

<http://ansinet.com/itj>

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSI*net*

Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

A Research Study: Using Data Mining in Knowledge Base Business Strategies

¹Girija N. and ²S.K. Srivatsa

¹Faculty Member, ICAFI Business School, ²MIT Campus, Anna University, Chennai, Tamil Nadu, India

INTRODUCTION

The power of Internet technology has heralded an era of information sharing. The availability of huge amount of information does not mean wealth of information. Filtering the data using various mining techniques gives the essence of valuable information called knowledgeable information. The knowledge based information system is a solution to manage the changing demands of the customers. As a result, Business Intelligence (BI) covers the process of transforming data from various data sources into meaningful information that can provide insights into business. It allows improving the decision-making at all levels by giving a consistent, valid and in-depth view of business. Based on the user's needs, there are diverse types of tools to be used for analysis and visualization of information. These tools range from query and reporting to advance analysis by data mining.

Data mining involves use of techniques to find underlying structures and relationships in a large database. Data mining is becoming more and more vital because of the variety of techniques such as neural networks, knowledge discovery, data visualization, fuzzy query analysis and case-based reasoning used in applications and services that can be built using the same.

KNOWLEDGE BASE

Knowledge management as a discipline came to the fore around the year 1994. By the year 1997, researchers realized that this was because knowledge management wasn't about technology; it was about people and human behaviour. The key challenges for knowledge management paradigm, especially for corporate are, global competition, reasonably priced and rapid growth of technology enhancement, increasing demand of the user and retaining market share in the global economy as shown in Fig. 1.

The focal point of knowledge management is to make business more valuable and the need to find ways to ensure that this happens. Knowledge is created and consumed across a wide range of activities such as

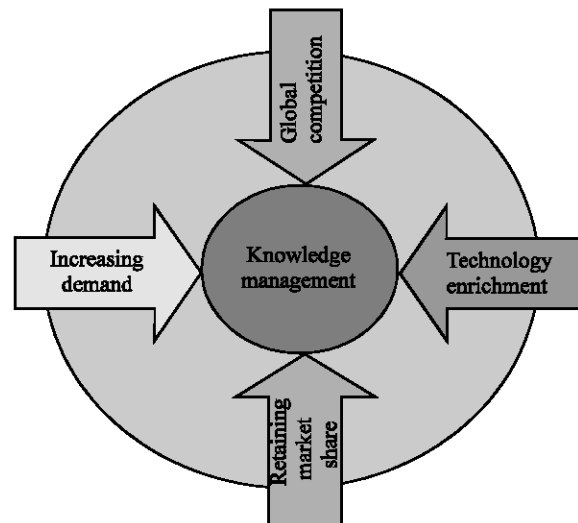


Fig. 1: Knowledge manegement challenges

individual conversations with one another, searches of information in a huge database, just-in-time learning like immediate yes/no ad-hoc query response, a continuous education and highly focused knowledge repositories. Knowledge management promotes and supports each of these activities. Knowledge management spins around the employee knowledge and experience is a vital intellectual capital. Knowledge management uses intellectual assets through knowledge sharing and documentation. There is no exact definition for knowledge management. If knowledge is defined too narrowly and with restrictions, then many opportunities to create and share knowledge will be lost.

Knowledge base is a collection of rules and facts of previous experiences. Knowledge can be classified as declarative and procedural (Grace and Butler, 2005). Declarative knowledge is a description of facts; it is information about real world objects, their properties and relationships among objects. Procedural knowledge includes problem solving strategies, analysis, calculations and inferential knowledge. Procedural knowledge manipulates the declarative knowledge to arrive at new declarative knowledge.

Characteristics of knowledge

The characteristics of knowledge can be stated as follows

Neutrality: Neutrality of knowledge consists of accuracy and reliability. Accuracy means that data or information is correct. Reliability implies that the information is a true indicator of the variable that it is intended to measure.

Accessibility: Is concerned with the availability of knowledge to potential users?

Relevance: Knowledge available to an individual must be appropriate to the task in hand. Its relation to completeness is normally judged in relation to a specific task or decision; all of the material information that is necessary to complete a specific task must be available. In addition, the level of detail or granularity of the information must match that required by the task and the user.

Recent information: Maintenance and periodic updates of the knowledge base also play an important role. Any information has its own life cycle. For decision making purpose, especially data warehouse concepts give key importance to historical information preservation.

Structure and organization: All knowledge has a structure. Structure is important to understanding. This cognitive structure is reflected in the way in which individual's structure information in their communications in the form of verbal statement, text and graphical representations. Two important features of this structure are

- The way in which items are grouped into categories
- The relationships between these categories.

