FROM HARD DATA TO SOFT DECISION

Dr. Zoltán BARACSKAI

Associate Professor at the Budapest University of Technology and Economics, Faculty of Social and Economics Sciences, Department of Industrial Management and Business Economics.
E-mail: baracskai@imvt.bme.hu, Tel.: +36(1) 463-1092, Fax: +36(1) 463-1606
Address: H-1111, Budapest, Mûegyetem rkp. 9. bld. T.

Dr. Géza CHIKÁN

Senior Geologist at the Hungarian Geological Survey, Geological Institute of Hungary.
E-mail: chikan@mafi.hu, Tel.: +36(1) 251-6878, Fax: +36(1) 251-0703
Address: H-1143, Budapest, Stefánia út 14.

Viktor DÖRFLER

Assistant Professor at the Budapest University of Technology and Economics, Faculty of Social and Economics Sciences, Department of Industrial Management and Business Economics.
E-mail: dorfler@imvt.bme.hu, Tel.: +36(1) 463-2438 Fax: +36(1) 463-1606
Address: H-1111, Budapest, Mûegyetem rkp. 9. bld. T.

Jolán VELENCEI

Scientific Researcher at the Budapest University of Technology and Economics, Faculty of Social and Economics Sciences, Department of Industrial Management and Business Economics.
E-mail: velencei@imvt.bme.hu, Tel.: +36(1) 463-1092 Fax: +36(1) 463-1606
Address: H-1111, Budapest, Mûegyetem rkp. 9. bld. T.

Abstract

It is impossible to create model of decision process, as we know nothing about the original decision process. Although it is possible to build models that can get us to the spaces where our fitness is strong enough. These models can contain hard data and soft information as well.

In the background of the widely accepted solutions there are transformations of soft information into hard data. This leads us to the world of quantitative decision support. This step is very dangerous! The decision maker uses logic not arithmetic in his thinking process.

DoctuS© Knowledge-Based System uses logic. The latest version is also capable of data mining. Using a cluster-analyzing algorithm it can transform the relations between hard data into soft information, which will be used for deduction in reasoning. The number of clusters is given by the user. The cluster-analyzing algorithm makes the clusters using learning example. When running the data mining the clusters remains unchanged and the new data will be transformed. The clusters can be handled using logic.

For illustration we use an example of taking decision about location for a power plant.

The required knowledge

Nevertheless, today's programs with the epithet of intelligence are stupid. It is simple to allude to upcoming artificial intelligence. Knowledge-based systems with inductive thinking and genetic programming possibilities are on standby for a while now. Of course there are partial attempts, which are good starting points.

„Technology is on standby – the problem lies somewhere else. People can't break away from the thought that teaching requires a talking head and hairy arms".[i]

[i] Interactive Week '00. okt. 9.

In future we can expect bigger and faster changes in business knowledge. Fast changes are challenging future experts to life long learning. Now we don't know (and it is impossible to know) what will we need to know in future. The world advances toward the life long learning model, which gradually extrudes the present narrow, concentrated learning model.

The new keystones for knowledge increase:

Customized (body-tailored) – body-tailored knowledge acquisition means that a "tailor" has to get to know the "body". It is very important to get to know the tailor too. For the web mediators it is of crucial importance to observe one thirsty for knowledge. Plural, in this case doesn't exist. Instead of concentrating on average mass, we should keep an eye on each one separately. Perpetual learning causes perpetual body-tailoring. Future requirements will differ from the present ones.

Fresh – It is important for knowledge increasing mediators to mediate the freshest knowledge and to assume them before the competitors do. False knowledge is more dangerous than ignorance. And the web is a perfect place for phantomic pipe dreams armed with fresh dates.

Who-to-learn-from – In knowledge-based organizations the conventional relationships between generations has radically changed. Expertise does not lie anymore just in heads of elder, experienced co-workers, but young people too who are open to new technologies and solutions. This can especially cause problems if co-workers have to learn new things from each other. The web provided learning (doesn't claim personal presence) neutralizes communicator's personal style, thereby it helps in softening the prejudice towards adopting youth values or solutions. Hereby we set up such a learning field where people focus on essential elements. It becomes irrelevant who provides the idea.

