Usage of Expert Classification in Diagnostic Expert Systems' Knowledge Bases Construction

Aleksandr Koval[1], Safwan Al Salaimeh[2], Oleh Andriichuk[3,1][0000-0003-2569-2026]

[1] National Technical University of Ukraine "Igor Sikorsky Kyiv Polytechnic Institute", Kyiv, Ukraine
[2] Aqaba University of Technology, Aqaba, Jordan
[3] Institute for Information Recording of National Academy of Sciences of Ukraine, Kyiv, Ukraine
avkovalgm@gmail.com, safwan670@yahoo.com, andriichuk@ipri.kiev.ua

Abstract. Studies of resent years in the field of artificial intelligent led to a number of major achievements. The most significant of these was the development of powerful computer systems, known as systems based on knowledge. They are based on the program, designed to represent and apply actual knowledge from subject domain to problem solving. Diagnostic belongs to weakly structured domains, so that the essential (sometimes only) source of information in this domain is experts.
In this paper the possibility of computerized construction of diagnostic expert systems' knowledge bases is described on the basis of solving the problem of expert classification. The described in the work approach provides a complete build of consistent knowledge bases of diagnostic expert systems.

Keywords: knowledge base, artificial intelligent, problem solving, expert system, decision-making, possibility.

1 Introduction

Studies of resent years in the field of artificial intelligent led to a number of major achievements. The most significant of these was the development of powerful computer systems, known as systems based on knowledge. They are based on the program, designed to represent and apply actual knowledge from subject domain to problem solving. A general overview of intelligent computer systems is translated into [3]. Example of development in recent years in [2] the possibility of using expert system (ES) is considered for teaching students using a computer, an expert rating system in commercial banks is considered. In [4] the principles of constructing expert systems of real time and for the problem of diagnostics are considered and management and environmental monitoring.

Diagnostic decisions are influenced by numerous quantitative and qualitative criteria, factors, and parameters. It is problematic to provide a formal mathematical (analytic) description of these factors. This leads to the fact that diagnostic belongs to weakly structured domains [5, 6].

2 Features of Weakly Structured Subject Domains

As shown in Fig. 1, there are following properties of weakly structured subject domains: absence of functioning goal which could be formalized, absence of optimality criterion, uniqueness, dynamics, incomplete description, presence of subjective human factor, impossibility of analytical model building, lack of benchmarks, high dimensionalities.

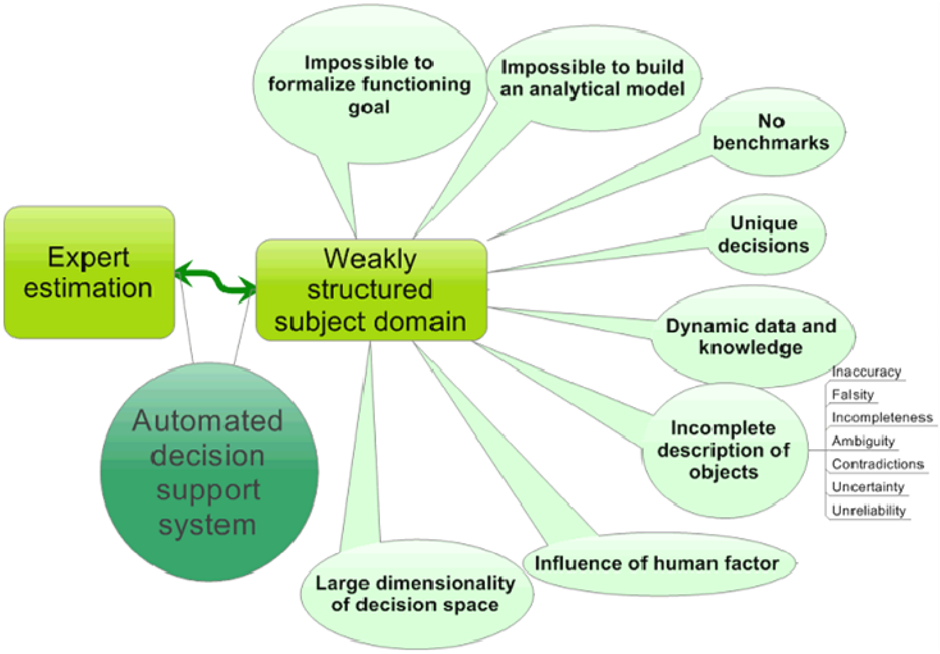

Fig. 1. Features of weakly structured subject domains

Let's consider in more details the characteristics of mentioned above weakly structured subject domains. Objects in such domains are unique. In fact, management systems for these areas are created once, in order to solve real problems; the transfer of such models to other objects is costly or simply impossible.

