Models and Technology for Medical Diagnostics

Vladimir A. Obraztsov(✉) and Olga V. Shut

Belarusian State University, 4 Independence Avenue, 220030 Minsk, Belarus
obraztsov@bsu.by, olgashut@tut.by

Abstract. This paper presents a novel decision support approach for the medical diagnostics in sports traumatology. In particular, the corresponding pattern recognition problem is solved by means of a special hybrid algorithm.

Keywords: Pattern recognition · Decision support · Medical diagnostics

1 Introduction

The decision support computer systems are widely used in the modern medical diagnostics. The specific implementation of such systems depends on the medical problem being solved as well as the core mathematical models and computer technologies.

As a rule, the mathematical tools, providing solution to the medical diagnostics problems, are related to the standard deterministic and statistic pattern recognition problems. The rapid development of computer-aided design technologies opened up new possibilities in this field, such as extended model integration, complex indeterministic information processing and structuralization.

The important area of research in medical diagnostics is integration of logical and precedent-related models of representation and derivation. This makes the extended functionality of computer diagnostics systems such as results validation and explanation possible. In many decision scenarios, it is important to provide the object comparison mechanism within the framework of different models. This can be achieved by eliminating some types of indeterminacy. Additionally, the transition from logical to precedent-related representation requires the construction of a large multidimensional binary objects' collection. Therefore, the information structuring is an important component in such systems, demanding particular features of the computer technologies applied. The paper proposes the solution to the decision making in medical diagnostics, while taking into consideration previously mentioned model features and information properties.

2 Medical Diagnostics as a Pattern Recognition Problem

Let us formulate a pattern recognition problem for the medical diagnostics.

Let $S_1, \ldots, S_n$ denote the set of features (symptoms). The features are used to specify patients, diseases, etc. in medical diagnostics problem. Thus, every feature must belong to a finite set of values provided by the problem context, i.e. the examination of patients.

V.V. Krasnoproshin and S.V. Ablameyko (Eds.): PRIP 2016, CCIS 673, pp. 150–159, 2017.
DOI: 10.1007/978-3-319-54220-1_16

As a result, every feature S_i ($i = 1, \ldots, n$) corresponds to a value from a set D_i. The mapping

$$x\colon S_1 \times \ldots \times S_n \to D_1 \times \ldots \times D_n \tag{1}$$

represents an object in space X, which is the space of possible descriptions of states of patients.

The sets of descriptions of patients can be grouped on different bases. One of the obvious bases in medical diagnostics is diagnosis, specifying the different states of patients (healthy, ill, etc.).

Let $X_1, \ldots, X_l$ ($l \in N$) denote the groups of diagnoses, satisfying the following conditions:

$$\bigcup_{i=1}^{l} X_i \subseteq X, \tag{2}$$

$$\forall i, j \in N \ \ (i \neq j \ \Leftrightarrow X_i \cap X_j = \emptyset), \tag{3}$$

The condition (2) means that the state of every patient belongs to the space of diagnoses. Obviously, several diagnoses can correspond to one patient.

Thus, the construction of space X is specified and a binary mapping on the set $X \times L$ is given, where $L = \{1, \ldots, l\}$. For each element (x, i) of the mapping on $X \times L$, it holds that $x \in X_i$, $i \in L$.

In practice, the decision-making is characterized by the level of information uncertainty. In the medical diagnostics, it causes the incomplete mapping specification $X \times L$ and therefore results in errors for some $x \in X$. The information uncertainty leads to other consequences, too, but the incompleteness of information is the most important factor. The following components can be undefined:

- for some $x \in X$ the second component isn't defined;
- for some $i \in L$ the first component isn't defined.

The latter case means that the description of partition of X into subsets $X_1, \ldots, X_l$ is redundant. One can eliminate it by means of removing the classes $i \in L$, for which $X_i = \emptyset$. As a result, the problem is reduced to the initial one.

In the first case, let us suppose that there exists $X^0 = \emptyset$, for which to each $x \in X^0$ there is the corresponding second component $i \in L$. It is clear that X^0 satisfies the condition $X^0 \subset X$ as well.

Now let us specify the construction of X^0. On the one hand, knowledge can be represented by the theoretical data (e.g. from a manual) and, on the other, can be deduced from the practice. In the first case, it contains rules or regularities, which control objects $x \in X$ represented as (1). In the latter case, it contains the descriptions of the same objects (1), for which the diagnosis has been verified and, as a result, a label $i \in L$ corresponds to each such an object. These options of the set X^0 description correspond to generally accepted ways of set descriptions, i.e., they are based on the convolution principle and direct specification of the set elements.

