International Journal of Parallel and Distributed Systems
Volume 1, Issue 1, 2011, pp-01-04
Available online at: http://www.bioinfo.in/contents.php?id=107

Knowledge and Data Mining in Decision Making Process: A Business Model

[1]S.B. Jagtap, [1]H.S. Fadewar and [2]G.N. Shinde
[1]*Swami Vivekanad Institute of Technology and Management, Udgir Dist. Latur(M.S.)*
[2]*Indira Gandhi Mahavidyalaya, CIDCO, New Nanded (M.S.)*

Abstract—This paper highlights the importance of tools that can be used during the decisionmaking process, for example, Data Mining (DM) and Knowledge Data Discovery (KDD). The relevance of these technologies is closely linked to the culture of the organization implementing them and to a profound understanding of the ways by which decision-makers turn data into information and information into knowledge. These are the principles that govern data mining.

In this paper we propose a decision making model for business estimation.

Keywords: Data Mining, Business Modeling, Goals Strategies, Decision-Making.

I. INTRODUCTION

Decisions can be defined as activities we perform everyday. Activities related to our daily life are routine actions and they are relatively well defined. However, really important decisions -the strategic ones- can be difficult to make, as they need a lot of information and, consequently, tools supporting the decision-making process.

Decision making tools make it possible to complement people's intellectual resources and the computers 'capacity of improving decision quality, especially when address non-structured problems. Problems are non-structured when their objectives come into conflict, their possible solutions are difficult to be identified and the selection of an alternative implies a high level of uncertainty.

Decision-makers should not be concerned about tools design but, on the contrary, they should be concerned about their effective and strategic application to improve the quality of identification of problems and their solution. For that purposes, tools must have at least the following characteristics:

- They can be employed in non-structural decision contexts.
- They support decision-makers but do not replace them.
- They support all the phases of the decision-making process.
- They are focused on the effectiveness of the decision-making process.
- They are under the decision-maker's control.
- They use data and models.
- They facilitate the decision-maker's learning process.
- They are interactive and friendly.
- They use an evolutionary and repetitive process.
- They provide support from the highest executive level to the line levels.
- They support multiple decisions, whether independent or interdependent.

Support tools are expected to serve at least to expand people's capacities of processing a great volume of information during the decision-making process. Even though many decision components are structured, they can be very complex and time-demanding for the decision-maker. In that event, the support tool should serve to free decision-makers, so that they can use that time and their knowledge to work on that part of the decision that is non-structured. Support tools can only incorporate the knowledge of their designers and (which is quite worthy) serve to process a limited set of abilities, but they can also "think" or "learn", as analyzed below.

However, the creation and incorporation of new knowledge by means of data mining and KDD occur in the cultural environment characteristic of each organization. This organizational culture shapes the way people ask themselves questions about their work, generating new answers. We will analyze the concept of organizational culture and a brief example of the way it affects the selection of tools to support decisions.

II. ORGANIZATIONS FRAMEWRK

Taking into account that people work within the framework of their organization, we can consider the organizational culture on the basis of the following characteristics:

- The degree of responsibility, freedom and independence of the individuals belonging to the organization.
- The degree in which employees are persuaded to be aggressive, innovative and risk-takers.
- The degree to which the organization creates clear objectives and development expectations.
- The degree to which the different organization units operate in a coordinated way.
- The degree to which managers provide clear information, assistance and support to their employees.

- The extent of regulations to control employees' behavior.
- The degree of identification with the organization as a whole.
- The degree of development of the employee in relation to his/her performance.
- The degree to which employees are persuaded to exhibit their conflicts and
- express their criticism.
- The degree to which organizational communication is restricted.

Both the individual style and the organizational environment affect decision quality. Consequently, the influence of organizational culture on the type of decision to be taken and on the convenience of applying certain decision strategies and tools must be seriously taken into account when implementing decision-support tools. For example, an organization oriented to collaboration and team work and engaged in a CRM (Customer Relationship Management) effort will prefer the implementation of a data mining system which offers the general characteristics of a standard system as well as the tools to incorporate workgroup management in their projects and other tools for collaboration work that make it possible to design large scale systems for the whole organization.

