

Data Mining and Management Decisions

Antonella Reitano, Fabrizio Di Maio, Salvatore Seminara

Abstract: *The information systems management sets the attention to the importance of data and above all the activities of selection to individualise this data. The decisive lines to the aforesaid problem can be individualised in the activities of Data Mining (DM). DM procedures can be implemented through different techniques and formalities. Among all, the most properly employed in the solution of business problem list are: Neuronal Nets, techniques of Case-Based Reasoning (CBR), Intelligent Agents (IA).*

Key words: *Data Mining, Information Systems' Management, Neuronal Nets, web site, intelligent agents, Case Based Reasoning.*

1. Techniques of Data Mining: evolutionary trials and peculiarity.

The information systems management sets the attention to the importance of data and on activities of selection that are finalized to the individualization of those data

Currently, even rather simple business monitoring system can bring a huge quantity of data to the attention of enterprise managers among which not all is of critical importance, even though the monitoring systems tend to underline the more useful data. Therefore, there is a need to develop a synthesis process that underlines the critical informations that can aid business decision making. To this end, it is useful to build:

- a data warehouse - a data store that contains information which is extracted by other systems used by the firm and make it accessible to business customers
- various data mart - a specialised data warehouse which is used by the department or a group of business customers for special requirements

The information, apart of being easily accessible it also has to be useful. In between data acquisition and data use, there is an interval process, which allows error correction, essential data extraction, as well as extraction of permanent data from provisional one.

The monitoring process of internal and external activities of an enterprise has as a result the development of enormous information database. The information extraction is a rather complex¹ activity and at the same time of great interest.

The way to solve the problem stated above is to follow Data Mining (DM) activities, or rather "the efficient discovery of valuable, nonobvious information from a large collection of data"². This brief definition expresses accurately the DM meaning, which emphasises two terms: 'discovery' and 'efficient.'

¹ "Il existe une différence aussi grande entre les données et les informations qu'entre les Montagnes Rocheuses et un lingot d'or pur à 99,99%. Nous savons tous que les Montagnes Rocheuses recèlent de l'or, mais nous ne savons pas exactement où. On consacre beaucoup de temps et d'efforts à trouver ce minéral d'or, clef de richesse. Il en va de même des données", v. A. AINSLIE, X. DREZE, *Le data-mining et l'alternative modèles classiques/réseaux neuronaux*, in *Décisions Marketing*, n. 7/1996, pag. 77.

² R. J. BRACHMAN, T. KHABAZA, W. KLOESGEN, G. PIATETSKY-SHAPIRO, E. SIMOUDIS, *Mining BusinessDatabases*, in *Communications of the ACM*, vol. 39, n. 11/1996, pagg. 42-48; J.P. BIGUS, *Data Mining with Neural Networks*, McGraw-Hill, U.S.A., 1996, pag. 9.

In the above definition, discovery states that the DM procedures focus on the automated detection of facts and relationships existing among data. The main idea that underlies this concept is that the large quantity of raw data presented by the enterprise is analysed by the DM algorithm in order to extract valid information. A strict analogy of this is the activity of gold and/or oil extraction.

DM is implemented through six different activities³, all based on the extraction of meaningful information:

- classification
- estimation
- forecasting
- grouping for affinity or associative rules
- clustering
- description and visualization

The process of DM is articulated through three principal phases. The first one consists of the preparation of the data, or rather the selection, the purification and the pre-elaboration under an expert's guidance. The second phase consists of further data elaboration, data compression and transformation, to extraction of valid information easier, even though it is not directly noticeable. The third phase consists of the analysis, which is in the evaluation of the DM output, with the purpose to verify the knowledge domain and determine the importance of the facts produced with the use of the DM algorithms.

The DM procedures can be implemented through different techniques and formalities. Among all, the most appropriately elaborated in resolving business problems are: Neuronal Nets, techniques of Case-Based Reasoning (CBR), Intelligent Agents (IA).

2. Neuronal Nets

The Neuronal Nets or Neural⁴, simulate the operation of human brain - mainly the storage and the elaboration of information – and have the tendency to reproduce decisional making.

The Neuronal Nets are applied to different fields of business and research:

- market segmentation⁵
- consumer analysis and behaviour prediction⁶ and therefore sales prediction
- pricing strategy⁷

³ M. J. A. BERRY, G. S. LINOFF, *Mastering Data Mining*, Wiley, U.S.A., 2000, pagg.8-11.