Knowledge management plan: The business goals and business processes are core objective for effective knowledge management. The effective knowledge management creates a new knowledge instead of merely sharing existing knowledge. This creation is helpful when the staff were really needs it. To attain share knowledge effectively with others, knowledge management must have the feature such as information, knowledge artefacts and knowledge content. Knowledge content represents an explicit decision to make knowledge accessible and usable by others. In order to achieve this, knowledge artefacts must be accompanied by a context, which explains why those artefacts were constructed in a particular way. Explicit knowledge content is not the only form in which knowledge can be expressed. The other method such as intelligence assistance is provided by knowledge base to support different decision processes,

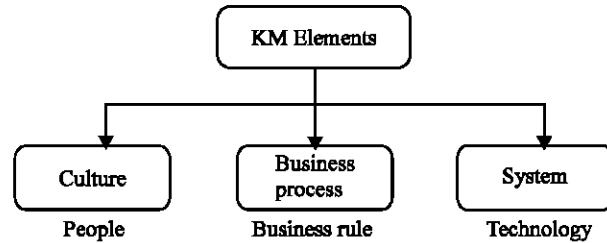


Fig. 2: Knowledge management elements

such as support decision analysis and selection methods. The knowledge base providing guidance in the following processes like forward guidance, backward guidance, preventive guidance, rectification guidance, compulsion-receive guidance and choice-receive guidance. The decision-maker can select any or all guidance mechanisms to complete the decision-making.

Although each company's implementation will be unique, three fundamental elements must be addressed in any knowledge management plan viz. people, business process and system as shown in Fig. 2.

Effective knowledge management may require going beyond initial knowledge capture, to facilitate strategic decisions about how to extend previously captured knowledge. Decision-making is a necessary process in most of the business fields. Quick but precious decisions in many corporations are important to achieve competitive advantages. Nevertheless, these are time-consuming and labour-intensive tasks due to the overwhelming amount of data that are required to be processed and understood to make the necessary decisions. Thus, an automated tool for analyzing and mining large amount of data in order to provide quick and correct decisions is greatly needed. The core focus is Knowledge Discovery in Databases (KDD) through automation of data analysis tasks. KDD assists the strategic decision-making process. Knowledge discovery is defined as "the non-trivial extraction of implicit, unknown and potentially useful information from data." Knowledge discovery process takes the raw results from data mining, which extracts trends and generates patterns from data. Next, KDD carefully and accurately transforms them into useful and understandable information.

A MINING FAMILY ROOT

Classical statistics is the foundation of data mining techniques (Fig. 3). It is mainly used to study data and data relationships in a large database. It includes Regression Analysis, Cluster Analysis and Discriminate Analysis etc. The Classification And Regression Tree (CART) is a tree structured statistical analysis, used to

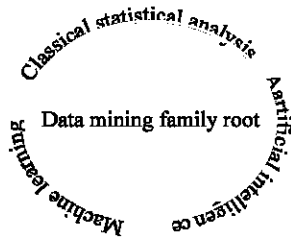


Fig. 3: Data mining family root

generate predictive models which can be applied in profiling customer and targeting the customer, particularly in direct mailing. Multivariate Adaptive Regression Splines (MARS) is a predictive model used to build business intelligence solutions for problems such as forecasting the customer interest (Fig. 4). It is helpful for strategic decision making based on marketing research challenges.

The second family root of data mining is artificial Intelligence or AI. The AI is built upon heuristics. The AI has taken an effort to apply human-thought-like processing to statistical problems. AI uses techniques for writing computer code to represent and manipulate knowledge (<http://www.stottlerhenke.com/ai-general/index.htm>.) which exactly fits in the computer processing in today's business environment. The agent paradigm used Artificial Intelligence (AI) is the mimicking of human through and cognitive processes to automatically resolve complex problems. Agents are software programs. It is capable of autonomous, flexible and focused to analyze and investigate the action in pursuit of objective. In particular, AI is used to predict the customer behaviour in risk management and fraud analysis. The on-line transactions changes the customer purchasing pattern and transaction processes are through credit cards, debit cards or plastic money transactions. On the other hand, they are useful for identifying the customer interest based on each and every transaction of the customer. This customer transaction information is useful for market basket analysis. The result of the analysis gives the pulse of the transaction pattern of the customer and is also useful for planning customer retention (Kimball, 1999). It uses the techniques like Neural Networks, Genetic Algorithms, Fuzzy Logic, Case-Based Reasoning (CBR), Bayesian Networks, Common Sense Reasoning and Cognitive Task Analysis.

The third family root of the data mining technique is machine learning. It can be more accurately described as the union of statistics and artificial intelligence. The advancement of AI could be considered as machine learning. Since, Machine learning has combination of AI heuristics with advanced statistical analysis. The machine

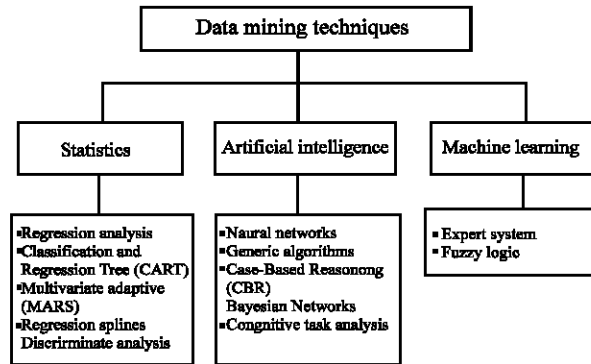


Fig. 4: Various data mining techniques

learning makes an effort to let computer program learn about data study. Based on attributes of data study, the machine makes various decisions. It is possible using base as statistics and the integration of advanced AI heuristics using algorithms to achieve its goals.