The culture of organizations, as well as culture in general, evolves together with new knowledge that is, at the same time, cause and effect of cultural transformations. Before going further, we would like to study in depth the process of knowledge and its limitations: because, beyond any kind of support tool that can be used, success depends on knowledge about the process applied by decision-makers and, especially, on how this process is integrated within the organization's framework. In the following section, we will analyze the way elements such as data and information are combined by knowledge to generate a summary of concepts that give sense to decisions support systems.

III. Data, Information and Knowledge

We live surrounded by reports on the importance of knowledge and information to obtain economic advantages in a globalized world. We could write many pages trying to define the term "knowledge", but we will only provide some brief ideas on the meaning of the terms "data", "information" and "knowledge".

A. Data

They refer to facts, measures or bservations that can occur (or not) in a given context. data without context are: 60, 62, 66, 72. The same data now contextualized could mean Laura's, Ana's, Juan's and Pedro's weight in kilograms, respectively. Data validity and effectiveness are mainly determined by their accuracy.

B. Information

It refers to data organized in a certain way, so as to be useful and relevant for the decision-maker who has to solve a problem. The key criterion to evaluate information is its usefulness.

C. Knowledge

It refers to a combination of instincts, ideas, rules, processes and information applied by decision-makers to guide their actions and decisions. Knowledge is our mind's interpretation of things, which is valid as long as the interactions of a problem with its context can be properly explained. From the above definitions, two conclusions can be inferred. The first one is that information is personal and the second one is that knowledge in not static: even more, it should change when the decision context changes.

IV. Data Mining

Data mining is a set of activities performed to find new, hidden and unexpected contexts in the data. Using information contained in a data warehouse, data mining can answer the queries that a decision-maker would not ask unless she/he had those tools.

The term Knowledge Data Discovery (KDD) is increasingly being used as a synonym for data mining. It is a more descriptive term and can be applied to all the activities and processes related the discovering of useful knowledge from the data. Using a combination of techniques - including statistic analysis, neuronal logic, diffuse logic, multidimensional analysis, data visualization and intelligent agents- the KDD can discover useful patterns to develop models that can predict behaviors or consequences, in a large variety of knowledge spheres.

A few years ago, decision support tools were based on the concept of *verificatio*n. A relational database could be used to obtain dynamic answers for properly asked queries. The key was in the previous knowledge that the person had, which could be extended and verified by the search result. The concept of verification was getting outdated. As there was an increasing need to *discove*r, the demand of designed techniques such as KDD is also increasing to discover new and non-classified behavioral patterns in the data.

The basic categories of data mining techniques currently used can be classified in: *classification, association, sequence and cluster*.

A. Classification

It includes data mining processes that look for rules to define if an item or event belongs to a determined

International Journal of Parallel and Distributed Systems
Volume 1, Issue 1, 2011, pp-01-04
Available online at: http://www.bioinfo.in/contents.php?id=107

s*ubset* or to a class of data. This technique, probably the most commonly used, includes two sub-processes: the elaboration of a model and the prediction. In general terms, classification methods develop a model composed of IF-THEN rules and are perfectly applied, for example, to find purchase patterns in customers databases and build roadmaps which link customers attributes to those of the purchased products. With an appropriate set of predictive attributes, the model can identify those customers who are likely to purchase a determined product during the following month. A typical case of classification is dividing a database of companies into homogenous groups working on variables such as "credit possibilities" and values such as "good" and "bad".

B. Association

It includes techniques known as linkage analysis, used to identify the patterns of operational transactions with high probabilities of repetition, as it happens when analyzing a basket in the search of similar products. An associative algorithm is developed and it includes the rules used to correlate a series of events with another one. For example, a supermarket could require information about the purchase habits of its customers to place the products generally bought together, place the new products in the best place, or offer promotions.

