all of them, which increase their marketing cost-effectiveness.

Customer segmentation can help companies innovate in designing and pricing of new products that are more attractive to their consumers. An example is the design of insurance plans for the health industry. Health insurance companies can offer consumer driven health plans tailored to the risk profile of customers, compared with the traditional two broad types of insurance plans the HMO and the PPO. Similarly, financial services companies can reduce their risks in their lending operations by using mining of their customer data. In the past, loans were granted to only those customers who had a sound history of repayment and were extended to borrowers who could secure their loans with collateral. However, mining of the data could help create newer strategy for determining credit worthiness such as customers without a credit history could have the potential to be good borrowers.

In general, profiling and segmentation of customers helps companies to efficiently align resources with the specific needs of customers. They can set prices, choose channels and design communication strategies based on the character of specific segments of customers.

Table III lists the steps one goes through to use advanced analytics. Each of these steps is iterative and interactive. The goals or business objectives best suited to the advanced analytics process are those involving prediction or seeking an explanation of behavior. For example, goals might include determining best customer profiles and predicting customers who are most at risk to churn. Once initial objectives have been defined, the next step is to determine what data are needed to address the objectives and how the data will be obtained.

Table III.
Steps showing how to use advanced analytics

Steps involved in using advanced analytics	
1.	Developing an understanding of the application domain and the goals of advanced analytics (data and/or text and/or web mining) process
2.	Acquiring or selecting a target data set
3.	Integrating and checking the data set
4.	Data cleaning, preprocessing, and transformation
5.	Predictive model development (such as decision trees, regression analysis, neural networks) and hypothesis building
6.	Choosing suitable data mining or text mining or Web mining algorithms
7.	Result interpretation and visualization
8.	Result testing and verification
9.	Using and maintaining the discovered knowledge

Data warehousing technology and a multidimensional enterprise data warehouse are keys to successful advanced analytics use in organizations. Data warehousing technologies include the ETL (extraction, transformation, and load) functions to move data in and out of legacy systems and disparate data marts into a comprehensive enterprise data warehouse. Data cleansing may be needed which includes cleaning, standardizing and linking the data as it is loaded from the legacy systems. Data enhancement if needed will involve adding external data such as demographic or spatial information. Customer profitability if needed is the application of identifying historical, current and projected value of the customers and then using it to improve segmentation and to implement customer strategies.

Data, text, or web mining algorithms are chosen and applied to data extracted from the data warehouse to produce predictive models. In general, before this is done, the data is first enriched by adding additional attributes to the data warehouse, which are statistically derived from the data. Generally, a large company possesses a variety of predictive models and scores about customers and their transactions, some produced internally and some purchased from third parties. Next, these different predictive models are analyzed and combined if necessary to produce a single aggregate model.

Advanced analytics applications should easily integrate into information delivery systems so that results can be shared with upper management and decision makers. This type of technology include a wide portfolio of algorithms, visualization techniques, flexible data exploration and manipulation capabilities that produce accurate models. Results should be delivered in a format that even non-technical users can understand – using everyday business terms to explain the results.

Therefore, in the implementation of the above mentioned profiling and segmentation examples, a business analyst might first use clustering to identify relatively homogeneous groups of customers which demonstrate similar buying behavior, that is, segment the customer database. When these segments are demarcated, predictive or statistical models can be developed to forecast their purchase behavior, for example, apply regression to predict buying behavior of each cluster. Each of these groups would then receive product and services relevant to their profile which saves costs of mailing catalogues sent to disinterested customers.

The implementation of advanced analytics could be done within a range of technology infrastructure. Current advanced analytics applications can be executed on a range of systems from mainframe to client/server to PC platforms. The system prices range from several thousand dollars for smaller applications to more than a million dollars for larger. The price is primarily dependent on two factors:

(1) the size of the data warehouse or database; and
(2) the complexity of queries that are executed.

The price is directly proportional to these factors.

Relational database storage and management technology is adequate for many advanced analytics applications which are relatively smaller. Excel, with its advanced data functions, is adequate for small companies and Access for data storage. However, the infrastructure needs to be significantly enhanced, such as SAS, which are expensive to license, to support larger applications.

Two other factors that must be considered when implementing advanced analytics:

(1) The quality of the data must be high if results of advanced analytics are to be used for decision making. This is especially true when data is purchased from external vendors.
(2) The hiring of business analysts to effectively implement and use advanced analytics is quite expensive. Advanced analytics requires a portfolio of skills – in data management, statistical analysis and business decision making – which is hard to find.

7. Managerial implications

Organizations deploying advanced analytics typically use one of the following three development methods:

(1) packaged analytic applications (industry verticals);
(2) analytic application development platforms; or
(3) customized application development. Each method has its advantages and disadvantages.

