[Sowa] Sowa J., Knowledge Representation: Logical, Philosophical and Computational Foundations. In http://www.bestweb.net/sowa/ontology/gloss.htm

[Studer et al, 1998] R Studer, V.R. Benjamins, D. Fensel. Knowledge Engineering: Principles and methods. In Data & Knowledge Engineering 25, 1998, p 161–197

[Wielinga et at, 1994] Wielinga, B., Schreiber A.T., Jansweijer W., Anjewierden A. and van Harmelen F. Framework and Formalism for Expressing Ontologies (version 1). ESPRIT Project 8145 KACTUS, Free University of Amsterdam deliverable, DO1b.1, 1994. http://www.swi.psy.uva.nl/projects/NewKACTUS/Reports.html

Author's Information

Alexander S. Kleshchev – kleschev@iacp.dvo.ru

Irene L. Artemjeva – artemeva@iacp.dvo.ru

Institute for Automation & Control Processes, Far Eastern Branch of the Russian Academy of Sciences

5 Radio Street, Vladivostok, Russia

METHODS OF ADAPTIVE EXTRACTION AND ANALYSIS OF KNOWLEDGE FOR KNOWLEDGE-BASE CONSTRUCTION AND FAST DECISION MAKING

Alexander Kuzemin, Darya Fastova, Igor Yanchevsky

Abstract: An approach for knowledge extraction from the information arriving to the knowledge base input and also new knowledge distribution over knowledge subsets already present in the knowledge base is developed. It is also necessary to realize the knowledge transform into parameters (data) of the model for the following decision-making on the given subset. It is assumed to realize the decision-making with the fuzzy sets' apparatus.

ACM Classification Keywords: I.2.5 - Expert Systems; I.2.6 - Knowledge acquisition

Introduction

The problem of knowledge representation in the process of the expert system (ES) design is the central one as the knowledge base (KB) main function implementation, i.e. new knowledge gaining, depends on its successful solution. Starting from this the structure and form of models and methods for knowledge representation making the decisive action on the ES efficiency and external information

In the majority experts' opinion the expert system power is defined by the volume of the knowledge the given system offers. Despite the fact that a lot of instrumental means helpful in gaining knowledge has appeared recently this problem still remains poorly defined and laborious one. Knowledge gaining is inseparably connected with the process of their check-out consisting in detection of insufficient knowledge and their introduction to the system, the KB check on non-inconsistency and completeness, check of the managing mechanism, the ES analysis and modification.

The process of compatible ES development implies creation of the specialized instrumental systems. Such systems support execution of the life cycle main stages, they commonly fix presentation of the used information, the knowledge presentation language, the knowledge interpreter (display) and a set of software instruments intended for a number of problems solution. However, these systems are oriented to the support of a user from the knowledge engineer class [2,3,4] and not information carriers (experts). Thus, they do not take into account a modern approach to creation of means for information processing (MIP), which consists in exclusion of a knowledge engineer, from this process as a redundant mediator.

By gaining knowledge is meant the process of gaining knowledge from experts or some other knowledge sources and their transmission to the ES whose efficiency completely depends on the gained knowledge quality and correctness of their presentation. The complexity is stipulated by a great volume of knowledge used by the expert, which are not always realized by him as such or formulated dimly.

In terms of the ES interaction with an expert and a knowledge engineer in the process of knowledge gaining it is possible to single out the following main phases:

1. The preliminary phase. Knowledge engineer obtains from an expert or some other knowledge source a general information about software (the main concepts, relations, sub-problems, data structure) and forms a general concept of the ES design principles on their basis. Starting from the aims of the ES being created the corresponding instruments is chosen: envelopes, empty ES, development medium, knowledge presentation language.
2. The initial phase. Knowledge engineer fills the system with knowledge about presentation, i.e. the values defining the organization, structure and means of the KB presentation.
3. The accumulation phase. Gaining of the main knowledge on software envisages the following:
   - detection of the knowledge incompleteness or discrepancy;
   - extraction of new knowledge controlling a definite discrepancy, incompleteness or incorrectness of the KB;
   - transformation of new knowledge into the form intelligible to the ES;
   - correction of the KB.

The process of gaining knowledge reduces to the following problems' solution:

1. definition of the necessity to modify or widen knowledge; if there is no such a necessity then the process of gaining knowledge then the process of gaining knowledge is finished;
2. extraction of new knowledge on software;
3. transformation of knowledge into the form intelligible to the ES;
4. modification of the knowledge system.