In systems that are not created by a human (such as biological), there is no goal of functioning that can be formalized. The purpose of the functioning of such systems is their efficiency in general, the support of some parameters in the specified limits, but it is usually impossible to formalize such a goal in the form of a certain criterion. For instance, in the biological system, all factors affecting its functioning are so numerous and the connections between them are so complex and not obvious that it is impossible to set a certain function to describe the purpose of its functioning.

Due to the lack of a functioning goal that is subject to formalization, it is impossible to construct a function whose optimization would provide the best mode of operation of the object. Objective optimization function does not exist, you can specify only some of the factors that can be optimized. However, it is impossible to optimize each of these factors separately, since they are closely interconnected, and their connections will be disturbed during the process of optimization. This can lead to a violation of the process that regulates or maintains the system in a stable condition relative to the changing environment in which the system operates, and may lead to catastrophic events and irreversible changes in the system.

Since in the weakly structured subject domains there is no goal of functioning that can be formalized, and it is impossible to construct a function whose optimization will provide the best mode of operation of the object, then it is impossible to construct an analytical model of this subject domain.

Dynamism is due to the fact that the structure and functioning of the object change over time, that is, the object evolves. Management of such systems should be adaptive, able to change when the object changes.

The incompleteness of the description is due to inaccuracy, incompleteness, falsity, ambiguity, contradictoriness, uncertainty, and unreliability of the data describing the object.

Characteristics of objects is problematic to describe quantitatively, therefore in weakly structured subject domains it is inappropriate to speak of the existence of benchmarks of these characteristics.

The large dimension of the decision space is due to the large number and heterogeneity of the criteria that characterize the subject domain.

The objects of management can be people who have free will. It is often impossible to predict human behavior as an object of control or a component of a system. A person operates in the system, taking into account his personal goals and interests. Therefore, when modeling the object of human behavior management is difficult to consider.

Described above properties of the weakly structured subject domains lead to the fact that the essential (sometimes only) source of information in them is experts.

3 Problem Statement

A fairly compute analysis of the problems organization – related receipt and the interpretation of expert information when creating an expert system is contained in the work [7].

Wide class of task, for which expert systems are development, design diagnostic tasks. Diagnostic expert systems are essentially designed to solve the classification problem: each object (situation). From the subject area a sings his diagnosis. Properties, course of action large role in the experts work of their knowledge is plays with the manual of the organization of expert inquiry. Success in creating an information system became possible due to the fact that a certain expert apparatus was developed, allowing to largely overcome the above problems. The essence of this approach is use of data on the structure of the problem to be solved, about the relaxation between its other elements. Such information makes it possible to draw certain conclusions on the possible limits changes of a number of estimates and thus reduces the number of questions to the expert, control the consistency of information. Identify and eliminate appearing in it mistakes [8, 9].

Diagnosis of the object on the basis of its description is possible only if the different values of the signs have different degrees of characteristics and for diagnosed properties. In connection with these, when structuring the task of expert classification, hypnosis is used about the varying degrees of the characteristics of the individual values of each characteristics and for each property (or what is the same class). It is also assumed that, for each feature, the expert may order its value by their characters and for the corresponding class and this order doesn't depend on the value of other characteristics. The experiment of solving the problem of expert classification suggests that the assumption of ordering the meaning of the attribute is also valid for many practical problems, and accordingly, the considered formulation covers a wide class of the problem of expert classification, the possibility of obtaining valuable information about the objects membership class makes it possible to construct a rational procedure for expert demand in order to minimize the number appeal to him. In addition, detailed information allows you to identify possible errors in the expert's answers [10, 11].

4 The Solution of Problems

The formulation of the classification problems is as follows. Given a set independent properties which can have the object of research:

$$P = \{P_1, P_2, \dots, P_i\};$$

M – signs, characterizing from different sides the objects of research; set Q_m possible m-th value of the m-th characteristics:

$$Q_m = \{q_{m1}, q_{m2}, \dots, q_{mn}\}$$

and n_m – the number of these values; set of all hypothetical possible states of the object of investigation:

$$A = Q_1 \times Q_2 \times \dots \times Q_m;$$

in this state $a_i \in A$ is characterized by a vector:

$$a_i = (a_{i1,}a_{i2}, \dots, a_{im}),$$

$$a_{im} \in Q_m,\ m = \overline{\iota, M}.$$

Formally hypotheses a specifically can be described as follows, ordering characteristics values Q_m by their character for the property p_i allows to add to Q_m transitive and antifriction binary relations (linear order):

