
Medical and Diagnostic Systems

SYNTHESIS OF STRUCTURED MODELS OF COMPUTER SYSTEMS IN MEDICAL DIAGNOSIS

Anatoliy Povoroznyuk, Nataliya Bilous, Olga Kozina

Abstract: *Method for structural identification of objects of diagnosis in design of computer systems to support decision-making in medicine at all stages of information transfer is developed. This takes into account not only the structural and functional bases, but also the uncertainty of the model parameters.*

Keywords: *medical diagnostics, computer system, structural identification, the model, decision rule.*

ACM Classification Keywords: *1.5. Pattern recognition and J.3. Life and medical sciences*

Introduction

The introduction of new information technologies in medicine gives rise to transition from traditional medical information retrieval systems to intelligent computer systems to support decision-making in medicine (ICSSDM) with advanced mathematical apparatus and elements of expert systems that are capable of monitoring of diagnostic criteria, build an advanced computer diagnoses, implement and management the treatment process.

Developers of ICSSDM denote following actual optimization problems that require urgent solutions:

- distinguishing of diagnostically significant structural elements in the background of artifacts and interferences (structural identification) in the analysis of biological signals, and medical imaging;
- formalization, the evaluation of informativeness of diagnostic criteria and formation of the minimum set of informative features;
- synthesis of decision rules according to which process of diagnosis (classification) of patients inside a given set of diagnosed conditions has carried out.

Application of traditional mathematical methods (deterministic methods, probabilistic, logical-and-linguistic, based on pattern recognition theory) for solving optimization problems in medicine, which was noted before, does not entirely correct [Ахутин 2002, Поворозньук 2006] because the complexity and multiple-level interactions between diagnostic objects (subsystems of the organism), the heterogeneity of diagnostic criteria (numeric, ranking or dichotomous data) and the heterogeneity of approaches of their obtaining, for example, by questioning, examination, clinical studies, the analysis of biological signals and medical imaging are not accounted for these methods.

In traditional methods diagnostic criteria are discussed in the form of a linear vector, and algorithms for automatic distinguishing of structural elements of biological signals are designed for each type of signal, and have heuristic nature. In addition, such methods are critical to volume of training set and have limitation on the dimension of objects of diagnostics in their implementation. In this paper, methods for constructing ICSSDM which based on unified formal approach of structural identification of objects of diagnostics at all stages of information transformation are proposed.

Principles of structural identification of objects of diagnostic

Proposed method of structural identification is based on conception of evolutionary identification of structured models which are effectively used for solving the problems of forecasting, recognition, management, etc. [Букатова 1990]. As known, the sense of structured C -models is the transformation $F : X \rightarrow Y$ or, in other words, reflection of input vector X to output Y via a set of operators f from the set F ($f \in F$) in accordance with structure S , which determines the sequence of operators f . Thus, the C -model is given by the graph, nodes of which are the functional elements of a given class of models, and information connection of nodes is defined by structure S , i.e. $C = \{f, S\}$. Synthesis of C -models is a recurrent and a stochastic process for consistent improvement of C -models, evaluation of C -model due to some criteria Q , and the selection of locally effective C -model. Classes of C -models, their functional and structural bases, as well as the modes of change are described in [Букатова 1990].

Taking into account the above optimization problem in the construction ICSSDM, heterogeneity of diagnostic criteria, an iterative process of diagnosis and the need to using of expert assessments in the formation of diagnostic output, the following stages of information transfer in ICSSDM according to structural identification have been formalized:

- structural identification of physiological signals and images;
- formalization of the description of heterogeneous diagnostic criteria;
- synthesis of hierarchical structure of diagnoses;
- synthesis of hierarchical diagnostic criteria;
- synthesis of diagnostic rules;
- synthesis of individual diagnostic prognostic models;
- recommendations on choosing the optimal treatment.

At each of considered phases local optimization problem is decided for which its own set of input data, own criterion of optimality and optimization algorithm must be used. That is why expanded notion of C -model (C' -model) and the procedure for reconfiguring the model is proposed to use in this work.