C. Sequence

Time series analysis used to relate events to time. For example, as a result of this kind of model we can learn that people who rent a video film tend to purchase promotional products during the following weeks, or that a microwave oven is often purchased after determined previous purchases.

D. Cluster

It is frequently difficult or impossible to define the parameters of a class of data. In that case, clustering methods can be used to create partitions so that their members are similar to one another, according to a determined metric system. Cluster analysis could be used, among other applications, to study the products bought by credit card, to discover -let's say- that those products bought by a business golden card are purchased during the days of the week and the average cost is 152 pesos, while the same kind of product bought by personal platinum card is commonly purchased at the weekends and the average cost is lower but includes a bottle of wine in more than the 65 % of the cases.

As explained, questions that can be asked are endless in the data mining world and the methodologies used to answer them are diverse and increasingly varied. Just as there are many techniques to support the data mining process, there are different technologies to build models, as mentioned below.

V. ESTIMATING AND BUILDING THE MODEL

This process includes four parts:

- Select data mining task,
- Select data mining method,
- Sselect suitable algorithm
- Extract knowledge as can be seen in Figure .1

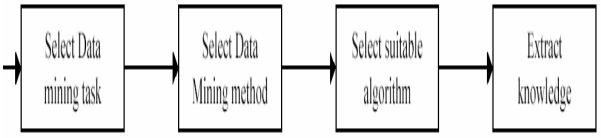

Fig. 1: Estimating and Building Blocks

Figure1 show that this process is divided into four parts these are:

A. Select Data mining task (s)

Selecting which task to use depends on the model whether it is predictive or descriptive.predictive models predict the values of data using known results and/or information found large data sets, historical data, or using some variables are tasks for predictive model. A descriptive model identifies patterns or relationships in data and serves as a way to explore the properties of the data examined. Clustering, summarization, association rules and sequence discovery are usually viewed as descriptive. The relative importance of prediction and description for particular data mining applications can vary considerably. That means selecting which task to use depends on the model whether it is predictive or descriptive.

B. Select Data mining method (s)

After selecting which task we can choose the method and assuming we have a predictive model and the task is classification while the method is Rule Induction, with Decision tree or Neural Network. In most research in this area; researchers estimates the relevant model this model to produce acceptable results. There are number of methods for model estimation includes these but not limited to neural networks, Decision trees, Association Rules, Genetic algorithms, Cluster Detection, Fuzzy Logic.

C. Select suitable algorithm

The next step is to construct a specific algorithm that implements the general methods. All data mining algorithm include three primary components these are:

(1) Model representation, (2) model evaluation, and (3) search.

D. Extracting Knowledge

This is the last step in building the model which is the results (or the answers for the problem solved in data mining) after making the simulation for the algorithm. This can be best explained by presenting an example of Auction Fraud [20] or fields in the data set to predict unknown, classification, regressions, time series analysis, prediction, or estimation

VI. STATISTICAL ANALYSIS

Let's suppose that 70% of the people who buy product X by credit card, also buy product Y and, in addition, product Y is never sold independently. Therefore, it is quite easy to build a model to predict the purchase of product Y with a 70% probability of occurrence. It will be certainly much more interesting to predict the purchase of product X. For these purposes, data mining needs statistical techniques capable of dealing with non-linear data, multiple outliers (data rarely far from the average) and non-numerical data, like data that can be found in a data warehouse. Lineal regression techniques, commonly used in many applications, cannot always be used in data mining due to the complexity of the data patterns and the lack of linearity.