Machine learning uses expert systems which are computer programs dedicated to solving problems and giving advice within a specialised area of knowledge. It is used in corporate planning, provision of an automatic customer help service, financial management, chemical analysis, medical diagnosis etc. (Copeland, 2000). The most essential ingredients in any expert systems are "Knowledge Base (KB)" and "Inference Engine". A significant feature of the expert system is that it works mutually with human users and enables human-computer interactions. Constructing an expert system is called Knowledge Engineering. The knowledge base of expert systems contains both factual and heuristic knowledge (Davenport and Prusak, 2000). Factual knowledge's are undisputed facts and heuristic knowledge results from good guessing.

MANY FACETS OF DATA MINING

The individual organization has to decide what information is eligible as intellectual and knowledge based assets. In general, knowledge-based information can be classified into two categories - Explicit or Tacit. The explicit knowledge consists of anything that can be documented, achieved and codified such as patents, trademarks, business plans, marketing research and customer information (Santous and Surmagez, 2005). However, in the case of tacit knowledge, recognition, formulation, sharing and management are real challenges such as email, instant messages, frequent web site visit, customer enquiry.

Knowledge base is a systematic process of finding, selecting, organizing, distilling and presenting information

in a way that helps the organization to gain insight, sustainable competitive advantage and capture markets from its own experience (Milton *et al.*, 1999). These activities entail an organization to resort to dynamic learning, strategic planning and decision making. At the same time, they make organizations to realize their intellectual assets and enterprise intelligence and enable increased flexibility.

Data mining is a synonym of the term Knowledge Discovery in Database (KDD). Data mining is a step in the knowledge discovery process and interacts with knowledge base (Hand *et al.*, 2005). It is the most significant cutting edge in database system and also promises interdisciplinary information needed by industry such as marketing, health care, geographical information and library information system. The data mining techniques are applied commonly to almost all information based subject areas wherein there exist huge amount of data for identification of unknown or hidden information. Therefore, data mining techniques used in WWW are called web mining. They are called multimedia mining when used in multimedia database, called biblio mining when used in digital libraries, spatial mining when used for maps and text mining when used only for text (Fig. 5).

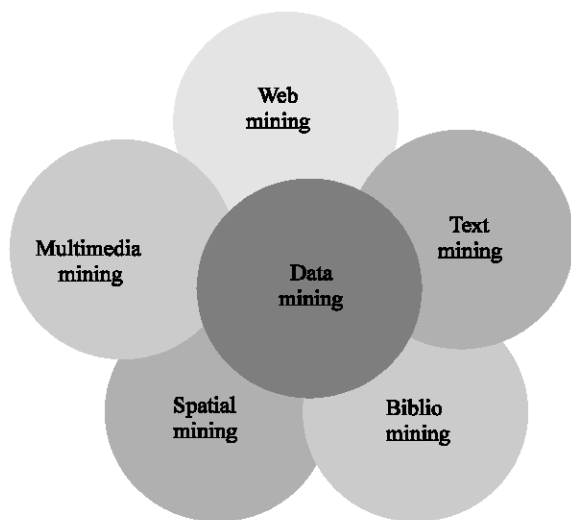


Fig. 5: Various facets of data mining

Web mining: The World Wide Web with amazing growth revolutionizes the business strategy called on-line Business or E-Business. One of the great opportunities of web interactivity is that it allows the marketers to reach precisely targeted customers and to build a relationship with them. It is possible only with right techniques for mining the huge amount of distributed information available in the WWW. The significant task for Web

mining can be listed out as Information Retrieval, Information Selection/Extraction, Generalization and Analysis. Web mining normally uses two types of approaches namely Agent Based Approach and Database Approach and broadly classifies into Web content mining, Web structure mining and Web usage mining (Kosala and Blockeel, 2000)

Web content mining: Web Content Mining is the process of extracting useful information from the contents of Web documents. It may consist of text, images, audio, video, or structured records such as lists and tables. Text mining and its applications such as topic discovery, extracting association patterns, clustering of web documents and classification of web pages. It uses Information Retrieval (IR) and Natural Language Processing (NLP) techniques.

Web structure mining: Web structure consists of Web pages as nodes and hyperlinks as edges connecting two related pages. Web Structure Mining is the process of discovering structure information from the Web. Based on web structure data, the mining is further used in hyperlinks and document structures. A hyperlink is a structure unit that connects a web page to different locations and is called an Intra-Document hyperlink and within the same page, is called an Inter-Document Hyperlink. Document structure focuses on automatic extraction tree structured format or Document Object Model of the tags within the page.

Web usage mining: Web Usage Mining is the application of data mining techniques to discover interesting usage patterns from Web data and identify Web users along with their browsing behaviour at a Web site. Web server data analyze the Web log entry in the Web server and collect the personalized customer information such as IP address, page references and access time of the users. It provides meaningful information and knowledge to an organization. Application server data are ability to track various kinds of business events in the E-commerce based application and log them in the application server logs.

Web mining applications mainly concentrate on Personalized Shopping, Web-wide tracking, On-line communities or user groups, Personalized Auctions, Personalized portal, Web user behaviour characterization and process the nature of Web traffic and Web resource usage, on-line based advertisement, web based business intelligence, Target Marketing, Campaign Analysis, Fraud Detection, Insurance Claims Analysis and Customer information service.

