An Intelligent Medical Differential Diagnosis System Based on Expert Systems

Valeriy A. Kamaev, Dmitriy P. Panchenko, Nguen Vien Le, and Olga A. Trushkina

Volgograd State Technical University, Russia
kamaev@cad.vstu.ru, {panchenkodp,nvien.vstu}@gmail.com,
vamp@volgograd.ru

Abstract. This abstract presents an approach to designing an intelligent medical diagnosis system based on the theory of expert systems. A formal model for a medical expert system on differential diagnosis is proposed. The architecture of the system is also presented.

Keywords: medical diagnosis, expert system, frame system, product system, inference engine.

1 Introduction

Medical diagnostics is one of the most difficult tasks of practical health care. Currently, there is a necessity of application of information-communication technologies in medicine, especially in the problems of creation of the telemedical diagnostics systems. The expert systems of medical diagnostics will execute advisory medical aid to patients in the most remote areas, thereby substantially reduce costs for patients. Of a particular importance is the possibility of creating a common knowledge base on the basis of highly qualified and specialized knowledge and experience of doctors from leading medical centers.

The system does the output of a diagnostic solution to the patient based on a set of incoming symptoms in the real-time mode. On the basis of diagnostic hypotheses, you can define the possible qualification of a physician, who needs to be contacted by the patient. The system must function even when the information is not enough, i.e. to be able to reason under uncertainty. In the case of the patient not accepting the result, he may claim the system solutions from the experts-doctors. The system will establish a dialogue with doctors in delayed mode via e-mail.

2 The Formal Model of the Medical Expert System for Differential Diagnosis

The formal model of the medical expert system for differential diagnosis consists of the following components: medical knowledge base; working memory; output control

A. Kravets et al. (Eds.): JCKBSE 2014, CCIS 466, pp. 576–584, 2014.

diagnostic solution; explanation of effective information; acquisition of medical knowledge; user interface [1, 2, 3].

2.1 Medical Knowledge Base

The base of medical knowledge is a key concept of the system. Expert systems are most effectively characterized based on the combination of fuzzy-production and frame models of knowledge representation. This approach also allows the representation of uncertainty of information in the description of the structure of the symptom. Frames are used for the representation of static knowledge on the current state of the field of diagnostics. The frame is considered as a set of slots, which are defined as structural elements that describe the properties of the frame. Fuzzy model of knowledge representation is recommended for the representation of dynamic knowledge about the transitions between different states of diagnosis.