Couple of hypothesis was taken. Essential is the belief that the knowledge increase supported by the web will differ from American training types, either in the case studies discussions or in the long distance learning:

- In future, the majority of training programs will content deep knowledge (oriented to problem exploration and solution).
- A "surplus teacher" is needed who can teach the knowledge thirsty ones the fundamental concepts, which are the basic terms in "the world of keywords".
- After training we should not collect the standard facts of yesterday, but current data to support our decisions (**data mining**).
- Relationship between data is meaningful (soft knowledge). This knowledge is stored in expert's long-term memory (knowledge bases). In case we need it, we can browse expert studies (**knowledge angling**).
- Benchmarking examines searches for the leading practice. Basic rule is: "don't copy". It is of crucial importance to understand and then to judge the current situation (**experience fishing**).

Knowledge increase

Organization cannot learn, but a man can and he learns differently within different organizations. Organizational culture defines the acceptable knowledge. Let's presume an open corporation.

Searching is based on keywords. Today the only ones who are able to inquire the web are those who know the keywords (basic concepts). We can select keywords from the existing knowledge depot.

According to Polanyi the knowledge can be divided into the focal and into the background knowledge. (Polányi, 1994)[1] The focal knowledge is what we are currently concentrating on. E.g. while reading, the meaning of the text is in the focus. The background knowledge is used in parallel but it remains unconscious. E.g. while reading the knowledge of letters.

Russel classified the knowledge by its origin, which can be divided into the knowledge acquired with or without the mediation of words. (Russel, 1961)[2] Without words we personally perceive or we will perceive in immediate future. All other knowledge is gained by the mediation of words.

Polanyi's famous classification divides the knowledge into codified and tacit. (Polányi, 1997)[3] These are not disjunctive categories. Once we learned a grammar rule we are able to put this knowledge into words. Later we forget the definition but while writing we use the rule perfectly.

Ryle divides knowledge into "knowing how" and "knowing that". (Ryle, 1999)[4] The same categories appear at Anderson as procedural knowledge and declarative or descriptive knowledge. (Anderson,1983)[5] It is essential that "knowing what to do" is included in "knowing how" not in "knowing that".

Investigating "knowing what to do" Minsky concludes that positive knowledge (knowing what to do) differs from negative knowledge (knowing what not to do). Both are essential. (Minsky, 1994)[6]

Minsky distinguishes the special knowledge from the common sense. (Minsky, 1982)[7] Common sense is out of our domain.

For the organization the knowledge can vary between valuable and hazardous.[i] The person may evaluate the knowledge from promising to ominous. (Table 1) If a particular new knowledge is valuable for the organization and promising for the person (no. 1 in the table) then it will be accepted. If it is hazardous and ominous (no. 4 in the table) then it will not be accepted. So, no. 1 and 4 are stable. What about no. 2 and 3? They are instable, so they have to move to no. 1 or 4.

		Organization	
		Valuable	Hazardous
Person	Promising	①	②
	Ominous	③	④

Table 1: Judgment of the new knowledge

The final judgment will be made by interaction of the value systems both personal and organizational. The two value systems affect each other by influence. Using norms the organization can influence the person to change his judgment. The person can influence the organization with attitude. Resistance works opposite with influence. No. 4 can also be changed but it requires third party influence. If the third party is master he can influence the person with attitude, so the no. 4 will move to no. 2. As no.2 is instable the person will try to influence the organization. If the third party is environment of the organization, it can use both norm (e.g. standards) and attitude (e.g. all the competitors did it). In this case the resistance of the person is neglected.

Knowledge creation

If we understand the point of the web then we can use the interactivity to comprehend the new knowledge on the basis of Linux (free software) development conception.[8] The easiest way to understand is to distinct free pub talk from free beer. Corporate University does not give away knowledge free, but we can freely talk about it or even modify it.

This idea is based on following:

- Anybody can freely access the knowledge shaped on the desk.
- Anybody can freely customize (body-tailor) the knowledge shaped on the desk.
- Anybody can freely spread either body-tailored either the desk shaped knowledge.

[i] These two extremes are used as examples anything in-between is possible.

Couple of hypothesis was taken. Their starting point is in the belief that knowledge creation supported by the web differs from forum chats.

- While body-tailoring the knowledge maker can get a hold of anything from shaped knowledge.
- While body-tailoring the knowledge maker can swap anything from shaped knowledge.
- While body-tailoring the knowledge maker can add anything to an already shaped knowledge.

Building a knowledge base

The first step is the acquisition of attributes. Cases in Doctus KBS[9] are characterized by attributes. Attributes are actually points of view, in connection with which we have expectations in relation to cases. The definition of attributes essentially provides a measurement of the knowledge base.