Text mining: Text mining is the process of analyzing and structuring large sets of documents. It uses the statistical or computational linguistics technologies. Text mining techniques can range from simple like arithmetic average and intermediate complexity like linear regression, clustering and extremely complex neural networks. The broader area of text mining includes Information Retrieval (IR), Information Extraction (IE) and computational Linguistics. Effective information retrieval is measured based on number of relevant documents retrieved from text database (Koenig, 1985). Relevant documents are in the form of input query such as keywords or example documents. There are two basic measures for assessing the quality of text retrieval (Han and Kember, 2001) - Precision and Recall. Precision is the percentage of retrieved documents that are in fact relevant to the query (i.e. "exact" response) and recall is the percentage of documents that are relevant to the query and were, in fact retrieved.

$$\text{Precision} = \frac{|\{\text{Relevant}\} \cap \{\text{Retrieved}\}|}{|\{\text{Retrieved}\}|}$$

$$\text{Recall} = \frac{|\{\text{Relevant}\} \cap \{\text{Retrieved}\}|}{|\{\text{Relevant}\}|}$$

Most of the information retrieval systems support keyword-based and similarity-based retrieval. In keyword based information retrieval, a document is represented by a string, which can be identified by a set of keywords. A user provides a keyword or an expression formed out of a set of keywords such as "Dr. S.R. Ranganathan and five library laws". The similarity based retrieval systems find similar documents based on set of common keywords. The output of such retrieval should be based on the degree of relevance, where relevance is measured based on the closeness of the keywords, the relative frequency of the keywords and so on.

The primary goal of information retrieval has to index the text and search for useful documents in a collection (Harter and Hert, 1997). It uses the index methods called latent semantic index method, inverted indices and signature files. The Latent Semantic Index method uses the technique which helps in extracting the hidden semantic structure from the text rather than just the usage of the term. There are significant techniques used for mining the text database and these are called association analysis, documents classification analysis and clustering techniques. The association analysis collects sets of keywords or terms that occur frequently together and

then finds the association or correlation relationship among them. The document classification method classifies documents based on set of associated, frequently occurring text patterns. In particular, corporate environment are highly competitive. Search engines serve the general population of Web users. Hence, it is unlikely that a company's Web page would link to a Web page of its competitor. Search engines are powered by advanced indexing algorithms, which make use of automated programs, called robots or crawlers, to continuously visit large parts of the Web in an exhaustive fashion. It is a challenging task to mining topic-specific knowledge on the web. A topical crawler that is aware of such domain level characteristic may utilize it to its advantage. Topical crawlers, also known as topic driven, focused, or preferential crawlers, are an important class of crawler programs that complement search engines. It helps in extracting a small but focused document collection from the Web that can then be thoroughly mined for appropriate information using text mining, indexing and ranking tools (Pennock *et al.*, 2002). It also aids the user to learn in-depth knowledge of a topic systematically on the Web. It supports applications such as specialized Web portals, online searching and competitive intelligence. There are improved agent methods available such as adaptive agent, meta search engine and comparison agent. Adaptive agent is not just searching the information. At the same time, it uses the genetic algorithm to improve the quality of every search by filtering important words in the preferred documents. Meta search engine selects the search engines for a particular query and post-processing is applied to the results. Comparison agent monitors specific resources or sites to collect data and uses heuristics to imply relations among the data stored in a knowledge base thus increasing the utility of an information resource.

The major aim of Information Extraction (IE) to make use the IR system to transform the information that is more easily understandable and analyzed from huge collection of documents. The key distinction of IE and IR is, IE extracts only relevant facts from the document and IR selects only relevant document from the collection. In the text mining process, the information extraction taking role after the information retrieval step and before data mining techniques are performed. IE can also used to improve the indexing process, which is part of the IR process. There are two types of IE – IE from unstructured texts and IE from semi-structured data. The root of IE is Natural Language Processing (NLP). Indeed, IE could be called a core language technology.

Computational linguistics computes statistics over large text collections in order to discover useful patterns.

These patterns are used to inform algorithms for various sub-problems such as syntactic analysis, semantic analysis and discourse analysis within natural language processing. The major objective of the computational linguistics is to discover patterns to aid other problems within the same domain, whereas text mining is aimed at discovering unknown information for different applications.

Spatial mining: The geographic locations and distribution of customers are critical information that is usually missing from Customer Relationship Marketing (CRM) (Nicholson, 2003). Spatial mining is the branch of data mining that deals with spatial i.e., geo-referenced data such as graphs and maps. It refers to the extraction of knowledge, spatial relationships and interesting patterns not explicitly stored in the spatial database. Spatial data mining tasks can be grouped into description, exploration and prediction. To understand the data, spatial data and spatial phenomena have to be first described and analyzed. Hidden patterns and relationships among spatial or non-spatial variables have to be explored. Based on the current pattern of spatial distribution and the understanding of spatial relationships, future state and trend of the spatial pattern and spatial distribution can be predicted. Spatial Mining is useful for retail marketing promotion. Obviously, users tend to purchase close to home or work, since it is convenient. It provides the models that recognize spatial correlation and use spatial inference such as geocoded customer information for increasing promotional activity based on customer utility. The knowledge discovery tasks involving spatial data include finding characteristic rules, discriminant rules, association rules or deviation and evolution rules, etc. The key task for spatial data mining is the problem of determining aggregate proximity relationships. This problem is concerned with relationships between spatial clusters based on spatial and non-spatial attributes. Expert knowledge is applied for defining spatial hierarchy and setting thresholds at different levels for both spatial and non-spatial data abstractions. GIS has a long history of being used as a tool to visualize spatial, attribute and statistical data. Such uses of GIS include choropleth mapping for filling uniform colour or pattern in each spatial unit, dasymetric mapping technique is a useful solution for mapping population density relative land-cover and trend surface analysis as a mathematical method of analyzing data in map form. The availability of functions such as spatial and non-spatial query and selection, classification, map overlay, network analysis and map creation, make GIS a useful tool for spatial data mining. Visualization through GIS gives the

user the ability to spot errors that can otherwise be unnoticed by analyzing raw data and to aid visual analysis and detection of certain features distribution and their patterns.