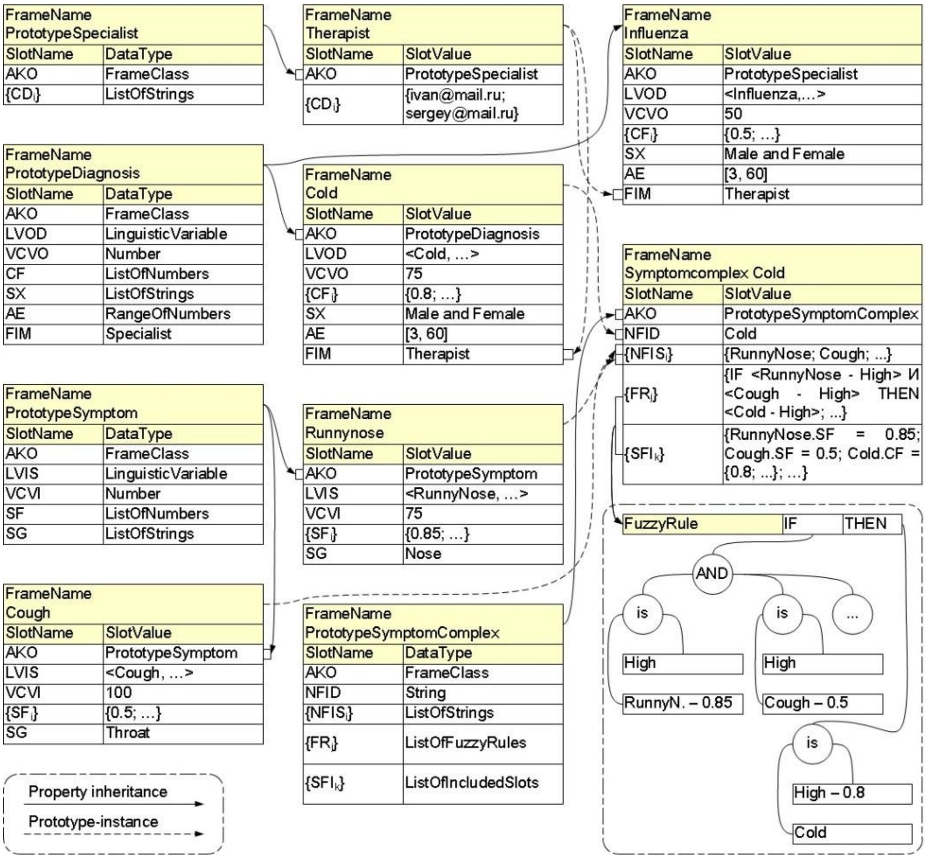


Fig. 1. The fragment of the frame hierarchy knowledge base

As an example, you can define ill-production rules of symptomcomplex cold as follows:

IF «Runny nose» = «High» (SF=0.85) AND
«Cough» = «High» (SF=0.5) AND
...
THE «Cold» = «High» (CF=0.8)

Here SF – specificity factors of symptoms from symptomcomplex; CF is the certainty factors of the rule. And for the assessment of values of the linguistic variables symptoms and diseases used a single scale qualitative terms: Low; Medium; High. The fragment of the frame hierarchy of the knowledge base is presented in Fig. 1.

2.2 Working Memory

Working memory system is intended to store information of a patient, effective diagnostic information and explanatory information. Under patient information, there are two types: personal information about the patient; quantitative assessment of incoming symptoms. Under effective diagnostic information is the quantitative integrated assessment of each disease. When the facts of receipt of the information of the patient or effective information are updated, the system creates entries (i.e. elements working memory) for each fact in the form of tuples of different length. Each entry can contain the whole fact or vice versa, the fact may be a system of records of fixed length.

2.3 Output Control Diagnostic Solution

Output control diagnostic solution is intended for diagnostic decisions based on quantitative estimates of incoming symptoms. Diagnostic hypothesis is referred to as a disease with the value of the output variable. Depending on a quantitative assessment, the disease is in one of three possible states: inactive candidate, potential candidate and active candidate. Output control diagnostic solution consists of the following four steps: mapping, conflict resolution, activation and action. Before inference engine, it is necessary to enter the facts of the symptoms in the working memory. Inference engine exists as follows [4]: procedures-methods are initialized during specification of diagnostic hypotheses that implement the backward chain to clarify the original values of the possible symptoms; when assigning initial values by slots, proceduresdemons function and are responsible for the forward chain that perform fuzzy output of a target values of slots diseases.

Backward Chain Strategy

Each symptomcomplex includes many symptoms, but some symptoms may simultaneously be included in several syndromes [5]. For further diagnosis, the generation of additional questions, concerning only possible diseases, is required. This provides higher efficiency of diagnosis. Refinement of diagnostic hypotheses consists of the following four steps. The calculation of areas of valid and invalid solutions for each disease is based on the odds of specificity of symptoms of symptom-complexes. To highlight a definite diagnosis for the disease from this set, three main criteria are used: the most minimal reliable solutions; the maximum current integral assessment of the disease; the maximum area of inaccurate decisions. When determining the leading symptom of selected diseases, the leading symptom corresponds with symptom having the maximum coefficient of specificity. At the generation stage of additional question based on the leading symptom, the patient should be asked simultaneous right questions to clarify quantitative estimates of possible symptoms.

Forward Chain Strategy

During literal implementation, the system checks the application of fuzzyproduction rules of each disease to each fact of symptoms in working memory, necessarily executes them and proceeds to the next disease, returns to the beginning after the exhaustion of all diseases.