Let X^{01} and X^{02} denote the parts of X^0, constructed as mentioned above. Similarly to the medical practice, both ways of constructing samples X^{01} and X^{02} exist independently. Let us suppose that $X^{01} \neq \emptyset$ and $X^{02} \neq \emptyset$, and the following condition holds:

$$X^{01} \cup X^{02} = X^0, X^{01} \cap X^{02} = \emptyset \tag{4}$$

Let $I_0(X^{01}, X^{02}, X^0, X_1, \ldots, X_l)$ denote all information being considered in the problem context, using conditions (1)–(4), the methodology of building the subsets, etc. It is necessary to specify an algorithm (a rule or a method) of the following type in order to solve the decision problem:

$$\forall x \in X \; (A{:}x \times I_0(X^{01}, X^{02}, X^0, X_1, \ldots, X_l) \to \mathit{result}) \tag{5}$$

It should be possible to interpret the result in (5) in terms of classification $X_1, \ldots, X_l$.

The above problem is called as the medical diagnostics problem with a combined training set.

3 The Basic Technological Principles of the Solution

Let us formulate the basic technological principles for solving the medical diagnostics problems with a combined training set. Initially, the most general scheme of system design, information processing and components interaction is shown (see Fig. 1).

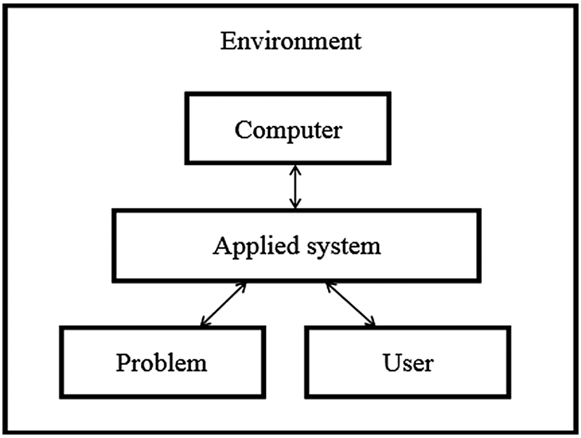

Fig. 1. The scheme of objects' interaction in the process of problem solution

This scheme does not depend on the problem domain and specifies the system components and the type of their interaction. The environment-specific limitations are common for any such system. The problem-specific features of the system are based on the interaction between the user and the information. The purpose of the applied system is to automate this interaction.

Therefore, in order to solve the technology development problem the detailed specification of the following level is provided (see Fig. 2).

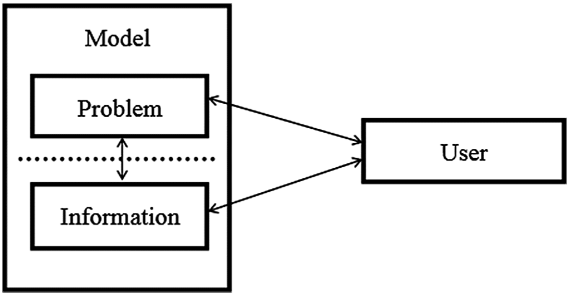

Fig. 2. The scheme of interaction between the user and information in the process of problem solution

Let's consider an arbitrary medical diagnostics problem with a combined training set, for which the type and properties of subsets X^{01} and X^{02}, the training and recognition processes requirements, the conditions for algorithms and the types of results are determined and the limitations to the cost and the ways of obtaining them are specified. Besides, the range and roles of users are defined. One needs to develop a technology for building computer systems that can provide the solution of the problem with the above limitations. The technology should:

- be based on the unified algorithmic kernel that provides possibility to use it in order to solve practical problems from different domains;
- contain means of formalized knowledge input and adjustment to a problem domain. They should include the possibilities of classes description with using different models, feature space description, training the system, etc., and provide the possibility to modify the knowledge base on the problem domain easily;
- provide the modularity of the system construction in order to guarantee the system flexibility, easy extension of its functionality, etc.;
- support the possibility to eliminate some types of indeterminacy that appear during the transition from logical to precedent-related representation of information and vice versa, and the structuring of prior information, directed at optimization of the solving problem process.

The solution of the above problem is provided at the theoretical, technological and applied levels.