Extended C' -model is defined as follows

$$C' = \{S, F, E, \varepsilon_\sigma, \varepsilon_\Delta\}, \quad (1)$$

where – $S = \{P, V\}$ – structure model specified through a set of nodes P and arcs V ;

F – functional basis of the model where the functions can be attributed as nodes f_p so and arcs f_v ;

E – expert estimations which if necessary can be added to components F ;

ε_σ – the uncertainty of parameters of F -basis that is determined by statistical properties of training set;

ε_Δ – the uncertainty of parameters of F -basis depended on accuracy of their determination (step of quantization).

Value ε_σ is used not only as an estimate of the model parameters (traditional approach), but is a parameter of optimization procedures in the synthesis model. Since some algorithms of optimization require quantization of the model parameters, then the task of selection the type of quantization (uniform or nonuniform) and selection of corresponding steps of quantization arises, thus ε_Δ can be considered like a parameter optimization procedures also.

Consider the application of the method of structural identification in the implementation mentioned earlier stages of information transfer in ICSSDM.

Identification of structural elements of biological quasi-periodic signals

Structural identification (allocation of the structural elements of the signal parameters which are diagnostic features) of quasi-biological signals (ECG, rheograms and others) is the most responsible and difficult phase of their processing, as well as errors in identification of structural elements (omission of structural elements or false identification) leads to gross errors in calculating of diagnostic criteria and errors in computer diagnosis (in cases when special means of exposure of these errors don't used, in particular, when the processing of suspected period is refused or confirmation of human-operator about correct structural identification is missed). The difficulty of identification is fact that the amplitude and temporal characteristics of structural elements holds information about the diagnosed subsystems of organism and have wide range of variability their values. In addition, these signals may include artifacts and can be registered against the background noise. That is why formalized procedure for identifying the structural elements of biological signals on the basis of individual space of parameters for each type of structural element using the methodology and computational procedures of Hough transforms proposed in this work [Поворознюк, 2003].

A classic application of Hough transforms is the approximation of contours points of two-dimensional monochrome images by analysis curves of first or second order. Thus, each i -th point of the image is transferred to the space of parameters (parameters are the coefficients of approximating curve), and it forms a subspace of admissible values of parameters M_i . Intersection area Ω of subspaces M_i for all contour points determines the true value of parameters [Toronto, 2007]. Hough transforms is a basis for implementing various types of recognition and classification algorithms whose main idea is to move into more informative space of parameters for this task. In the structural identification of physiological signals conversion from the original space (set of discrete points of the signal $X(t_i)$) in the minimum-required space of parameters $Y(P)$ is executed. Space of parameters $Y(P)$ is formed at description of structural element pattern by a limited set of basic reference functions (approximation of pattern). So comparison of analyzed signal with pattern and the decision on the presence or absence of a structural element of a given type is performed in the parameters space.

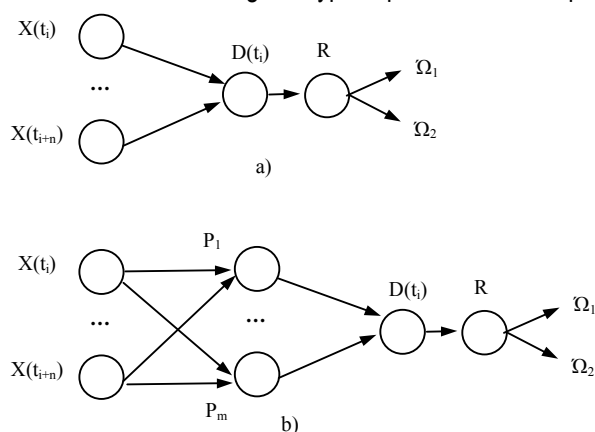


Figure 1. C' -model representation of structural element recognition:
a) in contour analysis; b) in proposed method

In considered cases scanning of whole signal and signal part analysis with an aperture n , which is equal to the length of the structural element, are performed. In contour analysis through complex algorithms for each type of structural elements in each type of biological signal function $D(t_i)$ which reflects the distinctive features of given structural elements, is constructed. Then function $D(t_i)$ is compared with boundary R and decision about whether is the part of signal structural element (class Ω_1) or not (class Ω_2) is accepted. In the proposed method, at first, the operator mark out a pattern's structural element of this type in the signal in learning mode.

After that the formalized procedure for approximation of the pattern by graphic primitives is started, and parameters of individual space with dimension m are formed by applying Hough transforms. In such case $D(t_i)$ is a function of differentiation of distance, in other words, each its point is value of remoteness from the pattern in parameters space $Y(P)$. Since parameters space $Y(P)$ must be discrete, the value ε_Δ is optimized in such way that center of the cluster Ω_1 has been the level of quantization and its effective diameter has been the step of quantization ΔP_i .