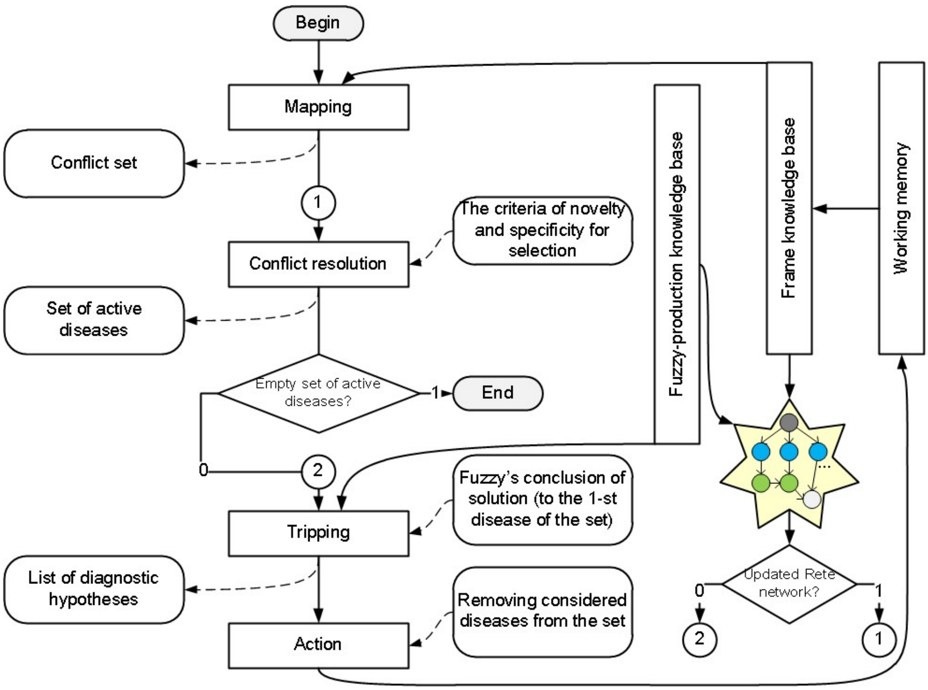


Fig. 2. The scheme of diagnostic output management solutions

Rete algorithm is used to ensure the speed at large knowledge base and a large number of facts in working memory (Fig. 2). When using the Rete algorithm, base of medical knowledge under translation is converted into Rete (or prefix tree) network, in the leaf nodes which are, on the one hand, procedures-demons, attached to the original slots, and on the other - procedures-methods to retrieve the values of the target slots when the truth of the premise of fuzzy-production rules, information which is stored in the intermediate nodes.

At the time of the assignment, all rules are not known under conditions of uncertainty. Therefore, a single network for all rules should not be built. This modification of the algorithm is called the Rete's fast algorithm [4]. In the modified Rete network the following components should be stored: activation list, where parent slots are stored, i.e. frames-prototypes slots; activation context, where links to current frames that caused the activation are stored, i.e. instance frames. When changing the value of some of the original slot located in the premise, all related procedures-demons that are directly trying to calculate the values of the target slot imprisoned are activated. Unification of rules with values in working memory as such is not made, and is replaced with implicit unification of inheritance, which is achieved through the call-demons procedure of all parent frames with the transfer of the current frame (caused by activation) as the calling context. Thus, the network implicitly established is attached to the demons slots, related to rules and fuzzy conclusion, in nodes whose intermediate results are remembered.

Comparison of symptom complexes with the available facts from working memory is performed after approval of incoming symptoms. As a result at this stage conflicting set consists of potential diseases.

Conflict resolution is carried out to select one or several most suitable diseases from the set of conflict. The result is a variety of active diseases, and also determines the order of their execution on the criteria of novelty and specificity. In addition, indistinctly production rules to existing facts should not be used.

Tripping. During tripping of values slots, ill-production rules associated with this slot are triggered, when assigned, i.e. in the left part of which the values for this slot are figured. The application of such rules matches a state change in the field of diagnostics. Fuzzy's conclusion of diagnostic solution on the Mamdani model includes the following steps [6, 7, 8]: fuzzification of the input variables, i.e., the conversion of quantitative estimates of symptoms to fuzziness; prerequisites aggregation, i.e. determining the degrees of truth prerequisites for each rule; conclusion activation, i.e. determining the activated membership function of assessment of the disease to a term of imprisonment for each rule; accumulation of conclusions, i.e. the receipt of the final fuzzy sets of each variable diseases; defuzzification of the output variables, i.e., the conversion of fuzzy variables diseases to the definition.

At the action stage, the update the state of the working memory is the result of triggering of ill-production rules. The composition of effective information includes not only the list of diagnostic hypotheses but also the list of active rules, and also possible medical specialists, to which the patient needs to contact. In addition, it ensures the possibility of the formation of questionnaires with responses to questions if the patient consents or waivers with diagnostic decision before exiting the system.

2.4 Explanation of Effective Information

Explanation of effective information is intended for storage of protocol output log, i.e. information about the behavior of the system of diagnostics of diseases. State memory is used to store log, each entry of which corresponds to one diagnostic solution based on previous symptoms.

2.5 Acquisition of Medical Knowledge

The acquisition of medical knowledge consists of the following 4 steps.