There exist several types of models of knowledge gaining (according to the degree of knowledge gaining process automation degree and degree of their independence from an expert):

1. Gaining knowledge only with the knowledge engineer help. At the early stages of work (interaction of the knowledge engineer with a PC) the above problems solution through their assimilation with the software expert help and the following development of the system only by the knowledge engineer is assumed. The shortages of this method are as follows:

- insufficient knowledge of software doesn't make it possible to ensure the completeness and consistency of the gained knowledge;
- absence of braking down of the system components to the KB and mechanism of input doesn't make it possible to preserve once achieved consistency at inevitable modifications of the system.

2. Gaining knowledge at the cost of organizing interaction of the expert and ES directly either with the knowledge engineer help at the expense of knowledge separation from the programs and knowledge presentation in the KB simple information structures. Such a model assumes that the first and the second problems of gaining knowledge are solved by the expert with the help of the knowledge engineer, the third problem is solved only by the knowledge engineer and the fourth one – by the ES. A great labor consuming can be assigned to shortages of the given approach as of four problems of knowledge gaining of this model only one problem is automated. Incorporation into ES of the intelligent editor possessing the dialog capabilities and an extensive meta-knowledge (knowledge of the KB structure) decreases significantly the labor consuming of knowledge gaining process and reduces to minimum participation of the knowledge engineer participation in this process. When using the intelligent editor, the expert solves the first and the second problems of knowledge gaining, the third and the fourth problems are performed with the ES. This model is widely used in practice.

3. The complete automation of knowledge gaining four problems based on application of the inductive generalizing programs. It is assumed that in the KS the concrete facts on software are kept in an explicit form. The Inductive program analyses the input data from some field of the expertise and forms automatically

meaningful deviations and rules describing software. The shortage of the model is that the inductive programs are capable to generate knowledge only on the basis of the structured rules.

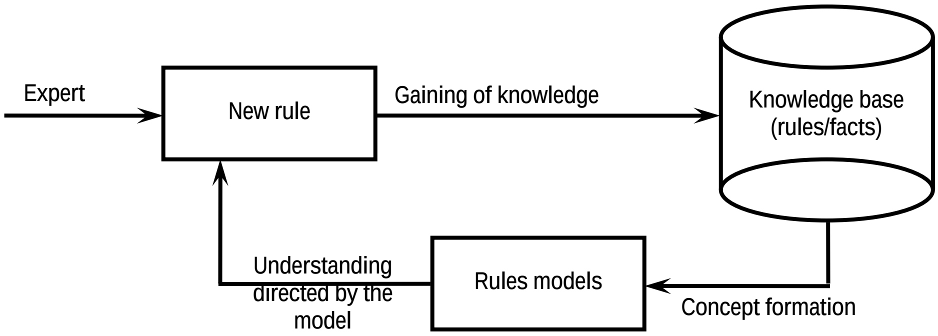

Fig. 1 – Rules derivation

Thus, to automate the knowledge gaining process the problem of gaining knowledge from the expert should be solved with the aim to develop further the means for inductive rules generation for the concrete software.

To solve such a problem it is expedient to define a priori some subset of information specifying the initial features to software. This assists software partitioning into parts with the aim of the expert's interrogation context directional narrowing.

ES successively inquires of the expert for examples of the object falling into every class called by him. Every example is analyzed for the elucidation of the following questions:

- attributes of the object in the example;
- values of these attributes in other objects belonging to the identical class;
- manifestation of these values in the objects of other classes.

The process of software rules derivation joins understanding directed by the model and the concept formation. The text of the new rule performs here the role of the signal, the program of knowledge derivation stands here for the signal processor; the new rule expressed in an internal representation is the interpretation and the model is the rules model.

Gaining of knowledge by this scheme represents the process with a feedback: the existing models of the rules acts on the process of new rules derivation and the new rules added to the KB stipulate the rules models re-calculation.

The main idea is based on the assumption that knowledge of software can be described at some generalized level in the form of the typical rules (rules models) which define the course of the knowledge gaining process. The set of the rules models is the KB model. The rules models are formed by the system automatically on the basis of the KB analysis, i.e. the system uses the rules models for limitation of the input message interpretation and the initial rule is used then for the rules model reshaping.

The availability of the rules models allows the system not only to derive new rules but also to give advice for modification of the rules introduced by the expert. Advice is given on the basis of comparison of the new rule with a model corresponding to it. If incomplete but only a partial comparison takes place then the system offers the expert to refine the rule according to those aspects of the model which turned out to be not represented in the new rule.