- The theoretical level is specified by the effective implementation of the developed models and algorithms.
- The technological level includes problems, related to the development of computer systems for solving medical diagnostics problems with a combined training set. The

whole lifecycle of the information processing for the problems considered is provided.
- The applied level consists of using the developed library of components configured to the problem domain for solving the practical medical diagnostics problem.

For solving the above problem, a technology, which is based on the following idea, has been developed. The medical diagnostics problem has been previously considered in different definitions. In the Information Control Systems department of the Faculty of Applied Mathematics and Computer Science, a technology, which contains all necessary components and aims to solve the medical diagnostics problems, has been developed and successfully used. The implemented scheme of problem solution is different from the one proposed in this paper and is out of our scope. The idea is to extend the implemented technology in order to provide the possibility of solving the diagnostics problem and thus extend the functionality of the existing computer systems to a new level.

Several modules such as decision-making support modules, dialog organization and interface module have enhanced the technology. Below goes the brief introduction to the modules.

The following components have been added to the decision-making support module:

- Modules for transition from logical to precedent-related representation of information and vice versa.
- Object normalizing module.
- Algorithm training module.
- Module for optimizing the obtained solution based on branch-and-bound method.

The following components have been added to the dialog organization and interface building module:

- Coding module for scaled and continuous features.
- Explanation module.
- Interface module.

From the architectural point of view, the structure of the technological basis corresponds to the one that has been developed before [1–3]. The following steps outline its main modules operation.

The representation of a decision-making algorithm:

Stage 1. The analysis of information.

Step 1.1. The training set, which is described using logical model, is represented using precedent-related model.
Step 1.2. The description of classes (diagnoses) is also represented using precedent-related model.
Step 1.3. Normalization of objects.

Stage 2. The examination.

Step 2.1. The input data, including the description of objects and classes are formed (see Stage 1).

Step 2.2. The results of diagnostics are calculated and interpreted.

Stage 3. Stop.

The scheme of practical system usage:

Step 1. The information about the patient is entered (personal data and the results of the preliminary examination).
Step 2. The localization of the disease is determined.
Step 3. The diseased area is examined and additional symptoms are found.
Step 4. The diagnosis is determined.
Step 5. The results of the examination are represented in the user-friendly form.
Step 6. The treatment scheme is formed.
Step 7. Stop.

To verify the modules of training and optimization of algorithms, two mechanisms are used. The first one consists of executing the standard tests on the verified information. The second one is common in decision support systems and is based on the following idea: the inverse implication algorithms are implemented in the system and used to prove the monotony of direct deduction algorithms with the verified information. In order to implement these algorithms, a derivation that can answer the following questions is used in the explanation system:

(1) How many diagnoses correspond to the current examination of the patient?
(2) What is the maximal value of the diagnosis confirmation degree that can be obtained during the following examination of the patient?
(3) What are the symptoms that correspond to the diagnosis?
(4) What are the diagnoses that correspond to the chosen symptoms?
(5) What is the list of features, for which the value of the diagnosis confirmation degree can be equal to 1?

The questions 1, 3, 4, 5 in this list can be easily implemented. The tables that show the correspondence between features and diagnoses are used for this purpose. For the question 2, an additional implementation is needed. Two inverse implication algorithms are used to answer this question and are shown below.

The algorithm for binary features:

Stage 1. Obtaining the input data: the patient vector x, the diagnosis number i and the vector x^i from the training set, for which the current estimate for the diagnosis i is calculated.
Stage 2. Performing the procedure of building a new vector x': those features of vector x that have already been examined with the value 1, are extended by the remaining part of vector x^i.
Stage 3. The calculation of the upper estimate. Execute the following procedure:

Step 3.1. Perform the elimination by vector x'. As a result, a group containing diagnosis i is determined.

Step 3.2. Calculate estimates for all diagnoses from the obtained group. As a result, the upper estimate S' for diagnosis i is obtained.
Step 3.3. If this estimate is maximal in the group, output it. Otherwise, go to step 3.4.
Step 3.4. Replace all values 0 in vector x' by indefinite ones and calculate the lower estimate S for the obtained vector. Output estimate S.

Stage 4. Stop.

For all cases output the patients, for which this estimate is obtainable.

The algorithm for scaled features:

Stage 1. Obtaining the input data: the patient vector x and the diagnosis number i for the group of diagnoses containing i.