Building a Knowledge Base

At the stage of building a doctor's knowledge base, with the participation of an engineer by field knowledge, the area of medical diagnostic and state diagnostic are described in the form of a hybrid model of knowledge representation. Formalization of hybrid model of knowledge representation is a preparatory stage. Knowledge engineer creates the frame and ill-production model of knowledge representation. At the stage of forming the frame hierarchy, doctors describe the current state of the field of diagnostics as follows: definition of specialties, diseases, groups of symptoms, symptoms; creation of interactions between objects region; display of quantitative estimates of incoming symptoms and quantitative integral evaluation of each disease. At the stage of formation of fuzzy-production rules, doctors describe transitions between the states of the diagnosis as follows: description of input and output variables (symptoms and diseases); the work of membership functions for the generated variables; creation of a vaguely-production rules.

Knowledge Extraction from the Sent Questionnaires

This approach is based on the formation of responses to an email sent by the patient, according to the prepared questionnaire, and the extraction of new knowledge from them, describing its contents. In the capacity of knowledge source, diagnostic solution is formed by doctors for medical consultations, in which the system could not deliver the final diagnosis. The formation of the prepared questionnaire is the preparatory stage. To do this, select the list of the concepts necessary to describe the structure of the questionnaire (possible specialty; questions-answers; and others) then the hierarchy of concepts are formed based on their connections. At the stage of filling out the questionnaire and sending to doctors, all fields of questionnaire are filled in with the information of medical diagnostics. Later the questionnaire in the letter is sent to physicians on possible specialty. At the stage of receiving the letter and response to questions, the doctor decides to accept the letter and answer questions. The answered letter is sent to the patient's email address. At the stage of forming new fuzzyproduction rules, answered questionnaires with a knowledge base to retrieve new facts are compared, which are stored in the form of rules.

Learning of Fuzzy Knowledge Bases

To increase the effectiveness of learning ill-production model of knowledge representation, solving the task of forming a knowledge base, using genetic algorithm is recommended [9]. To do this, first you need to set the encoding/decoding illproduction model of knowledge representation that defines some parameters (membership function parameters; the coefficients of specificity and confidence), which are reduced in a single vector. The value of one parameter lies in a specific surrounding area, which can be broken into 2^16 intervals. Then to encode non interval, the 16-bit value in the code warming can be used, in which adjacent numbers are characterized by a lower number of positions. To create the initial population of chromosomes randomly, generation 100 chromosomes with the initial initialization values genes in a given area performed. Then use the operation of composition to combine a set of genes in a single chromosome for assessing the fitness of chromosomes. Each chromosome of the population is put in conformity assessment of its suitability chromosomes in the population, the calculation of which is based on the training sample and vectors of the model parameters. The learning process is considered as complete if the condition that the estimate is greater than the threshold value. In the selection process, based on the principle of the roulette wheel, the more sectors on the roulette wheel (i.e. corresponding assessment of fitness chromosomes), the more the chance that this chromosome is selected, which in the future after the decomposition operation genetic operators are applied to create descendants for the next population. Application of genetic operators in chromosomes. In genetic algorithms for the transfer of genes of the parents to descendants the crossing operator is responsible. Operators of mutation and inversion are designed to support a variety of chromosomes in the population. The formation of a new population. Effective chromosomes should be placed in the population after the composition operation. To reduce the population to its original number of chromosomes, the operator reduction is applied. After shutdown of the genetic algorithm, a trained model is obtained, which is approximated to a given accuracy data from the learning sample and forming knowledge base, consisting of a system of fuzzy-production rules.

Testing of the Knowledge Base

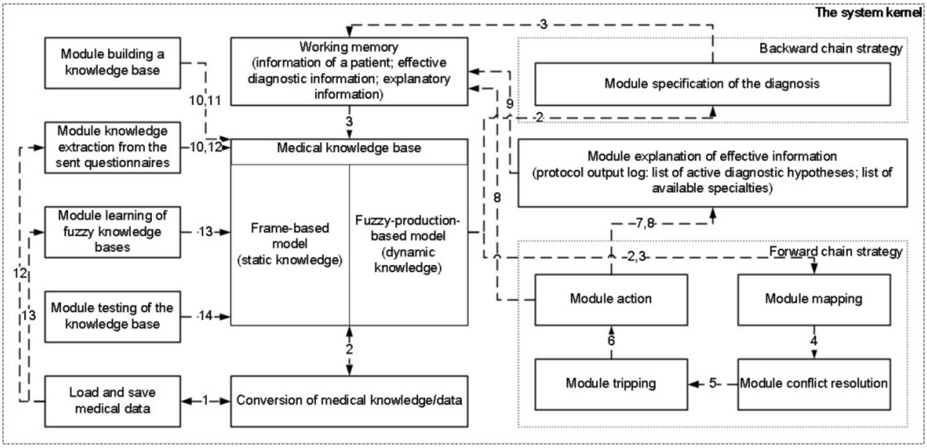
Testing of the knowledge base is designed to assess the completeness and integrity of the knowledge base by test sample [10].