As the models of the rules are constructed by the system itself on the basis of the current content of KB a number of important conclusions is observed:

1. the models are created automatically and the expert doesn't take care of their creation and even doesn't know about their presence;
2. the KB models (i.e. a set of rules models) accumulate the whole previous experience (previously introduced models are used for current models shaping);

3. the KB model evolves with the KB growth and represents variations in the base content due to modification of the rules' model when inputting every rule.

A set of all rules' models can be regarded as an abstract image of KB. The system when analyzing the KB model specifies content and limits of the accumulated knowledge. For every model, it is possible to introduce the measure characterizing this model power. It is considered that the higher the model power the higher its definiteness index and the greater number of rules serving the basis for this model creation. A small power value for some model can indirectly point to insufficiency of the knowledge on some fragments of software that should be taken into account when adding to the ES rules models.

The shortages of the offered model:

- the presence of some minimum volume of information in the KB (critical mass of knowledge), it is impossible to ensure the acceptable context for gaining knowledge without it;
- derivation only one rule at the current moment;
- ES functioning is adequate if a number of detected errors is small or their number is great but they depend only slightly on each other.

Prospects of creation of the ES assisting in the process of decision-making depends on such a source as the capability for understanding and reasoning through analogy. The key to the analogy realization with a computer consists in presentation of a new information in a structured form.

From the knowledge organization standpoint, it is expedient to classify knowledge in two aspects: by the concept levels and activity levels. Each of the aspects forms their own hierarchy.

To ensure that the ES could control the solution search process, gain new knowledge it should be able not only to use its knowledge but also to understand and study them. If knowledge on software is related to the zero concept level then meta-knowledge will get to the first level. This first level contains knowledge about the means of zero level knowledge presentation. It is precisely meta-knowledge which play an essential role in consultation control, the ES actions explanation and knowledge gaining as meta-knowledge are invariant to software.

It is possible to single out the second level of the concept containing knowledge about meta-knowledge i.e. knowledge about presentation of the fundamental concepts of the first level etc. Separation of knowledge by the levels of representation widens the ES application field.

The number of comprehension levels in many instances depends on the specificity of the problems being solved, knowledge volume and means of representation. As a rule no less than three comprehension levels are singled out, they represent the general, logical and physical organization of knowledge. Introduction of several comprehension levels ensures an additional degree of ES flexibility as it makes it possible to perform changes at one level without referring to another ones. Variations at one comprehension level can only result in variations at the same level necessary to ensure matching of data and program structures.

Separation of the comprehension levels for considering knowledge with different degree of detail.

One of the main problems in terms of mathematical simulation in the process of creating the information systems for situational management is the knowledge presentation means, the decision is made in the concrete situation on its basis. Difficulty in knowledge presentation consists in realization of the information fragments transition into terms of the databases (DB) and knowledge bases (KB) structures. In terms of facts and processes resulting in changes in DB one should consider semantics and syntax of such presentation. By syntax is meant a set of rules for joining symbols into logically concrete expressions, and by semantics is meant the method of interpretation of expressions formed as a result of the concrete realizations of syntax rules.

Urgency of the problem is stipulated by the necessity of the information intelligent system development on the BD and KB basis which will be able to shape and add the database and knowledge base itself and perform logical derivation for the further decision-making on subsets.

Aim of the work consists in reduction of the time needed for decision-making on some definite subset at the cost of derivation and analysis of a new knowledge arriving to the KB input.

To gain the specified aim it is necessary to solve the following problems:

– to realize predicate inquiries construction and their modification which are the apparatus for description and investigation of the processes of updating and modification of databases and knowledge bases,

- to define rules of logical derivation based on databases an knowledge bases,
- to calculate the shortest way between knowledge needed for referring the new knowledge to the nearest subset.

By the database is meant a set of facts. The main ideas of such an approach are considered in the frameworks of the SQL concrete realizations or WWW and WEB realizations.

But in the above realizations the function of inquiries logical substantiation is displaced to the database user, in Prolog-program the knowledge base construction by the engineer is envisaged and constant support of the user during the logical derivation session is applied.