Multimedia mining: The digitization technology is used for satellite images and remote sensing data. Methods for mining the images, digital objects, texture, colour, comparison, shape are called Multimedia Mining. It mines the WWW text markups, linkages and interactive audio-video database in the Internet database. Traditional data mining techniques have been developed mainly for structured data types. The image data type does not belong to this structured category. Content of an image is visual in nature and the interpretation of the information conveyed by an image is mainly subjective, based on the human visual system. Most of the activities in mining image data have been in the search and retrieval of images based on the analysis of similarity of a query image or its feature(s) with the entries in the image database. The image retrieval systems can be broadly categorized into two searches use either description of an image or its visual content.

The descriptive based images are user-defined texts. The images are indexed and retrieved based on these rudimentary descriptions, such as their size, type, date and time of capture, identity of owner, keywords or some text description of the image. This method is also called Text-Based Image Retrieval. In the visual based method, image are searched and retrieved based on the visual content of the images. The visual contents such as colour, texture, pattern, image topology, shape of objects, their layouts and locations within the image, etc. are retrieved. These images features can be extracted and used as index or basis of search. This method is also called Content Based Image Retrieval (CBIR).

The content based image retrieval uses three steps for retrieving the images. They are - visual content or feature extraction, multidimensional indexing and retrieval. Colour histogram has been found to be very effective in characterizing the global distribution of colours in an image and it can be used as an important feature of image characterization. To define colour histograms, the colour space is quantized into a finite number of discrete levels. Each of these levels becomes a bin in the histogram.

The size, shape, colour and orientation of the textons can vary over the region. The difference between two textures can be in the degree of variation of the textons. It can also be due to spatial statistical distribution of the textons in an image. Texture is an innate property of virtually all surfaces such as bricks, fabrics, woods, papers, carpets, clouds, trees, lands, skins, etc. It contains

important information regarding underlying structural arrangement of the surfaces in an image. Shape is essential to segment the image to detect object or region boundaries. Techniques for shape characterization can be divided into boundary-based, using the outer contour of the shape of an object and region-based, using the whole shape region of the object. The topological property of a digital image is known as Euler number. The Euler number is usually computed in a binary image. The Euler number is defined as the difference between number of connected components and number of holes in a binary image. Euler numbers have strong discriminatory power because a digital image may readily be distinguished from other digital images in its class. This implies that the Euler number may be used for more efficient searching or matching of digital images. This technique is useful for medical diagnosis such as the detection of malaria infected cells which are often different from that of a normal cell. Multidimensional indexing is an important component of content-based image retrieval. Before indexing, it is very important to reduce the dimensionality of the feature vector and to find an efficient data structure for indexing and to find suitable similarity measures.

The content based image retrieval techniques can be extended, in principle, to video retrieval systems. A video is not only a sequence of pictures, it represents the actions and events in a chronological order to convey a story and represent moving visual information. Each video clip can be considered as a sequence of individual still pictures and each individual frame can then be indexed and stored using the traditional content-based image retrieval techniques. The indexed feature data and the corresponding key frames are stored in the metadata database. In a video retrieval system, the query processing depends upon the query image; it is then matched with each and every key frame stored in the metadata database. The matching technique is repeated, based on the extraction of different features from the query image, followed by the matching of these features with the stored features of the key frames in the database.

Biblio mining: Data mining not only mines the information which is in the structured format in database, but also uses the technique in semi-structured data like email message, electronic publication in the digital library web pages. Text retrieval is a key focus of the text mining. Based on the similarity of text, it retrieves documents from the database. This is further enhanced and when used mainly for digital library, is called biblio mining. Biblio mining (Nicholson, 2005) deliberately supports Information Retrieval (IR), which has focused on user query and transaction processing of structured digital

library database. The biblio mining is the combination of data mining, biblio metrics, statistics and reporting tools used to track patterns in authorship, citation and extract patterns of behaviour. It could be expected that digital libraries would develop techniques to forecast and respond dynamically to changing user preferences over time; however, to evaluate how well a digital library responds to the needs and preferences of its user community requires that such preferences can be accurately and ultimately, dynamically determined. Although the evaluation of digital library collections can be approached from a number of perspectives, an assessment of the nature of digital library user communities should be an essential part of any digital library evaluation strategy.