The primary condition of rule based reasoning is that attributes should be subordinated to each other with given rules. The deductive graph describes the dependency relations of attributes. Nodes depend on nodes (factor attributes) connected from below. Deduction is, technically, a matter of upward tracking on the graph. The top of the graph is the final conclusion. On the basis of subordination, two types of attribute can be distinguished: attributes depending from other attributes are called dependant attributes. The values of these are determined by rules given by user. Attributes which do not depend on anything are called input attributes, or independent attributes. The values of these attributes are given by user.

The second step is acquisition of rules. A rule determines the value of its attribute for a given value combination of factors. So the domain of rules is, therefore, the complete range of possible permutations of a stock of factor values. By way of illustration: this range is a k dimension field, the extent of which is determined be the number of factors (k). Rules are placed in this field so that, in good cases, they will fill it completely, and so every case is covered. Rules, which are valid for just one selection of factor-values, are called elementary rules. The valid range of a rule can be more extensive, in as much as a certain factor value may be of concern to other neighboring values. These rules are called complex rules. They can be considered as a rule assembled from elementary rules.

The data mining concept

The concept of data mining lays on the following observation: organizations record and store a huge amount of data about their activity but they use most them for nothing. Maybe there are hidden relations (patterns) among the data that could be useful for the organization. These patterns can be identified with data mining tools.

It is easy to find all patterns among data stored in our databases and data-warehouses. It only needs computers, which are fast enough. By doing so we get nowhere. We will have a huge amount of rules instead of huge amount of data, both useless.

Therefore the aim is to find as less rules as possible, but to find the needed ones. For this purpose the expert of the discipline is needed. The expert selects the benchmark, which means to select an attribute according to which's values Doctus KBS will classify the cases of our experience.

The classification gives us a rule-base with few attributes only. The form of appearance is a directed graph called modeling graph. The modeling graph can be fine tuned by the expert. This is necessary to get rules, which give us the most useful new knowledge.

The inductive algorithm of Doctus KBS is capable of finding the rules as described above. But one more problem has to be solved: how the hard data (measurable) and the soft information (immeasurable) are to be handled together. This is necessary because the soft knowledge, experience, intuition of the expert is to be handled together with numeric data from databases. The soft information can be represented using nominal or ordinal scales, while the numeric data can be represented using interval or proportional scales. To handle these two together it is necessary to use a single scale. Nobody solved the problem of representing soft information using interval or proportional scales. We believe it is impossible. For that we should be able to express how much or how many times is "beautiful" better then "ugly". But there is no reason not to represent numeric data using ordinal scale. However it

would not be good to get few hundred or few thousand values for numeric attributes. Therefore Doctus KBS executes a cluster-analysis on input data putting them into classes (clusters). The number of clusters is given by the expert.

This way the hard and soft knowledge is either represented in our knowledge-base. For easy use Doctus accepts input data not only form user interface but also directly from databases or data-warehouses and from long-distance-users through Internet or intranet.

Illustration

The use of energy is increasing in Hungary so building a new power plant will be necessary. We took a part in the project of evaluating locations for a new Nuclear Power Plant. Decision maker has to consider social, economical and environmental aspects. We have built a knowledge-base using Doctus KBS for investigating the environmental aspect. Three depending attributes were used at the first level: “infrastructural”, “geographical” and “geological”.

Infrastructure considers international phone cables, regional gas and water pipes and transmission lines. This branch will have higher value if the infrastructure is poor.

Geographical position is better if there are no major settlements or border in the area, if the population has low density, if the traffic relations are good and hydrography is of highest significance (big river).

Geological factors are classified in three groups: the changeable (the level and the oscillation of ground water, the geological surface, the water supply and the disquisition), the expensively changeable (the geomorphology, the mineral deposits, the drinking water supply and the waste disposal), and the unchangeable (the earthquakes, the basement geology, the tectonics and the recent movements) attributes.