Neuronal networks, genetic algorithms and diffuse logic: neuronal networks are mathematical structures capable of learning and are developed to show, by means of equations, the way the human brain recognizes data and information patterns. Thus, it is possible to develop non-linear predictive models which learn by combining variables and studying the way they affect the data. The techniques of learning machines, among which genetic algorithms and diffuse logic are included, can convey meaning to complex and inaccurate data and they can also extract patterns and detect tendencies. Genetic algorithms combine the notions of survival of the fittest and a structured and random exchange of characteristics among the individuals of a group of possible solutions in order to make up a search algorithm that can be applied to solve optimization problems in different fields. By imitating the mechanism of biological evolution, genetic algorithms work on a group of possible solutions to a problem. Each element of the group is referred to as a "chromosome". A chromosome represents, within the genetic algorithm, a possible solution to a problem. The way chromosomes code the solution is referred to as "representation".

Diffuse logic is a kind of logic where variables can have several levels of truth or falseness represented by ranks of values between 1 (true) and 0 (false). The result of an operation can simply be expressed as a probability and not necessarily as a certainty. For example, apart from the values "true" and "false", a result can adopt values such as "probably true", "possibly true", "possibly false" and "probably false". Diffuse logic is a mathematical approach to deal with the inaccurate nature of everyday language and the surrounding world. It is commonly used, for instance, to select human resources, where one has to select the candidate who meets a greater number of the qualities required on the basis of the best adequacy coefficient.

VII. CONCLUSION

In Knowledge society, people compete at speeds unknown by all preceding societies. Universal access to knowledge makes it possible for everyone to have the possibility of improving organizations' performance.

In this paper we proposed framework for decision making model for taking the effective decisions retrieving the knowledge and information. In this framework we apply the information mining techniques for building the model.

REFERENCES

[1] Berry, Michael J. A. and G. Linoff, 1997, Data Mining Techniques: for Marketing, Sales, and Customer Support, John Wiley and Sons.
[2] Halkidi, "Quality assessment and Uncertainty Handling in Data Mining Process"htttp://www.edbt2000. unikonstanz.de/phd-workshop/papers/ Halkidi.pdf.
[3] Fayyad, U. M., G. P. Shapiro, P. Smyth. "From Data Mining to Knowledge Discovery in Databases", 0738-4602-1996,AI Magazine (Fall 1996): 37–53.
[4] Jiawei Han, Micheline Kamber. "Data Mining: Concepts and Techniques", Morgan Kaufmann Publishers, Champaign: CS497JH, fall 2001, www.cs.sfu.ca/~han/DM_Book.html.
[5] Bowers, D. G., and S. L. Seashore, 1966, "Predicting Organizational Effectiveness with a Four Factor Theory of Leadership", Administrative Science Quarterly, 11 (September).
[6] Clemen, R. T., 1991, Making Hard Decisions: An Introduction to Decision Analysis, Belmont, Duxbury Press.
[7] Culman, M. J. and B. Gutek, 1989, "Why Organizations Collect and Store Information" in Stohr, E. A. and B.R Konsynski, Information Systems and Decision Processes, Los Alamitos, IEEE Computer Society Press.
[8] Fayyad, Usama, Gregory Piatesky-Shapiro and Padhraic Smyth, 1996, "The KDD Process for Extracting Useful Knowledge for Volumes of Data", Communications of the ACM, pp. 27-34.
[9] Golub, A. L., 1977 Decision Analysis: An Integrated Approach, New York,Wiley.
[10] Gray, P., 1997, "The New DSS: Data Warehouse, OLAP, MDD and KDD", in : Proceedings of the Americas Conference on Information Systems, edited by Carey, Jane M., Phoenix, Association for Information Systems.
[11] Harrison, E. F., 1995, The Managerial Decision-Marketing Process, Boston, Houghton Mifflin Co.
[12] Inmon, W. H. J. D. Welch and K. L. Glassey, 1997, Managing the Data Warehouse, New York, Wiley.
[13] Watson, H. J., G. Houdeshel and R. K. Rainer, 1997, Building Executive
[14] Information Systems and Other Decision Support Applications, New York, Wiley.
[15] Yolis, E., P. Britos, J. Sicre, A. Servetto, R. García-Martínez and G.
[16] Perichinsky, 2003, "Algoritmos Genéticos Aplicados a la Categorización Automática de Documentos", ITBA.