In biblio mining, the data mining applications like web mining, text mining and spatial mining are used to retrieve various digital objects like text, images and links. Especially text mining plays a vital role in keyword-based and similarity-based text retrieval from digital library database. Similarity-based retrieval finds similar documents based on a set of common keywords. The digital objects have heterogeneous structure which makes it difficult to categorize, filter or interpret documents. Therefore, a more intelligent tool is needed for information retrieval such as intelligent agents. These intelligent agents have been developed that search for relevant information using domain characteristics and user profiles to organize and interpret the discovered information especially handling knowledge query to respond to the intelligent query answer. The technology used in the intelligent agent search engines are crawler, indexer and ranker. The crawler is called an agent to automatically scan and collect documents based on the query. From the links, the pattern is created and other related documents are found. To visit the link, the documents are indexed by depth first and breadth first approach. The ranker performs the ranking function and hence determines the order in which the documents must be displayed to the user. These technologies are using various algorithms like page rank algorithm, Fuzzy c-Means clustering algorithm.

DATA MINING IN MARKETING BUSINESS STRATEGY

Customers consistently demand increasingly innovative products and services, delivered more quickly and cost effectively. The only way for an organization to respond to the increased complexities of these demands is by maximising the efforts of its most experienced and capable employees. Effective knowledge management implementations will accelerate how rapidly knowledge

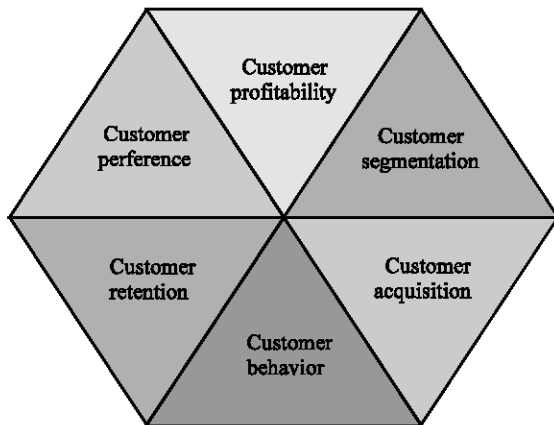


Fig. 6: Customer relationship elements

can be brought into the business process environment, accelerating the pace of innovation and delivery. A strategic plan revolves around a set of management decisions about what the organization should do to be successful. The strategic planning process - defining the organization's vision and mission through the framework of long-range goals - provides a platform to identify the much-needed resources in capital and manpower.

The business values are measured by single sales model, to discover the most profitable customer and maintaining customer loyalty (Berson *et al.*, 2000). This model can be improved through understanding the influencing customer behaviour and meaningful communication. As a result, customer service and support retain and improve the customer retention, customer loyalty and customer profitability.

Data mining uses well-defined statistical and machine learning techniques to construct the models that predict the customer behaviour. Recent technology using AI and machine learning enhances the automated mining process. Integration of the customer information in the business (using data warehouse) gives a rewarding business strategy. It revolutionises the role of data in marketing and transfers the customer data from management to the hands of business analyst and users. The market analysis is not an easy task; it should give understandable picture as to the market size, customer behaviour details of competitors and future trends.

The web technology opens a new opportunity for visionary companies; it has a set of new standards for responsiveness, quality and getting individual customer information. The knowledge about customer information as shown in Fig. 6 is precious and is the fundamental foundation for business value. While competition is increasing, there is need to better understand customers

and too quickly respond to their individual needs and wants. Every time a web visitor clicks a link to see another page, several pieces of data are stored in the web server's log files (Zaiane *et al.*, 1998). This data can be used to build and maintain databases that provide valuable marketing opportunities. Web usage mining deals with this type of database stored in the log files for determining the customer behaviour. The general term for this information is called "Click Stream Data" (Kimball, 2000). Web usage mining technique interact and collect the information with web server access logs, proxy server logs, browser logs, user profiles, user queries and user transactions etc. (Zaiane *et al.*, 1998). It primarily concentrates on the customer navigation patterns. While browsing or shopping in an online environment, a customer has typically several different types of input fields for interacting with the underlying system. Text fields allow the input of keywords and choices allow the selection of static or dynamic predefined values of an attribute. This customized usage tracking helps to build data mining model using response behaviours (Berson *et al.*, 2000). Consequently, the idea behind is to move closer to individual customer based service in reality and also predict the profitable customer.

Individual identification of user behaviours and customer personalization are very complex. The ideal strategy to satisfy and move closer to customers is customer segmentation. The key objective was to enrich the knowledge about individual customers leading to new strategic customer segments. Customer segmentation is the process that segments customers into smaller groups called segments. Segments are homogeneous within and desirably heterogeneous in between. The segmentation could be categorized as demographic or psychographic. The psychographic can be easily identified through similar interest of the user using web usage mining behaviour. Using spatial data mining, one can analyze the demographic segmentation based on data using Geographical Information System. It enriches the identification of customer segments using clustering. The customer demographics are measured by time spent in the last visit, frequency of the visit and average transaction value. Clustering is a computerized technique in which objects with demographic, geographic, psychographic and behavioural properties are grouped together. This process is equivalent to that of customer segmentation. So clustering is widely used for customer segmentation purposes. Clustering based on neural network known as Self Organizing Maps (SOM) (Mitra *et al.*, 2003) is used mainly for clustering large data set. SOM is how our brains process information and recognize patterns and features. It works well with both numerical and categorical