Adaptation of identification of two-dimensional diagnostic structural elements is result of description of two-dimensional patterns by different types of spatial adaptive developments where transition from original rectangular system of coordinates i, j ($X_{i,j}$ – values of brightness in rectangular system of coordinates) into the space I (X_I – values of brightness along the line scan) is performed. In contrast to traditional arrangement of structural identification of two-dimensional images (filtering, contouring, the search for structural elements), proposed approach give abilities to define pattern directly to the grayscale image.

Synthesis of hierarchical structures of diagnostic criteria S_X and structures of diagnoses S_D based on their self-informativeness

Synthesis of hierarchical structures of diagnostic criteria S_X and diagnoses S_D based on their self-informativeness is considered like hierarchical clustering based on the analysis of correlations through transformations $F_1: X \rightarrow S_X, F_2: D \rightarrow S_D$. To implement the conversion F_1 the procedure of hierarchical clustering of diagnostic criteria based on representation of clustering in view of a streaming model is proposed. Thus, initial characteristics become nodes of fully connected graph. Arcs of such graph can be equivalents of pair correlation coefficients. Then, task of hierarchical clustering of nodes is reduced to a consistent procedure for finding of the minimum cut of the graph in the streaming task with limitations. To solve streaming task adaptation algorithm of "defect" is proposed.

In streaming tasks with limitations each arc is characterized by the following parameters:

- f_{ij} – flow on arc (i, j) ;
- L_{ij} – lower bandwidth arc (i, j) ;
- U_{ij} – upper flow capacity of arc (i, j) ;
- C_{ij} – the cost of passing the unit flow from node i to node j .

For considered problem of clustering of diagnostic criteria of the algorithm on the basis of algorithm of "defect", value L_{ij} is equal 0, and U_{ij} is equal to criterion of self-informativeness (pair correlation coefficient). If apriori information is absent, then for all arcs C_{ij} can be equal 1 or their values can be received from expert estimations in contrary. Values of streams f_{ij} are determined when the algorithm of "defect" is completed and initially may be equal to 0. To reduce the dimension of the problem it is possible to consider only arcs for which self-informativeness criterion is significant in the sense of Student's criterion.

For implementation of algorithm of "defect" it is necessary to define initial and finish node (source S and sink). Results of its work are the calculation of streams circulation which minimized the total cost of streams on all arcs according to limitations of arcs capacity (L_{ij} and U_{ij}). The correct choice of source and sink is heuristic task, therefore, at first estimated clustering must be carry out by method of galaxy correlation, and as source and sink are chosen the most connected nodes in each cluster. After determining of streams on all arcs the minimum cut of graph R is determined, for which the true relationship:

$$\begin{aligned} f_{ij} &= U_{ij} \quad \forall (i, j) \in (N_c, \overline{N_c}), \\ f_{ij} &= 0 \quad \forall (i, j) \in (\overline{N_c}, N_c). \end{aligned} \quad (2)$$

The minimum cut divides all nodes of the graph into two disjoint sets ($\overline{N_c}$ include an initial node of S and $\overline{N_c}$ include and finish node T), thus, streams saturate all direct arcs of cut and zero arcs of cut in inverse direction (2). In this case, the sum of streams of all arcs of the cut equals the maximum circulation of stream.

Result of hierarchical clustering is hierarchical structure, on lower level of hierarchy into which separate diagnostic criteria, if they are informative and independent, or clusters of correlated features can be present.

Synthesis of hierarchical system of diagnoses – transform F_2 – is performed on a similar way of submission problem like streaming model with limitations and finding the minimum cut with help of algorithm of "defect". However, there are the following differences:

- top capacity of arc i, j – value U_{ij} – is intercluster distance of diagnoses D_i and D_j in space of criteria;
- dichotomic clustering is performed without compression, so the above criteria for linear clustering and optimality in this procedure is not used.