Packaged analytic applications are relatively easy to use and often allow adopters to secure core functionality without having to reinvent the wheel. In most cases, however, they are unable to support all of an organization's decision making needs – at least not directly off the shelf. Application development platforms allow greater customization and also offer tools and templates that accelerate the application development process. Ultimately, however, the quality of the production application depends on the capability of the developer. The customized "build-from-scratch" approach often yields the most innovative solutions – unconstrained by the functional limitations of development platforms – but can result in redundancy or inconsistency if multiple applications are developed for use in different parts of an organization.

In order to decide between these options and get the most from their investment, managers should assess the decision making environment within their organization. Which business processes would benefit most from decision making assets? How many distinct applications are needed? Will they be deployed to users across the enterprise or concentrated in the hands of a select few? What analytic resources and IT support is available within the organization? How will responsibility for application development be shared by IT and line-of-business stakeholders?

There are usually three stages at deploying advanced analytics technology in an organization. In the first stage, the potential of advanced analytics is discovered. First naïve studies are performed, often by external consultants (which are data mining specialists). Once the profitability of advanced analytics is proven, in the second stage it is used on a regular basis to solve business problems. Users usually are teams of analysis experts (with expertise in advanced analytics technology) and domain experts (with extensive knowledge of the application domain). In the third stage, full exploitation of advanced analytics technology is performed within the organization. End-users are enabled to perform their own analysis according to their individual needs. The necessity of this stage is clearly recognized, although it has not been realized yet.

The different users at these three stages have different demands and they bring different prerequisites. Most of the available tools are aimed at analysis experts, requiring a significant amount of training before being useful to novice end-users. Typical business end-users are for example marketers, engineers or managers. These users are less skilled in complex data analysis and have less knowledge of the nature of the data available, but have a thorough understanding of their occupation domain. Furthermore, they are usually not interested in using advanced analytics technology themselves, but only in getting clear, rapid answers to their everyday business questions. There is a need for user-friendly operational support for both analysis experts and novice users.

End users need simple-to-use tools that efficiently solve their business problems. Existing software packages lack sufficient support for both directing the analysis process and presenting the analysis results in a user-understandable manner. If not, they are restricted to a very limited set of techniques and problems. Ideally, a better usability by novice users would have to be achieved without giving up other desirable features such as flexibility and/or analysis power.

The factors that are typically used to evaluate an advanced analytics software suite include the ability to access a variety of data sources; the ability to access data both online and offline; the underlying data model used such as object-oriented; the maximum number of tables/rows/attributes allowed; the database size the software suite can comfortably handle; the attribute types the software suite can handle such as non numeric; cost; ease of learning; functionality; ease of use and effectiveness. The criterion to assess the ease of learning could be based on whether the software suite had: a demo version, tutorials, user's manual, sample solutions, and online help.

Both SAS (Matigon, 2007; SAS, 2006) and SPSS provide the most complete view of a customer through the combined analysis of structured data in databases, unstructured text, and web usage and survey data. They provide an integrated environment for predictive analytics and descriptive modeling, data mining, text mining, web mining, forecasting, optimization, simulation and more for informed and guided decision making for managers.

Most analysts separate mining software into two groups: mining tools and mining applications. Mining tools provide a number of techniques that can be applied to any business problem. Mining applications, on the other hand, embed techniques inside an application customized to address a specific business problem. Mining tools are used to ensure flexibility and the greatest accuracy possible. Essentially, mining tools increase the effectiveness of mining applications. Since no two organizations or data sets are alike, no single techniques delivers the best results for everyone. Not only do mining tools deliver in-depth techniques, they also deliver flexibility to use combinations of techniques to improve predictive accuracy.

8. Conclusion

Market dynamics – including deregulation, globalization, commoditization and political uncertainty – create an environment where successful companies have to learn the importance of BI and how to turn BI into a competitive advantage.

Advanced analytics, also known as predictive analytics, help turn operational data into strategic information, which is used in decision making to gain competitive advantage. In this paper, the opportunities and challenges presented by advanced analytics that are driven by the three emergent mining technologies – data, text, and web – were investigated to determine their application and managerial benefits.

These three technologies help companies by providing insights that are used to adjust business rules and react to customers in a relevant, personalized manner. Since companies are increasingly conducting their businesses with lesser face-to-face exchanges, getting to know and understand the customer has become even more complicated. The roles of advanced analytics therefore become increasingly significant.

References

Adomavicius, G. and Tuzhilin, A. (2005), "Towards the next generation of recommender systems: a survey of the state-of-the-art and possible extensions", *IEEE Transactions of Knowledge and Data Engineering*, Vol. 17.