variables. There is no need for data transformation and weight calculations. More importantly, it works well with customer data without problematic data transformations. Clustering is also very important in multimedia data mining, in order to cluster similar multimedia contents together for efficient indexing and storage in multimedia database. It is useful in on-line based product promotion activity. The key feature of on-line shopping is based on multimedia based product display and demonstration. A business analyst determines the optimal follow-up activities for every customer based on the relevant customer segmentation combinations. Industry experts are predicting that web enabled self-service applications will dramatically improve service and reduce cost. The gathering of Internet information about customers allows messages to be finely targeted via email and web based ads. Using specially designed customer intelligence service package enables the most profitable customers and also addresses the customer's business needs directly. It provides the company with new revenue opportunities. This spatial mining helps to construct a model like customer loyalty segmentation model, competition analysis model and target media model to reach the customer. The core idea of spatial mining is to identify the opportunity of product promotional activities based on the segmentation containing customers with high profit potential. Predictive web analytics tools automatically discover segments and profiling brings the leading predictive techniques based on the actual customer behaviour recorded in the log files data.

The growth of business involves the acquisition of new customers. Gaining enough knowledge and awareness about the business of the particular customer and targeting the customer is very important. Business Intelligence focuses on delivering the strategic view to the user in an easy-to-use, up-to-date format that can be employed immediately by all users to support business decisions (Geiwitz, 2001). Probabilistic techniques like Bayesian Network may be used in a large database to diagnose the target customer planning. On the other hand, retaining customers is most significant for business, whether a profitable customer or not. As per marketing strategy there is nothing more expensive than acquiring a new customer and nothing cheaper than maintaining a good customer. This challenge can be tackled by building the best way to improve customer loyalty. It will lead to Customer Life Time Value Model.

The customer loyalty needs intelligent dealing and meaningful communication, especially in Web environment. Maintaining a constant and healthy relationship with the customers has found its importance in today's businesses, since loyal customers drive profits

and sustainability of the business. Marketing strategies are designed based on an analysis on the needs of the customers and their loyalty levels. While need analysis is done before the process of segmenting and targeting, loyalty levels are analyzed by understanding the purchase behaviour process of the customers. The loyalty captures the customers response based on what customer knows about the seller and what the seller knows about the customer interest. The reason for giving importance to loyalty effect is that loyal customers tend to spend more, cost less to serve and refer others. The core business strategy of customer loyalty is to acquire, grow and retain profitable customer relationships with the goal of creating a sustainable competitive advantage. It is possible by identifying the customer preference. The majority of the customer preferences are receiving a service that better matches their needs. Cross-selling is the process that provides new offers, products and services related to existing purchasing behaviour of the customer. It can be done by marketing optimization. The marketing optimization based on the customer preference, gives the best possible offers which is real profit to the customer and also to corporate to increase the sales. The important business activities are product recommendation in attracting customers. On-line helps the company to acquire customers preferences and recommend products accordingly on a one-to-one basis in real time and more importantly, at a much lower cost. This type of system is called recommendation system and it can be categorized as content-based and collaborative system. The content based system provides recommendations by matching customer interests with product attributes. The collaborative system utilizes the overlap of preference ratings among customers for product recommendation. The content-based system uses machine learning method called Support Vector Machine (SVM) by analyzing the relationships between the preference ratings and the corresponding product features. Another technique called naïve Bayes classifier determines using probabilistically independent preferences and recommendations by applying rule-based systems.

Market basket analysis is useful type of data analysis for marketing. It determines what products customers purchase together; it takes its name from the idea of customers throwing all their purchases into a shopping cart i.e., Market Basket. Once an enterprise understands the customers who buy one product are likely to buy another, it can market the products together or make the purchasers of one product the target prospects for another. It provides a distinct advantage in the highly competitive and often opportunistic world of consumer behaviour. This strategy facilitates impulse buying and

helps ensure that customers who would buy a product do not forget to buy it simply because they did not see it. It is also useful for product promotion. Contact with the existing customers with information about new products shown to sell well with the products they have already bought will probably ensure that they will be interested. Market basket analysis uses the data mining technique such as Association Rules. The Clustering for Association Discovery (CLASD) (Aggarwal, 2002) algorithm has been effective in discovering localized associations, from individual segments, that expose a customer pattern which is more specific than the aggregate behavior.

Customer information is a domain for any business. The effective information retrieval plays a vital role in taking business decisions. Most of the information is interactive based like image, video and picture. Digitization has not only changed the quality of the information, has also invoked the technology for information preservation. The recent semantic web environment revised the term information into Object. The biblio mining need never be a special for digital library purpose alone. It also uses intelligent agent for retrieving the knowledge needs for business which helps to develop and maintain a competitive, cutting edge in workforce (Zaiane *et al.*, 1998).

CONCLUSIONS AND RECOMMENDATION

The increasing focus on globalization, competition has become worldwide. Therefore, the knowledge base concepts and activities have the potential to make a significant impact in corporate environments. It has raised the awareness of customer information issues in organization to compete effectively in the knowledge economy to change their values and establish new objectives on creating and using intellectual assets. This organizational transformation is being further accelerated with the techniques called "Data Mining". The foremost technique of data mining is to suit any form of information. In knowledge base, using data mining approaches are now moving from statistical and probabilistic to visualization approach. It further extends multi-paradigmatic called "Hybrid Approach", which combines more than one data mining approach to achieve the desired knowledge base. The multi-agent is using this approach. It is especially effective and efficient information retrieval with good user query response. Information extraction concentrates on developing intelligent agent based languages like OWL, OIL and SHOE etc. At the same time, research also needs to be carried out to develop techniques for analyzing how

digital library user usage behaviour evolves over time. From an experimental human behaviourist's viewpoint, the Web is the perfect experimental apparatus. Research needs to be done in developing the right set of Web metrics and their measurement procedures, so that various Web phenomena can be studied. (Srivastava *et al.*, 2005).