Reconfiguration of structures of diagnostic criteria S_x

Structures of diagnostic criteria S_x is obtained by the criterion of self-informativeness of criteria and can not be used for creating of additional diagnostic specification directly because the resulting structure S_z must contain minimum necessary amount of informative diagnostic criteria for given diagnostic problem, i.e. given set of diagnoses. The structure must be submitted with structure S_D and procedure of clusters replacement at all levels of hierarchies S_x should be developed. Therefore, reconfigure of structure S_x is proposed in order to create the structure S_z . Criteria of diagnostic value of parameter x_i with regard to given set of diagnoses $\{D\}$ can be used for this purpose [Поворознюк 2007, 1]. Diagnostic value $I_p(x_i)$ is amount of information introduced into the system after patient observation to parameter x_i . For all that transformations $F_3: S_x \rightarrow S_y$ and $F_4: S_y \rightarrow S_z$ are completed.

If diagnoses set are formed whole group of incompatible events – only one diagnosis D_i correspond to one patient in training set, no patients with multiple diagnoses), the uncertainty of diagnoses set is estimated entropy:

$$H(D) = - \sum_{i=1}^n P(D_i) \cdot \log_2 P(D_i), \quad (3)$$

where $P(D_i)$ – a priori probability of diagnosis D_i .

Expression (3) show estimation of maximum possible amount of information that can be introduce in set of diagnostic criteria, i.e. for any system of diagnoses D and any set of diagnostic criteria $X = \{x_1, x_2, \dots, x_p\}$ inequality is correct:

$$I_D(X) \leq H(D), \quad (4)$$

which becomes an equality only for set of deterministic parameters.

Transformation F_3 determine informativeness of original space of criteria X according to the system diagnoses set $\{D\}$. For this, clusters of lower-level hierarchy of structure S_x – subsets of correlated features – are

replaced to most informative and other clusters of remaining level of S_x are replaced to integral features. In addition, in synthesis of S_y set of heterogeneous elements S_x is given to a single scale – every original parameter x_i is represented like disjoint set of diagnostically significant intervals.

Changing range of numerical value $\Delta = [x_{j\min}, x_{j\max}]$ for sequence of intervals is in fact the replacement of theoretical law of distribution for histogram. From this point of view to obtain an acceptable accuracy of approximation of distribution law, the number of intervals must be large enough. But it is necessary to increase training set, because at constant sample size and increasing the number of intervals relatively small number of points gets in each interval, which reduces the reliability of statistical estimates. Thus, with limited training set, the task of dividing Δ on intervals is optimization task in which integral error ε_k is minimized [Поворозняк 2007, 2]:

$$\varepsilon_k = \frac{\Delta_k}{\sqrt{6N_k}} \sqrt{\left(\frac{G_{k+1} - G_k}{\Delta_{k+1} + \Delta_k}\right)^2 + \left(\frac{G_k - G_{k-1}}{\Delta_k + \Delta_{k-1}}\right)^2}, \quad (5)$$

where m – number of nonuniform intervals Δ_k ;

N_k – number of points from training set which get in interval Δ_k ;

G_k – ordinate of the histogram into interval Δ_k .

Calculation of ε_k due to (8) requires apriory dividing Δ by Δ_k and calculation of values inside not only in the current k -th interval – Δ_k and G_k – but also in neighboring: $\Delta_{k-1}, G_{k-1}, \Delta_{k+1}, G_{k+1}$. Minimum of ε_k can be reached by iterative procedure for formation of Δ_k .

The goal of transformation $F_4 : S_y \rightarrow S_z$ is agreement of topologies S_z and S_D in order to complete the method of diagnostic specification and to ensure the optimal plan for individual patient diagnostic observations. Thus, each pair of S_D with total parent has agreement with one element of S_z . Elements of structure S_z are formed from elements of S_y on the basis of their informative completeness and diagnostic value so that they could perform a differential diagnosis for each level of diagnoses hierarchy.

Set of diagnostic criteria X relative on diagnoses set D is described by coefficient of informative completeness $k_{IC}(X, D)$:

$$k_{IC}(X, D) = \frac{I_D(X)}{H(D)}. \quad (6)$$

Similarly, coefficient of informative completeness for each parameter x_j is determined:

$$k_{IC}(x_j, D) = \frac{I_D(x_j)}{H(D)}. \quad (7)$$

Since process of diagnostic criteria measurement requires to using medical equipment and some resources (time, material, financial, etc.), given the complexity of combined measurement, each parameter is characterized by the coefficient of diagnostic value:

$$k_{DV}(x_j, D) = \frac{k_{IC}(x_j, D)}{r_i(x_j)}, \quad (8)$$

where $r_i(x_j)$ – the total ratio of measuring complexity of parameter x_j .