1. KB description

All decisions in the universe of discourse are made on the basis of experienced experts' analysis conclusions. The information system knowledge base is considered according to [1,2] as a set of information essences of atomic predicates from some information space $\Re$. All changes taking place in the knowledge base are considered as a result of the modifying predicate inquiries Q_m . The basis of the predicate inquiries taken alone is a set of the modifying predicate rules:

$$Q_m \leftrightarrow (K_B)^{<<} \left\| \begin{matrix} K_{B-(X)} \\ K_{B+(X)} \end{matrix} \right. << \qquad (1)$$

where $X \in \Re$, $K_{B+(X)}$ means that the atomic predicate o should be incorporated into the knowledge base K_B ; K_B means that X should be eliminated from the knowledge base; $(K_B)^{<<}$ means the knowledge base modification at the level of the predicate rules logical coherence as a consequence of the rules incorporation and elimination operations use; $K_{B\pm(X)}$ means the possibility of modification not only of the knowledge base but security of the user as well on the basis of descriptors; $<<$ is considered as a complex arrow, its features are studied with the categories theory.

2. Knowledge Gaining

Knowledge can be presented as production rules of the type [3]:

«if $X_1 \& ... \& X_K$, then $X_{K+1} \& ... \& X_{K+L}$ »,

where $X_1...X_K$, $X_{K+1}...X_{K+L}$ - some predicates.

Definition 1. The set $W = \Pi_1 \times \Pi_2 \times ... \times \Pi_{K+L}$ is called the contents of the knowledge «if $X_1 \& ... \& X_K$, then $X_{K+1} \& ... \& X_{K+L}$ ». The random element of this set is called the element of the knowledge contents.

The set $W_1 = \Pi_1 \times \Pi_2 \times ... \times \Pi_K$ is called the contents of knowledge condition. The random element of this set is called the element of the knowledge condition contents.

The set $W_2 = \Pi_{K+1} \times \Pi_{K+2} \times ... \times \Pi_L$ is called the contents of the knowledge result. The random element of this set is called the element of the knowledge result.

Definition 2. By the probability p_i of the knowledge contents element $w_i \in W$ is meant the event probability, which consists in that all predicate constants being a part of w_i will take the logical meaning «and» when substituting the objects from the truth domain of variable predicates making part of this knowledge instead of arguments.

The element of the knowledge content w_j represents a vector, predicates variables making part of this knowledge are its components. Vector $z_j = z_j(1)$,..., $z_j(K+L)$ from R^{K+L} can be put in conformity with the knowledge content element w_j .

The function of knowledge distribution is the function of $K+L$ arguments: $F(y)=F(y(1), y(2),..., y(K+L))$, with the domain of definition R^{K+L} and assuming the values in the space R^1. It is defined by the formula $F(y)=\sum_{z_j \le y} p_j$, where z_j - image of the knowledge contents element w_j in R^{K+L}. Expression $z_j < y$ is understood as a fulfilling of the conditions: $z_j(i) < y(i),\ i=1,\ ...,\ K+L$.

Definition 3. By distance between the comparable knowledge 3H1 and 3H2 is meant Hellinguer distance $d(G,\ Q)$ between two probability distributions of their contents elements $G=\{p_{11},\ p_{12},\ ...,\ p_{1r},\ ...\}$ and $Q=\{p_{21},\ p_{22},...,p_{2r},...\}$, which is calculated by the following formula:

$$d(G,Q)=\sum_j (\sqrt{p_{1j}}-\sqrt{p_{2j}})^2 \qquad (2)$$

Calculating the distance between knowledge, the problem of a new gained knowledge distribution is solved. The distance between a new knowledge and all knowledge in the knowledge base and then a new knowledge is placed into the subset containing such a knowledge for which the distance assumes the least value.

3. Decision-making

Despite the fact that decision-making is realized in the chosen knowledge subset for the complex systems and processes the adequate mathematical description of decision-making is absent or represent rather cumbersome mathematical constructions whose optimization and practical use in the real time are impossible. This problem can be solved with using algorithms built with the models simulating the process of decision-making by an experienced expert. The theory of fuzzy sets can be used as a mathematical apparatus for a great number of decision-making models. When choosing decisions in the situation centers (SC) the aim of the designing stages consists in the choice of the design version or the parameter value from the rather small specified set defined, as it was mentioned above, with the formula (2). To simulate the decision-making process it is offered to use the decision-making models based on the fuzzy rule modus ponens, fuzzy inductive scheme of derivation and fuzzy expert information of the second type. In this case, the following inductive output scheme will be used [4]:

$$\frac{\tilde{L}^{(2)}=\begin{cases}\langle If\ \ \tilde{B}_1\ ,then\ \tilde{A}_1\rangle; \\ \langle If\ \ \tilde{B}_2\ ,then\ \tilde{A}_2\rangle; \\ \dots\dots\dots\dots \\ \langle If\ \ \tilde{B}_m\ ,then\ \tilde{A}_m\rangle; \\ A'-true\end{cases}}{B'-true} \qquad (3)$$