$$(q_{ms}, q_{mi}) \in r_m,$$

if q_{ms} more characteristics for this property than q_{mt}. On this basis of these relation it is possible to construct a binary relation of dominance by characteristic for each property on the set of the state on the object of investigation which are described as follows:

$$R = \{(a_s, a_i) \in A \times A |\ m = \overline{\iota, m},\ (a_{is}, a_{im}) \notin r_m\},$$

$$m \in M: 1 \le \vartheta \le m,$$

such that *a*, have a reflexive transitive relation of strict dominance:

$$a_{s\vartheta}, a_{i\vartheta}, a_{t\vartheta} \in r_\vartheta.$$

Required based on expert knowledge for each state from *A* identify the presence of relevant and properties from the set *P* and thus build a classification of the set *A:*

$$\mathrm{A} = \cup_{\mathrm{i}=0}^{\mathrm{k}} \mathrm{k_i},$$

such that, the state $a_i \in A$ belongs to the class K_i, if the object in this state has the experts opinion and properties P_i. To class K_o, in this case, there are such states in which the object doesn't possess, in the experts opinion, not one consideration. It is customary to call this problem the main task of classification [12,13,14].

The procedure for constructing the classification is described as follows: we put in correspondence to each state $a_i \in A$ to sets C_i^+ – set of the numbers of classes belong to the state a_i and C_i^- – set of class numbers, to which the state a_i cannot belong. Let *K* – multiple numbers of generated classes:

$$\mathrm{K} = \{0, 1, \dots, \mathrm{L}\}.$$

Status will be considered classified, if the following two conditions will be done:

$$\mathrm{C_i^+} \cap \mathrm{C_i^-} = \emptyset,$$

$$\mathrm{C_i^+} \cup \mathrm{C_i^-} = \mathrm{K}.$$

Denote by $A^o \in A$ a subset of all classified state. Before the beginning of the expert survey $\forall_{a_i} \in A$ suppose:

$$C_i^+ = C_i^- = A^o = 0.$$

The procedure for an expert survey and when $A^o = A$. In the course of the survey the next required expert advisor is selected:

$$a_i \in A \backslash A^o.$$

The expert makes conclusion on this condition in the form of the list of classes belonging (object properties, is in this state). Thus for this state a_i the set is defined explicitly C_i^+ and an implicit set:

$$C_i^+ = K \backslash C_i^-.$$

After this condition a_i is classified and:

$$A^o = A^o \cup a_i.$$

Information received regarding the state a_i, allows reducing uncertainly for a number of other states [15 - 20].

Along with the main task of expert classification on solving practical problem, there may be a need to clarify the severity of the diagnosable properties, which can be sensible ordered. With this definition of groups the state of the objects, in which he has the same degree of severity and diagnosable properties, means the spitting of the state of an object into a set of defined classes. It is customary to call this task the expert classification of the problem of order to classification.

The problem of the order of expert classification for a certain property can be defined as follows: there are many independent of properties, which may have the object of study:

$$P = (P_1, P_2, \dots, P_n)$$

For some properties of P determine the set of its values:

$$\{P_1, P_2, \dots, P_n\}.$$

On this set the ration of linear order is determined $\check{R}$ such that (i.e. value set are ordered from larger than the degree of expression of this property):

$$(P_i, P_j) \in \tilde{R}, \text{if } i < j$$

Let Q – set of symptom, describing the state of the object of study:

$$Q = \{Q_1, Q_2, \dots, Q_m\}.$$

On the scale of each feature Q_m $(m = \overline{1, M})$. A certain ratio of linear order:

$$(q_{ms}, q_{mt}) \in r_m,$$

if q_{ms} more characteristic for this property, than q_{mt}, the set of all possible states of the object of investigation:

$$A = Q_1 \times Q_2 \times \ldots \times Q_m$$

On this set a certain relation of strict domination *R* (similar to the considered above).

Required based on expert knowledge. Determine the severity of this device for $\forall a_i \in A$ and thus to construct a partition of the set *A*:

$$A = \bigcup_{n-1}^{N} Y_n (Y_i \cap Y_j = \emptyset, \; i \neq j, \; i, j = \overline{1, n})$$

such that $a_i \in A$ belong to the class Y_n $(1 \leq n \leq N)$ if the objects in this state have this property of degree P_n.