Apte, C.V., Hong, S.J., Natarajan, R., Pednault, E.P.D., Tipu, F.A. and Weiss, S.M. (2003), "Data-intensive analytics for predictive modeling", *IBM Journal of Research & Development*, Vol. 47 No. 1, pp. 17-23.

Dorre, J., Gerstl, P. and Seiffert, R. (1999), "Text mining: finding nuggets in mountains of textual data", *Knowledge Discovery and Data Mining*, ACM, San Diego, CA, pp. 398-401.

Emery, P. (2005), "Advanced text mining in ZyIMAGE", *White Paper*, February, available at: www.zylab.com

Fan, W., Wallace, L., Rich, S. and Zhang, Z. (2006), "Tapping the power of text mining", *Communications of the ACM*, Vol. 49 No. 9, pp. 77-82.

Feldman, R. and Sanger, J. (2007), *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, Cambridge University Press, New York, NY.

Froelich, J., Ananyan, S. and Olson, D.L. (2005), "Business intelligence through text mining", *Business Intelligence Journal*, Winter, pp. 43-50.

Hearst, M. (2003), *What Is Text Mining?*, available at: www.sims.berkeley.edu/ ~ hearst/text-mining.html

Jonas, J. and Harper, J. (2006), "Effective counterterrorism and the limited role of predictive data mining", *Policy Analysis*, No. 584, CATO Institute, Washington, DC, pp. 1-11.

Li, J. and Zaiane, O.R. (2004), "Combining usage, content, and structure data to improve web site recommendation", *Proceedings of the 5th International Conference on Electronic Commerce and Web Technologies (EC-WEB)*, pp. 305-15.

Lui, S.M. and Qiu, L. (2007), "Individual privacy and organizational privacy in business analytics", *Proceedings of the 40th Hawaii International Conference on System Sciences*, IEEE Computer Society, Washington, DC, p. 216b, available at: http://dx.doi.org/10.1109/HICSS.2007.268

McCue, C. (2005), "Data mining and predictive analytics: battlespace awareness for the war on terrorism", *Defense Intelligence Journal*, Vol. 13 No. 1 & 2, pp. 47-63.

Marsella, A., Stone, M. and Banks, M. (2005), "Making customer analytics work for you!", *Journal of Targeting, Measurement and Analysis for Marketing*, Vol. 13 No. 4, pp. 299-303.

Matigon, R. (2007), *Data Mining Using SAS Enterprise Miner*, Wiley, New York, NY.

Mobasher, B. (2005), "Web usage mining and personalization", in Singh, M.P. (Ed.), *Practical Handbook of Internet Computing*, CRC Press, Boca Raton, FL.

Mobasher, B., Cooley, R. and Srivastava, J. (2000), "Automatic personalization based on web usage mining", *Communications of the ACM*, Vol. 43 No. 8.

Oliveira, J.P.M., Loh, S., Wives, L.K., Scarinci, R.G., Musa, D.L., Silva, L. and Zambenedetti, C. (2004), "Applying text mining on electronic messages for competitive intelligence", *Proceedings of the 5th International Conference on Electronic Commerce and Web Technologies (EC-Web 2004), Zaragoza, Spain, August 30-September 3*, pp. 1-13.

Rao, R. (2003), "From unstructured data to actionable intelligence", *IT Pro*, November/December, pp. 29-35.

SAS (2006), "SAS Text Miner 3.1 – capitalize on the value hidden in textual information", *FactSheet*, available at: www.sas.com/technologies/analytics/datamining/textminer/

Srivastava, J., Desikan, P. and Kumar, V. (2004), "Web mining – concepts, applications and research directions", *Data Mining: Next Generation Challenges and Future Directions*, MIT, Cambridge, MA.

Volonino, L. (2003), "Electronic evidence and computer forensics", *Communications of the Association for Information Systems*, Vol. 12, pp. 457-68.

Wang, H. and Wang, S. (2008), "A knowledge management approach to data mining process for business intelligence", *Industrial Management & Data Systems*, Vol. 108 No. 5, pp. 622-34.

Weiss, S.M., Indurkhya, N., Zhang, T. and Damerau, F.J. (2005), *Text Mining: Predictive Methods for Analyzing Unstructured Information*, Springer, New York, NY.

Wu, Y.B., Li, Q., Bot, R.S. and Chen, X. (2006), "Finding nuggets in documents: a machine learning approach", *Journal of the American Society for Information Science and Technology*, Vol. 57 No. 6, pp. 740-52.

Corresponding author
Ranjit Bose can be contacted at: bose@mgt.unm.edu