⁴ L. FU, *Knowledge Discovery Based on Neural Networks*, in *Communication of the ACM*, vol. 42, n. 11/1999, pagg. 47–50.

⁵ D.S. BOONE, M. ROHEM, *Retail segmentation using artificial neural networks*, in *International Journal of Research in Marketing*, n.19/2002, pagg. 287-301; R.J. Kuo, L.M. Ho, C.M. Hu, *Integration of self-organizing feature map and K-means algorithm for market segmentation*, in *Computers & Operations Research*, n. 29/2002, pagg. 1475-1493.

⁶ P.M. WEST, P.L. BROCKETT, L.L. GOLDEN, *A Comparative Analysis of Neural Networks and Statistical Methods for Predicting Consumer Choice*, in *Marketing Science*, vol. 16, n. 4/1997, pagg. 370–371.

- fraud prevention
- scheduling and routing in the manufacturer sector and in transport
- financial analysis
- character recognition and text recovery

The Neuronal Nets appear beneficial, particularly in situations where vast databases of examples are available, where even experts cannot use specific rules to combine the immense amount of input data. Neuronal Nets can execute operations usually performed by human brain. Such tasks are important for human survival and involve the simultaneous elaboration of vast amount of data (sight, hearing, touch) where fast and accurate pattern recognition is required.

3. Case-Based Reasoning (CBR)

The Case-Based Reasoning (CBR)⁸ is a method for decision-making, based on comparing the current situation to past similar cases in which the decision has to be obtained.

The CBR systems store therefore numerous cases related to the matters considered. When a new situation is delineated the system identifies the characteristics of the same situation and receives a list of similar cases that have been analysed in the past. Often the past cases produce important indications that can answer the question directly. In case that a particular problem is resolved, the system adds it as an additional example to the database. In such a way, a CBR system, which is initially empty, or includes limited cases, can gradually grow larger and become a precious resource.

The prospects of the CBR systems are proportional to the size of the database. Even though such systems can work with database of modest dimensions, good results can be reached when the database contains large amount of examples.

The CBR⁹ systems are classified as:

- analytical systems
- synthetic systems

The first ones are used for classification, diagnosis and support decision-making. The last ones are used for planning, shaping and making plans. The CBR systems can secure an intelligent support for problem solving and can stimulate innovation in the domain of management.

In addition, the CBR systems can support the implementation of the followings cases:

- problem solving, based on the problem description and on managerial intuition

⁷ R.J. KUO, P. WU, C.P. WANG, *An intelligent sales forecasting system through integration of artificial neural networks and fuzzy neural networks with fuzzy weight elimination*, in *Neural networks*, n 15/2002, pagg. 909-925.

⁸ P.A. PHILLIPS, F. DAVIES, L. MOUTINHO, *Assessing the Impact of Market-Focused and Price-Based Strategies on performance: A Neural Network Typology*, in *Journal of Market-Focused Management*, n.5/2002, pagg. 219-238.

⁹ M. MEHDI OWRANG O, *Case discovery in case-based reasoning systems*, in *Information systems management*, Winter 1999, pagg. 74-78.

¹⁰ BO SUN, LI DA XU, XUEMIN PEI, HUAIZU LI, *Scenario-based knowledge representation in case-based reasoning systems*, in *Expert Systems*, vol. 20, n. 2/May 2003, pagg. 92-99.

- prediction and estimation, based on the description of the observed object certain peculiarities of the object can be estimated and predictions can be made about relative phenomena
- simulation, based on the existing situations and on the decisions that have to be taken to identify the possible consequences
- decision making assistance, consisting in recovering, modifying and reconstructing cases that support choosing amongst alternatives
- offering services, regarding education, computer science, telematics and knowledge acquisition

4. Intelligent Agents

The Intelligent Agents (IAs) represent one of the most recent evolutions of Artificial intelligence and therefore a field of research where knowledge is not yet consolidated. There are different points of view even about definitions currently acceptable¹⁰. Amongst them two are the most relevant:

- an IA is a software that assists users in implementing numerous actions. It can automate repetitive tasks, remember events, develop the addition of complex data, implement learning processes
- the IAs execute, continuously, three functions: perception of environmental dynamic conditions; actions that have impact on environment; reasoning development in order to interpret perceptions, solve problems, trace inferences and determine actions

The IAs¹¹ represent a spin-off of neuronal technology. Generally, an IA consists of a sensible element able to receive results regarding certain happenings, or to recognise, or classify what determines them.