2.6 User Interface

The user interface handles all messages exchanged between users and the system, and also performs roles of participants of the dialogue between them and the organization of their interaction in the process of medical consultations, and the acquisition of medical knowledge.

3 Information Structures of Medical Expert System for Differential Diagnosis

When designing the system, it is necessary to clearly distinguish the following levels: the presentation layer is intended for user interaction, information display and controls; logic layer is intended to implement the functionality of the kernel; the data access layer is designed to communicate with the data source used by the business logic layer. The system kernel is implemented as software modules. In Fig. 3, the structural scheme of the system kernel is represented.



1 - Medical information
2 - Medical knowledge
3 - Facts
4 - Potential disease
5 - Active disease
6 - Diagnostic hypothesis
7 - List of active diagnostic hypotheses
8 - List of possible specialists
9 - Effective information
10 - Knowledge and experience of doctors
11 - Frame-based and fuzzy-production-based model
12 - Sent questionnaires
13 - Learning fuzzy knowledge and learning samples
14 - Test samples

Fig. 3. The structural scheme of the system kernel

4 Conclusion

As an example of practical use of the developed model of the system, the process of medical differential diagnosis is described. Identification of quantitative expert evaluation of diseases and specialties on the basis of incoming signs of disease is needed. As evidence of diseases, 68 independent symptoms are applied, and the output - 89 diseases and 21 specialties. The developed system allows to increase the efficiency of medical diagnosis that she must not stop because of the fact that no part of the input information of the symptoms. Through a combination of ill-production and frame models of knowledge representation the problem of management of medical knowledge from different sources, having fuzzy, blurry structure is avoided.

References

1. Dvoryankin, A.M., Kizim, A.V., Jukova, I.G., Siplivaya, M.B.: Iskusstvennyy intellekt. Bazy znaniy i ekspertnye sistemy. VolgGTU. RPK “Politehnik”, Volgograd (2002)
2. Djekson Piter Vvedenie v ekspertnye sistemy. Per. s angl. Izdatelskiy dom “Vilyams”, M. (2001)
3. Djarratano Djozef, Rayli Gari Ekspertnye sistemy principy razrabotki i programmirovanie. Per. s angl. OOO “I.D. Vilyams”, M. (2007)
4. Soshnikov, D.V.: Metody i sredstva postroeniya raspredelennyh intellektualnyh sistem na osnove produkcionno-freymovogo predstavleniya znaniy: dis. kand. fiz.-mat. nauk. MAI, Moskva (2001)

5. Povoroznyuk, A.I.: Konceptualnaya model obekta diagnostiki v kompyuternyh sistemah medicinskoy diagnostiki. Sistemi obrob. informacії. Vip. 9, pp. 133–136 (2007)
6. Demenkov, N.P.: Nechetkoe upravlenie v tehnicheskih sistemah. Izd-vo MGGU im, M., N.E. Baumana (2005)
7. Lyu, B.: Teoriya i praktika neopredelennogo programmirovaniya. Per. s angl. BINOM. Laboratoriya znaniy, M. (2005)
8. Fomenkov, S.A., Davydov, D.A., Kamaev, V.A.: Matematicheskoe modelirovanie sistemnyh obektov. VolgGTU. RPK "Politehnik", Volgograd (2006)
9. Katasev, A.S., Ahatova, C.F.: Neyronechetkaya model formirovaniya baz znaniy ekspertnyh sistem s geneticheskim algoritmom obucheniya. In: Problemy upravleniya i modelirovaniya v slojnyh sistemah: Trudy XII Mejd. konferencii, pp. 615–621. Samarskiy nauchnyy centr RAN, Samara (2010)
10. Popov, E.V., Fominyh, I.B., Kisel, E.B., Shapot, M.D.: Staticheskie i dinamicheskie ekspertnye sistemy. Finansy i statistika, M. (1996)