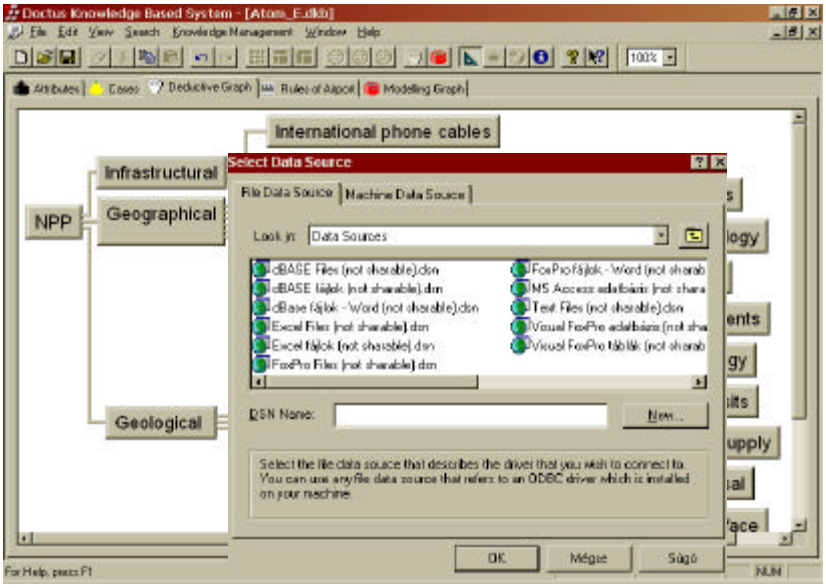

Figure 1: External data sources for the knowledge-base

After fine-tuning the knowledge-base we got rules like: if the “geomorphology” is “more than 50% plane” and the “waste disposal” can be solved with “no transport” the relevance of the “mineral deposits” and the “drinking water supply” is low. The “expensively changeable” can still be “prosperous”. But if the “geomorphology” is about “50% plane” these other attributes become relevant.

Some of these attributes are numeric. Data for the numeric attributes can be retrieved from external tables or databases. (Figure 1)

The cases for the knowledge-base can be collected from the experts using web pages. (Figure 2)

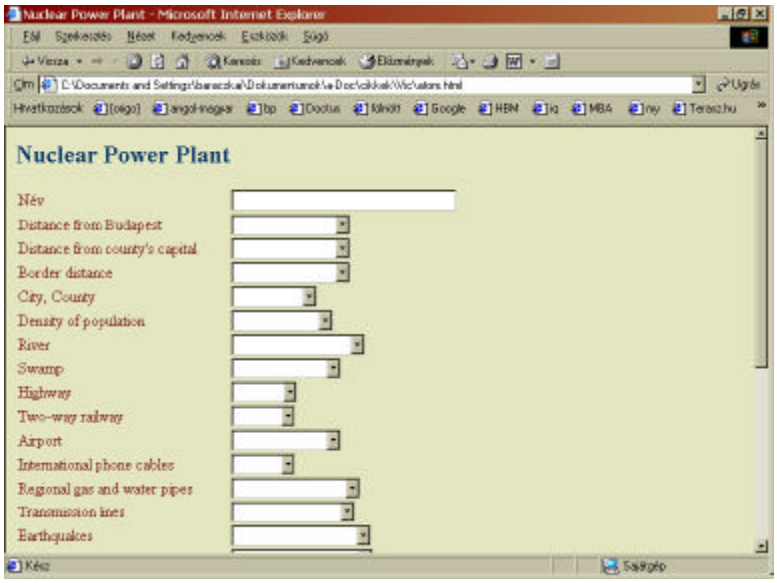

Figure 2: Web page as external input for the knowledge-base

---

[1] POLANYI Michael [1962]: **Personal Knowledge**. University of Chicago, Chicago, Illinois.

[2] RUSSELL, Bertrand [1956]: **Human Knowledge - Its Scope and Limits**. London.

[3] POLANYI Michael [1966]: **The Tacit Dimension**. Doubleday & Company, Inc. Garden City, New York.

[4] RYLE, Gilbert [1969]: **The Concept of Mind**. Hutchinson and Co. (Publishers) Ltd.

[5] ANDERSON, John R. [1983]: **The architecture of cognition**. Cambridge, Harvard University Press.

[6] MINSKY Marvin [1994]: **Negative Expertise**. In: International Journal of Expert Systems Vol. 7 No. 1 pp. 13-19 (http://www.ai.mit.edu/people/minsky/papers/NegExp.mss.txt).

[7] MINSKY Marvin [1982]: **Why People Think Computers Can't**. In: AI Magazine Vol. 3 No. 4 (http://www.ai.mit.edu/people/minsky/papers/ComputersCantThink.txt).

[8] RAYMOND, Eric Steven [2000]: **The Cathedral and the Bazaar**. http://www.tuxedo.org/~esr/writings/cathedral-bazaar/cathedral-bazaar

[9] http://www.doctus.hu/en/index.html