The procedure for experts can be described as follows: let G_i – a set of class numbers Y_B, which correspond to the state $a_i \in A$. Before the survey for:

$$\forall_{a_i}, G_i = \overline{1. N},$$

the classification can be considered complete, when:

$$\forall_{a_i} |G_i| = 1.$$

Let the expert determined, what the condition is $a_i \in A$ corresponds to the value $P_n (1 \leq n \leq N)$ by the degree of expression of the property under consideration, i.e. $a_i \in Y_B$. Consequently, in this case, the state described by a set of characteristics is no less characteristics for this property, it cannot have a lesser degree of its expression that is, if $a_i \in A$ and $(a_i, a_j) \in R$, then:

$$a_j \overline{\in} Y_n, k > n.$$

Similar condition, described by a set of characteristics value of the characteristics most characteristics for this properties, no more characteristics of this property can have a greater degree of its expression, that is, if $a_i \in A$ and $(a_i, \; a_j) \in R$, then:

$$a_j \overline{\in} Y_n, k < n.$$

There are various ways to determine the in formativeness of the object to the expert. It is also possible to estimate the informative state is the number of indirectly classified states based on the characteristics relationship. For each state, you can find this number with every possible expert answer and calculate in the middle value or the minimum of them.

Using these indicators it is possible to compare all unknown states of an object by informative their presentation to the expert and choose the most informative. When presentation the most informative status to the expert, we will receive on average the largest amount of information with any expert answers.

Whenever any procedure for interviewing experts should consider the possibility of errors in the answer. These errors can be detected by inattentive, expert fatigue, and also the complexity of the problem being solved. Since the knowledge base must be non-selective, information analysis needed, obtained from an expert, controversy detection. The possibility of indirectly defining the classes of accessories of the state allows you to check the consistency to expert assessments. If there is discrepancy in the indirect and direct evaluation of the status this indicates that there are errors in his answer. It is necessary to present contradictory answers of experts for their comprehension and choosing the right way to assess the series condition.

Possible two strategies for removing the contradiction when building a knowledge base. One of this to continuously compare information, receive from the expert, from the received early and check for inconsistency. If there is contradiction between the last expert answer and the previous information, this contradiction is presented to the expert for analysis and choice of the contradiction of politics. Another strategy is to obtain from the expert or a part, either all the necessary information, and only then the implementation and in it of the search for contradictions and their removal.

In this way. The procedure developed by the expert survey should, on the hand, minimize the work of the expert and on the other hand, allow him to analyze the information received from him, from the point of view of its consistency. If the traditional way of building a knowledge base for diagnostics systems, an expert has to solve the problem of synthesizing of their knowledge, then the proposed method corresponds to its usual task of analysis specific situations. So, he unconsciously uses many of his skills and techniques, which is difficult for him to formulate in an explicit form.

The methods for solving the problem of expert classification have natural limitations, due to their dimensionality. In problems of large dimension (dozens of the signs with a large number of possible boundaries and a large number of diagnosed properties) the laboriousness of the experts work also sharply increases and the computational complexity of the procedure increases the solution of the problem of expert classification, which makes it difficult to use them directly. In practical problems of classification, it is often possible to distinguish separate groups of attributes according to their semantic content in such a way that the characteristics to one group reflect on of the sides of the object under consideration. In these cases one of the possible ways of constructing the coverage of the initial state set is to decompose the original problem into subtasks less than the size, the used classification procedure will consist of several stages.

Let all sets of attributes Q broken into T groups:

$$S_1, S_2, \dots, S_T;$$

$$U_{S_i} = Q, S_i \cap S_j = \emptyset.$$

Each group of signs to a certain extent characteristics the presence of the object to be diagnosed to properties. Therefore, according to the values of the *i*-th group of characteristics $(i = \overline{1,T})$ the expert can make a preliminary conclusion about the possible classes of states, the characteristics described by these values.

Each group of characteristics gives incomplete, only a partial description of the state of the object of investigation. Accordingly, the evaluation of the expert, it is based only on the part of the object description, can has a probabilistic character and reflects the varying degree of the experts confidence in the availability or the absence of the object in the given state of the diagnosed properties. Consequently, at the first stage of solving the problem, the expert decides its main, and order any classification tasks within each group of characteristics, which allows to build an orderly classification, and for each properties for each group of each characteristics. This give opportunity for each state:

$$\mathrm{a_i} = (\mathrm{a_{i1}}, \mathrm{a_{i2,}}, \dots, \mathrm{a_{iM}})$$

and for each property P_i define a vector $(T_i, T_2, \dots, T_S)$, where T_S – degree of suspicion of *i*-th property in the state a_i by *j*-th group of attributes.