Here the clear statements A' and B' have the following form:

$$A':\langle \beta_W\, is\ \ w'\rangle;\quad B':\langle \beta_V\, is\ \ v'\rangle$$

$$w'=(x,\ y,\ z...)\in X\times Y\times Z\times...,\ \ v'\in V.$$

In the given derivation scheme the statements concerning the values of the input parameters are the premise for the scheme itself (statement A') and the result inside the system $\tilde{L}^{(2)}$ of the statement (statement A_j). The statements about the values of the output parameters are the result for the derivation scheme (3) (statement B') but it is the premise inside the scheme $\tilde{L}^{(2)}$ (statement $\tilde{B}_j$). That is why to choose the values of the output parameter V based on the rule modus ponens it is necessary to transform the derivation scheme (3) to the form:

$$\frac{\begin{array}{c}\widetilde{L}^{(1)}\\ A'-true\end{array}}{B'-true}.$$

To do this it is offered to transform the second type statements system into the first type system equivalent to it using the contraposition rule according to which the statements “IF A, THEN B” and “IF¬B, THEN¬A” are equivalent for the random expressions, i.e.

$$\langle If\ A\, then\, B\rangle \equiv \langle If\ \neg B, then \neg A\rangle,$$

Here expressions $\neg A$ and $\neg B$ are negative expressions A and B.

Applying the contraposition rule to the expressions $\tilde{L}_j^{(2)}, j=1,...,m$ of the second type system we will obtain

$$\langle If\ \widetilde{B}_j\ then\ \widetilde{A}_j\rangle \equiv \langle If\ \neg\widetilde{A}_j, then \neg\widetilde{B}_j\rangle,$$

where the statements $\neg\tilde{A}$ and $\neg\tilde{B}_j$ can be regarded as statements $\langle \beta_W\ is\, \alpha_k\rangle$ and $\langle \beta_V\ is\, \alpha_l\rangle$, $k=W_j^*$ and $l=V_j^*$ where the values α_k and α_l are defined by the belonging functions μ_k and μ_l, being the complement to μ_n and μ_p, where $n=W_j$ and $p=V_j$:

$$\mu_k(w)=1-\mu_n(w), \forall w\in W=X\times Y\times Z\times ...$$

$$\mu_l(v)=1-\mu_p(v), \forall v\in V.$$

Conclusions

Scientific novelty of the developed approach consists in that the information intelligent system based on DB and KB makes it possible to shape and complement the database and knowledge base. After that, it in its own transforms the analyzed knowledge into parameters of the model for making decisions on a definite subset.

Practical value – in the average the given approach shortens the time needed for making decision on the subsets for 20%. This result was obtained at the cost of using Hellinguer distance at the stage of new knowledge comparison and classification.

Bibliography

1. Sheketa V.I. Modification Predicate Applications // Programming Problems. Institute of Program Systems of NAS of Ukraine. – 2004. - №2-3. – 339-343pp. //Special Issue. Proceedings of the 4-th SPC “UkrProg2004”, July 1-3, 2004.– Kyiv, Cybernetics Center of NAS of Ukraine.
2. Sheketa V.I. Predicate Queries Modification, as a Tool to Work with the Knowledge Bases of Oil and Gas Subject Domain //Proceedings of 6-th International Scientific Conference “Modern Problems of Radio Engineering, Telecommunications and Computer Science – TCSET’2004”. – Lviv-Slavsko. – 2004. – February 24-28. – P.315-319.
3. V.N. Koval, Yu.V. Kuk. Knowledge Gaining and Analysis //Artificial Intelligence. – 2004. - №3 – 293-304pp.
4. N.G.. Malyshev, L.S. Bershtein, A.V. Bozheniuk. Fuzzy Models for Expert Systems in SAPR. M.: Energoatomizdat, 1991, 136 p.

Authors' Information

Kuzemin Alexander Yakovlevich, Cand. Techn. Sciences, Prof. Department of Information Sciences, Kharkov National University of Radio Electronics. Address: KNURE, 14, Lenin Ave., Kharkov. Tel.: (057) 702-15-15. Scientific interests: systematic approach in designing and programming

Fastova Darya Vladimirivna, post-graduate, Department of Information Sciences Kharkov National University of Radio Electronics. Address: KNURE, 14, Lenin Ave., Kharkov. Tel.: 702-15-15. Scientific interests: information systems design, systems analysis.

Yanchevsky Igor Vladimirovich, General director of JSC PJ “CT and SA”, competitor of the Department of Information Sciences, Kharkov National University of Radio Electronics.