Stage 2. Calculating the lower S and upper S' estimates by a vector, which allows obtaining the maximal estimate g in the diagnosis i:

Step 2.1. Replace every feature, as soon as its value is 0 or indefinite, by 1. If the estimate g rises, perform the correspondent replace in the patient vector. As a result, a new patient vector is obtained.
Step 2.2. Calculate estimate S' for the given patient vector by class i.
Step 2.3. Calculate estimate S by class i and the patient vector that differs from the built one only by positions, which have value 0: one need to set an indefinite value in these positions.

Stage 3. Identifying the possibility to show the upper estimate S':

Step 3.1. If $S' = 1$, then output S'.
Step 3.2. If $S' < 1$, then extend the initial (current) vector by those values 1, that appeared during stage 2, and perform a new division into groups.
Step 3.3. Calculating estimates S and S' by all diagnoses. Afterwards, analyze the estimates to determine, which of the two estimates can be shown.

Stage 4. Stop.

As in the algorithm for binary features, for all cases output the patients, for which this estimate is obtainable.

In the process of solving the medical diagnostics problem the dialog organization and interface building module, the decision-making support module and some third-party modules like browser or Web-server are used. The sequence of their work, based on the developed technology, can be described as follows.

The scheme of using the technology for solving medical diagnostics problems:

1. The dialog organization and interface building module provides the dialog with the user.
2. The decision-making support module solves the problem using the knowledge file and the algorithmic kernel.
3. The dialog organization module outputs the results and their estimates.
4. Stop.

This scheme is shown in Fig. 3.

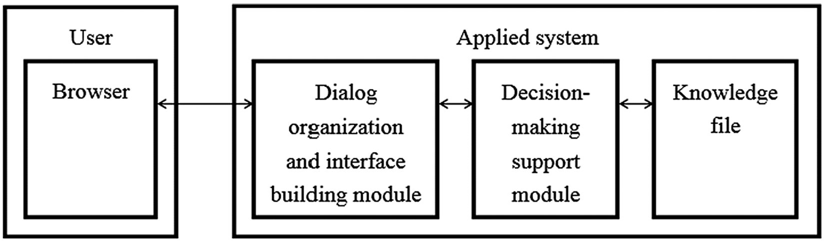

Fig. 3. The scheme of using the technology for the pattern recognition problem solution in medical diagnostics

Initially, the dialog organization and interface building module converts the requirements of the feature space to the questions addressed to the user and converts the answers to the feature values. The user can inquire the system about the additional information. This module is also used to output the solution as a list of found diagnoses with their estimates and allows using the means of explanation of the obtained results.

The decision-making support module represents the central part of the system and implements the algorithmic kernel. In the process of solving the problem this module controls the examination and calculates the estimates. The necessary information is obtained from the knowledge file, which is created during the development. In this file, features and classes are formally listed, the description of classes is given as a set of logical rules and training sets, additional options of algorithms are specified.

The standard methodology of using the algorithmic kernel consists of the following steps: first the algorithm is chosen and then its configuration and verification are performed. If the results are erroneous, the algorithms are corrected. The number of objects in the prior and test samples is usually limited by the size of information that is available to the researchers. Further, during the commercial operation stage, the base algorithms can be corrected in the case of receiving new information.

4 The General Description of System EXTRA

The system EXTRA has been developed based on the above technology for solving problems of diagnostics orthopedics diseases. In this chapter, the system is described similarly to [2–6]. The main purpose of the system is support of decision-making in the areas of orthopedics, sporting traumatology and rehabilitation, providing the doctor with the continuous informational help during diagnostics and choosing the treatment tactics based on general and special medical knowledge.

The base information in the system is the diagnosis with fixed one or several (alternative) treatment schemes. The final rehabilitation protocol is made depending on the current state of patient. The information is of a recommended character and can be corrected at user's discretion.

The architecture of the system is based on the client-server technology and distributed data processing. The client part includes end-user-oriented functions. The server components of the system provide database services and knowledge bases controlling.

Several users can work simultaneously. The architecture allows unification of the treatment and rehabilitation methods exchange of experience between users and scientific data mining with their further interpretation.

The main functions of the system are:

- Identifying all diagnoses including any given symptom/symptoms;
- Identifying all symptoms typical for this diagnosis;
- Identifying all symptoms that are typical for any given anatomical structure;
- Searching the description of the treatment methods based on the diagnosis and symptoms for a specific patient.

Figure 4 shows the general structure of system EXTRA.