The next stage of the solution consists in construction generalized classification for each property based on the other of their classification by separate groups of characteristics. In this case, the signs of the second level are also the degree of the experts confidence in the availability of the *i*-th property for each of the *T* group of initial signs. It is necessary for each property P_i $(i = \overline{1,l})$ on the basis of the value of the attributes of second level, construct an order classification, determine whether or the absence of each of the property under consideration and its degree of serenity.

If the expert cannot make a diagnostic conclusion for a certain group of signs. Not having information about the result, received for another group of characteristics, the survey is conducted as follows: if the diagnosis is on *i*-th group, the expert must build *L* order of classification for *i*-th group of characteristics with the addition of the result for property p_i by *j*-th group of characteristics, number of hierarchy levels (as the decomposition levels) can vary in different tasks. It is determined the classification problem.

Conclusion

In this paper the possibility of computerized construction of diagnostic expert systems' knowledge bases is described on the basis of solving the problem of expert classification.

The described in the work approach provides a complete build of consistent knowledge bases of diagnostic expert systems.

References

1. Safwan Al Salaimeh The Optimal Management of Information Servicing Logistics System / Institute Mathematics and Computer Science Journal – India, 2003. – pp. 75-80.
2. Safwan Al Salaimeh Information Technologies of Distributed Applications Design / Institute Mathematics and Computer Science Journal – India, 2003. – pp. 99-103.
3. Safwan Al Salaimeh, ZaferMakadmeh Multi-Criteria Synthesis of Logistics Systems Through the Hierarchy Analysis / Journal of System Sciences // Poland University of Technology, Poland. – pp. 107- 115,
4. Safwan Al Salaimeh, Khaled Batiha Business Process Simulation with Algebra Event Regular Expression / Information Technology Journal – Volume 5, Number 3 – Pakistan 2006 – pp. 583-589.
5. Taran T.A., Zubov D.A. Iskusstvennyy intelekt. Teoriya i prilozheniya (Artificial intelligence. Theory and applications) // Lugansk: V. Dal VNU, 2006, - 239 p. (in Russian).
6. Glybovets M.M., Oletsikiy O.V. Shtuchnyy intelekt (Artificial intelligence). – K.: "KM Akademiya" pulblishers, 2002 – 366 p. (in Ukrainian).
7. Khaled Batiha, Safwan Al Salaimeh E-Learning / Leonardo Electronic Journal of Practices and Technology – Romania, 2006. – pp. 1-4.
8. Khaled Batiha, Safwan Al Salaimeh, Khaldoun Al Besoul Digital Art and Design / Leonardo Journal of Science – Romania 2006. – pp. 1-8.
9. Khaldoun Al Besoul, Safwan Al Salaimeh The Structure of logistics organizational technological system / Journal information society – Vol.4, Num. 7 – Romania, 2007, pp. 126-129.
10. Howard Bandy Modeling Trading System Performance / Blue Owl Press, Inc., 2011. – 384 p.
11. Yan Houmin, Yin George, Zhang Qing Stochastic Processes, Optimization, and Control Theory: Applications in Financial Engineering, Queuing Networks, and Manufacturing Systems – Springer, USA, 2006. – 360 p.
12. Process dynamics modeling, analysis, and simulation by B. Wayne Bequette 1st Edition / Prentice-Hall PTR, Upper Saddle River, New Jersey 07458, 1998. – xviii + 621 pages.
13. Otto Bretscher Linear Algebra with Applications, 4th Edition / Pearson, 2008. – 504 p.
14. Software process modeling by Silvia T. Acuna, Natalia Juristo / Springer US, 2005. – xxiv + 208 pages.
15. Hossein Bidgoli Modern Information Systems for Managers / Emerald Publishing Limited, 1997. – 438 p.
16. Howard Bandy Modeling Trading System Performance / Blue Owl Press, Incorporated; 1 edition, 2011. – 400 p.
17. Salim Istyaq, Safwan Al Salaimeh, Amjad Miqdadi Decomposition Algorithm of the Model of Electronics Systems for Modeling in Conditions of Distributed Resource / International Journal of Emerging Technology and Advanced Engineering – Volume 8, Issue 3, March 2018. – pp. 29-31.

18. Daniel H. Pink When: The Scientific Secrets of Perfect Timing / Riverhead Books, 2018. – 272 p.
19. Dan Ariely Predictably Irrational: The Hidden Forces That Shape Our Decisions – HarperCollins Publishers, 2008. – 280 p.
20. Marjory Harris The Personal Power Roadmap: The Ultimate 7 Step System to Effectively Solve Problems, Make Decisions, and Reach Your Goals, Paperback – CreateSpace Independent Publishing Platform, 2016. – 168 p.