The issues in the technology domain relate mainly to the issue of integration of technologies and infrastructures across geographical and organizational boundaries. Virtual Teams (VTs) enable organizations to extend their reach to remote markets, bring dispersed talent from different organizations together and transcend geographic and cultural boundaries to work on challenging projects and enhance customer service. The major important aspect is communication and however, is only one of the major factors determining VT. In future, collaboration tools are another invention, which are designed to overcome communication difficulties among VT members. The significant challenge of these tools is to support multilingual service, global coverage, cost savings and quality.

Forthcoming business environment require more than technology. It should support more knowledge centric service. The maturity and evolution of outsourcing strategies is leading business to shift the off shoring of high-end processes as Knowledge Process Outsourcing (KPO). The KPO sector is likely to grow globally to face challenges such as patent and technology landscaping, patent claim mapping, intellectual property monitoring, patent mining and administration type of service to provide the organization with tools to meet global competition. The other challenges for KPO are to provide unique value proposition service integrating process delivery across the end customer life cycle. This can be called integrated customer life cycle management service.

REFERENCES

- Aggarwal, C.C., 2002. Towards effective and interpretable data mining by visual interaction, ACM SIGKDD Explorations Newslett., 3: 11-12.
- Berson, A., S. Smith and K. Thearling, 2000. Building Data Mining Applications for CRM. Tata McGraw-Hill Publishing Company Ltd., New Delhi.
- Copeland, J., 2000. Expert systems, http://www.alanturing.net/turing_archive/pages/Reference%20Articles/what_is_AI/What%20is%20AI07.html.
- Davenport, T.H. and L. Prusak, 2000. Working Knowledge How Organizations Manage What They Know. Harvard Business School Press, USA.

- Geiwitz, R., 2001. White paper: Business Intelligence, www.billinmon.com.
- Grace, A. and T. Butler, 2005. Learning Management systems: A new beginning in the management of learning and knowledge *Intl. J. Knowledge and Learning*, 1: 12-24.
- Han, J. and J. Kamber, 2001. *Data Mining Concepts and Techniques*. Morgan Kaufmann Publishers, USA.
- Hand, D.H. Mannila and P. Smyth, 2005. *Principles of Data Mining*. Prentice-Hall of India Private Ltd, New Delhi.
- Harter, S.P. and C.A. Hert, 1997. Evaluation of information retrieval systems: Approaches, issues and methods, *Ann. Rev. Inform. Sci. Technol.*, 32: 3-94. http://www.stottlerhenke.com/ai_general/index.htm.
- Kimball, R., 1999. The market basket data mart, intelligent enterprises magazine, Vol. 2, No. 6, http://www.intelligententerprise.com/db_area/archives/1999/992004/warehouse.jhtml.
- Kimball, R., 2000. The special dimensions of the clickstream intelligent enterprises. Magazine Vol. 3, No. 2, <http://www.intelligententerprise.com/000120/webhouse.shtml>.
- Koenig, M.E.D., 1985. Biblio graphic information retrieval systems and database management systems. *Information Technology and Libraries*, 4: 247-272.
- Kosala, R. and H. Blockeel, 2000. 'Web mining research: A Survey. *ACM SIGKDD*, 2: 1-14.
- Milton, N.N. Shadbolt, H. Cottam and M. Hammersley, 1999. 'Towards a knowledge technology for knowledge management. *Intl. J. Human-Computer Studies*, 51: 615-641.
- Mitra, S. and T. Acharya, 2003. *Data Mining - Multimedia, Soft Computing and Bioinformatics*, John Wiley and Sons, Inc., Hoboken, New Jersey.
- Nicholson, S., 2003. The Basis for Biblio mining: Frameworks for Bringing together usage-based data mining and biblio metrics through data warehousing in digital library services. *Information processing and Management*, <http://dlist.sir.arizona.edu/886/>.
- Pennock, D.M., G.W. Flake, S. Lawrence, E.J. Glover and C.L. Giles, 2002. Winners don't take all characterizing the competition for links on the web. *Proceedings of the National Academy of Sciences*, 99: 5207-5211.
- Santous, M. and J. Surmacz, 2005. The ABC of knowledge management. Knowledge Management Research Center. <http://www.cio.com/research/knowledge/edit/kmabcs.html>.
- Srivastava, J., P. Desikan, and V. Kumar, 2005. Web Mining -Accomplishments and future directions. <http://www.ieee.org.ar/downloads/Srivastava-tut-paper.pdf>.
- Zaiane, O.R., M. Xin, and J. Han, 1998. Discovering web access patterns and trends by trends by applying OLAP and data mining technology on Web log. *IEEE International Forum on Research and Technology Advances in Digital Libraries*, Los Alamitos, CA: IEEE Computer Society, pp: 19-29.