A primary aspect of IAs is the concept of authority delegation. In this case, the user commits the responsibility and the burden of some time consuming duty to specialised software. Considering such a delegation, the user is free to perform other activities and even to disconnect from the computer while the software agent laboriously completes the job that has been allocated with. The user does not need to understand the operations the computer performs.

According to some researcher, the characteristics of Intelligent Agents can be grouped into two categories internal and external.

Amongst internal characteristics the following are worth of consideration:

- autonomy in working or performing research alone
- reactivity in receiving data and providing feedback
- goal driving, that is the provision of detailed knowledge so that goals can be achieved
- awareness of peripheral activities in order to act proactively
- mobility which allows the navigation of agents through telematic architectures
- continuity, so that an IA continues to monitor the environment after having achieved its objectives

In addition, amongst external characteristics the following are worth of consideration:

¹⁰ H. C. DESOUZA, *Intelligent Agents for Competitive Intelligence: Survey of Applications*, in *Competitive Intelligence Review*, vol. 12/2001, pagg. 57-63.

¹¹ J.P.BIGUS, *Data Mining with Neural Networks*, cit., pagg. 115–128.

- cooperation, a group of agents work together to achieve a common goal, defines a multi-agent system
- communication, the agents have to be able to communicate with other agents and/or the user

Considering their abilities (and more often considering their performance) it is possible to identify different categories of IAs; among all, the ones that are closely related to DM are:

- the filtering agents
- the information agents
- the user interface agents
- the office agents, or work flow agents
- the interface agents
- the DM agents

The last category of IA specialise in assisting enterprises not to be overwhelmed by large amount of data.

Amongst the IA applications “shopbot” can included which enables an efficient and smooth online shopping. The shopping agents provide extra information, such as technical news about products. The shopbot can help customers to compare best price-quality combination. From their point of view, enterprises use shopbot to develop business strategies, such as the monitoring of competitors’ prices, with the main objective to improve their knowledge.

5. Web sites and DM

Relevant information can be obtained through research focusing on observing businesses web sites. Other ways of obtaining information include sending email to a predetermined target group. Other information could be extracted “observing” persons who visit web sites. From this kind of analysis, it is possible to extract a lot in terms of different variables, such as, search habits, preferences, consultation frequency and connection time. This could be particularly interesting in developing communication strategies that aim at a specific target group of web site visitors¹². Other digital tools for extracting information through internet¹³ are: yearbooks, robots, discussion forums.

There are also other advanced tools for data extraction that are as follows:

- web site aspirators, that allow to copy, integrally, a web site so that it could be analysed successively (see Teleport Pro (<http://www.tenmax.com/teleport/pro/home.htm>); www.map.com e www.cartia.com)
- meta-engines, that allow to interrogate, simultaneously, dozen of search engines with the purpose to have a better web coverage and consequently a more exhaustive data extraction (see the following examples Copernic, www.copernic.com and www.webseeker.com)
- semantic tools of analysis, for analysing texts and/or discussions, index, content (for example Zoom; Ethnos (www.acetic.fr) etc.)

¹² For example, a 60 years old woman that visits a car business web site, would like to have information about security and price while a young person is, generally, more interested in velocity.

¹³ B. MARTINET, Y.-M. MARTI, *L'intelligence économique. Comment donner de la valeur concurrentielle à l'information*, Éditions d'Organisation, Paris, 2001.

6. Areas of application

The DM techniques are widely and effectively used in different business areas, such as:

- distribution

In distribution, DM techniques are principally used to develop basket analysis, the sales prevision, the consumer's profile and planning and collocation of goods.

- the bank sector

In the bank sector, DM techniques can be used for identification of consumers' segments of distributors and buyers, the setting better price strategies and therefore better revenues, identification of frauds and prevision management of every customer's life cycle.

- telecommunications

In telecommunication business, DM techniques can be used for knowledge discovery technique in order to implement records analysis and to analyse customer's fidelity.

The application of DM techniques provide economic advantages consequent to cost reduction and maximization of revenue. Concluding, the application of DM techniques improve the decision making process and reduce the uncertainty in taking decisions.