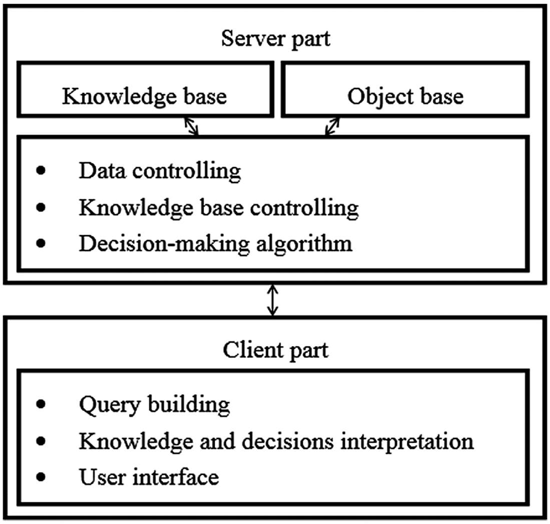

Fig. 4. The structure of system EXTRA

The process of solving, i.e. finding the diagnosis, is based on the following ideas.

The problem domain is divided into independent subsystems – localizations, where the decision support is made independently. There exist classical examination methods that allow obtaining the set of features, which is enough for the diagnostics. They are divided into subsets that in total describe some diagnosis. The localization is defined by the set of diagnoses. Each diagnosis is connected with a set of rules that determine some prior information for the diagnosis. To form the rules, logical operations are used. Each diagnosis corresponds to a set of possible actions for treatment.

The knowledge base includes descriptions of features, diagnoses, formulas, etc. The symptoms, which describe the state of patient, make a system of binary features. The features are ordered as a treelike structure, which corresponds to the order of the patient examination. The rules of calculating the diagnoses are stored as logical formulas using logical operations AND, OR and XOR in the feature space.

In the system EXTRA, the methods, which are based on conversion of the problem to a standard pattern recognition problem, are used. For this purpose, an algebraic structure of information is introduced. Localizations correspond to a list of diagnoses, a list of features and a map of examination, which represents a table that shows the usage of features in the descriptions of diagnoses. The part of training set, which is represented using the logical model, is converted to the precedent-related model. The resulting information is joined with the part that has initially been represented using the precedent-related model. Then a special pattern recognition algorithm is used on the joint information. The training set for this algorithm is built using the operations of the logic algebra. Thus, the solving of a problem with combined prior information is reduced to the solving of a pattern recognition problem that uses only the precedent-related model. The extension of the functionality of the system EXTRA has allowed improving the quality of diagnostics process while searching the diagnosis and choosing the treatment tactics in the area of sports traumatology.

5 Conclusion

The paper outlines a construction of algebraic base for the modeling and implementation of applied medical diagnostics support systems.

In addition, the component and functional structure of an artificial intelligence computer system EXTRA for diagnostics diseases is presented in the paper. The system implements the decisional kernel based on the algorithms for solving the pattern recognition problems with a combined training set.

References

1. Ablameyko, S.V., Krasnoproshin, V.V., Obraztsov, V.A.: Models and technologies for pattern recognition with application in data mining. Vestnik BSU. Ser. 1 Ph. Math. C. Sc. **3**, 62–72 (2011) (in Russian)
2. Krasnoproshin, V.V., Gafurov, S.V.: A program technology of building systems for solving pattern recognition problems with complicated structure. Artif. Intell. **1**, 30–37 (2008). (in Russian)
3. Krasnoproshin, V.V.: Theoretical Basis and Technologies for Constructing Recognition Systems by Precedence. Dr. Sc. Thesis 05.13.17. Minsk (2007) (in Russian)
4. Ablameyko, S.V., Krasnoproshin, V.V., Obraztsov, V.A.: The experience and main results of information control systems department (Belorussian State University) in the field of theory and practice of pattern recognition. In: Proceedings of Crimea International Mathematics Conference (CIMC-2013), vol. 3. Simferopol (2013) (in Russian)
5. Ablameyko, S.V., Krasnoproshin, V.V., Obraztsov, V.A.: Pattern recognition and image analysis: theory and experience of solution of practical problems. In: International Congress of Computer Science: Information systems and Technologies, pp. 434–444. BSU, Minsk (2013) (in Russian)
6. Krasnoproshin, V., Lositskiy, Ye., Obraztsov, V., Vissia, H., Gutnikov, S., Popok, S.: Intelligence system of decision support in sporting traumatology. Bulletin of NTU “KhPI”. Comput. Sci. Modeling **31**, 106–111 (2010) (in Russian)