REFERENCES

- Ainslie Andrew, Dreze Xavier, *Le data-mining et l'alternative modèles classiques/réseaux neuronaux*, in *Décisions Marketing*, n. 7/1996
- Assens Christophe, *Réseau neuronal et réseaux d'entreprises*, in *Revue française de gestion*, mars-avril-mai 1997
- Berry Michael J.A., Linoff Gordon S., *Mastering Data Mining*, Wiley, U.S.A., 2000
- Bigus Joseph P., *Data Mining with Neural Networks*, McGraw-Hill, U.S.A., 1996
- Boone Derrick S., Rohem Michelle, *Retail segmentation using artificial neural networks*, in *International Journal of Research in Marketing*, n.19/2002
- Brachman Ronald J., Kabaza Tom, Kloesgen Willy, Piatetsky-Shapiro Gregory, Simoudis Evangelos, *Mining Business Databases*, in *Communications of the ACM*, vol. 39, n. 11/1996
- Brondoni Silvio (a cura di), *Il sistema delle risorse immateriali d'impresa: cultura d'impresa, sistema informativo e patrimonio di marca*, Giappichelli, Torino, 2004
- Cohen Marcel, Stathis Costas, *Strategic change stemming from e-commerce: implications of multi-agent systems on the supply chain*, in *Strategic change*, May 2001
- Costes Yseulis, *Comprendre et mesurer le profil et le comportement des internautes*, in *Revue française du marketing*, n. 177-178/2000
- Desouza Hevin C., *Intelligent Agents for Competitive Intelligence: Survey of Applications*, in *Competitive Intelligence Review*, vol. 12/2001
- Kuo R. J., Ho L. M., Hu C. M., *Integration of self-organizing feature map and K-means algorithm for market segmentation*, in *Computers & Operations Research*, n. 29/2002
- Kuo R.J., Wu P., Wang C. P., *An intelligent sales forecasting system through integration of artificial neural networks and fuzzy neural networks with fuzzy weight elimination*, in *Neural networks*, n. 15/2002
- M. Mehdi Owrang O., *Case discovery in case-based reasoning systems*, in *Information systems management*, Winter 1999
- Martinet Bruno, Marti Yves-Michel, *L'intelligence économique. Comment donner de la valeur concurrentielle à l'information*, Éditions d'Organisation, Paris, 2001

- Phillips Paul A., Davies F., Moutinho L., *Assessing the Impact of Market-Focused and Price-Based Strategies on performance: A Neural Network Typology*, in *Journal of Market-Focused Management*, n. 5 /2002
- Reitano Antonella, *Marketing decision support system: dati, informazioni, interconnessioni di reti telematiche*, in *Economia e commercio*, n. 2/1996
- Reitano Antonella, *Telematica marketing strategico e business intelligence*, in *Onde*, n. 2/2001
- Reitano Antonella, *Telematica, concorrenza e comunicazione integrata d'impresa*, Giappichelli, Torino, 2003
- Rygielski Chris, Wang Jyun-Cheng, Yen David C., *Data Mining techniques for customer relationship management*, in *Technology in Society*, n.24/2002
- Smith Kate A., Gupta Jatinder N.D., *Neural networks in business*, Idea Group, Hershey, 2002
- Sun Bo, Xu Li Da, Pei Xuemin, Li Huaizu, *Scenario-based knowledge representation in case-based reasoning systems*, in *Expert Systems*, vol.20, n.2/May 2003
- Thieme Jeffrey R., Song Michael, Calantone Roger J., *Artificial Neural Network decision Support Systems for New Product Development Project Selection*, in *Journal of Marketing Research*, vol. XXXVII, November 2000
- West Patricia M., Brockett Patrick L., Golden Linda L., *A Comparative Analysis of Neural Networks and Statistical Methods for Predicting Consumer Choice*, in *Marketing Science*, vol. 16, n. 4/1997
- Wind Jerry, *Digital marketing*, in *Symphonya. Emerging Issues in Management*, n.1/2002
- Wind Jerry, Mahajan Vijay (ed.), *Digital Marketing*, ETAS, Milano, 2002
- www.acetic.fr
- www.arisem.com
- www.cartia.com
- www.copernic.com
- groups.google.com
- www.iee.org.uk/publish/inspec/docs/document
- www.map.com
- www.tenmax.com/teleport/pro/home.htm
- www.thesaurus.com
- www.webseeker.com

ABOUT THE AUTHORS

Antonella Reitano, Associate Professor, Faculty of Engineering, Department of Business Administration, University of Calabria, Phone: ++39 0984 49 2270, E-mail: a.reitano@unical.it.

Fabrizio Di Maio, Engineer, Assistant, Information and Telematics Area, University of Calabria, Phone: ++39 0984 49 6891, E-mail: f.dimaio@unical.it

Salvatore Seminara, Dott., Assistant, University of Calabria, Phone: ++39 0984 49 2270, E-mail: sseminara@tiscali.it