The formulation of S_Z begins from top-level element, i.e. from element $Z_{1,2}^1$, which should provide a differential diagnosis between states $D_{1,1}^1$, and $D_{2,1}^1$.

Initially entropy diagnoses $D_{1,1}^1$, and $D_{2,1}^1$ is calculated by (3) and required informativeness of $Z_{1,2}^1$ is calculated by (6) for a given ratio of measuring complexity. It should be noted that informativeness of any element of S_Z does not exceed the 1bit and reaches its maximum for equiprobable diagnoses, in accordance with (4).

Then coefficients of diagnostic values of all elements of S_y for pair of diagnoses $D_{1,1}^1$, and $D_{2,1}^1$ is calculate by (8) and elements of S_y is organized in accordance with

$$k_{DV}(y_i) \geq k_{DV}(y_j) \geq \dots \geq k_{DV}(y_q), \quad (9)$$

For each element of S_y coefficient informative completeness for pair of diagnoses $D_{1,1}^1$, and $D_{2,1}^1$ is defined, and then sequential process of including of elements y_i from ordered sequence (9) into $Z_{1,2}^1$ is repeated until not reached the necessary level of informative completeness of element $Z_{1,2}^1$.

In order to formulate elements of next level of hierarchy $Z_{1,2}^2$ and $Z_{2,3}^2$, elements informativeness of which has been exhausted during formulation of $Z_{1,2}^1$ are excluded from structure S_y .

Thus, every element of S_Z is ordered subset of elements of S_y and uniquely specifies order of their inclusion in the diagnostic procedure of diagnostic specification at each phase of implementation.

Conclusion

Thus, the methods of synthesis of computer systems to support decision-making in medicine based on the structural identification of objects of diagnosis is developed. Conception of extended structural model takes into account not only structural and functional bases, but also the uncertainty of parameters and also allows using expert estimations.

Construction of hierarchical structures from 9 parameters of clinical blood analysis and 10 different diagnoses are completed for 434 patients. The average probability of diagnostics has increased from 88.48% when using a standard cluster analysis to 92.8% when using the developed method. From the total number of negative responses differentiation for 25 negative responses to the preliminary diagnoses of different levels is received while using only 3.61 diagnostic parameters an average for each patient, that in 2,5 times less than the original 9 parameters which was used for standard procedures of diagnostic.

The developed methods allow removing restrictions on dimension of diagnostic criteria space, to improve the reliability of computer diagnosis and to adapt to specific objects of medical diagnosis.

Bibliography

- [Ахутин 2002] В.М. Ахутин, Шаповалов В.В., Иоффе М.О. Оценка качества формализованных медицинских документов. Медицинская техника. № 2 (2002), pp. 27-31.
- [Поворознюк 2006,1] А.И. Поворознюк Формализация этапов проектирования интеллектуальных компьютерных систем медицинской диагностики. Электронное моделирование. Т. 28, №1 (2006), pp.85-97.
- [Букатова 1990] И.Л. Букатова, Ю.И. Михасев, А.М. Шаров Эвоинформатика: теория и практика эволюционного моделирования.-М.:Наука, (1990), pp.212.
- [Поворознюк 2003] А.И. Поворознюк Применение преобразования Хока для структурной идентификации физиологических сигналов. Моделювання та інформаційні технології. Збірник наукових праць інституту проблем моделювання в енергетиці. Вип. 22, (2003), pp.143-149.
- [Toronto, 2007] N. Toronto, B.S. Morse, D. Ventura, K. Seppi. The Hough Transform's Implicit Bayesian Foundation. IEEE International Conference on Image Processing, ICIP 2007. Vol. 4, (2007), pp.377-380.
- [Поворознюк 2007, 1] А.И. Поворознюк Реконфигурация структуры диагностических признаков при синтезе компьютерного диагноза. Системи обробки інформації. Вип. 2(60), (2007), pp. 135 – 139.
- [Поворознюк 2007, 2] А.И. Поворознюк Формирование диагностических интервалов численных признаков при дифференциальной диагностике. Вісник Хмельницького національного університету. №3, Т.1, (2007), pp. 106-109.

Authors' Information

Anatoliy Povoroznyuk – Ph.D., lecturer of Computers and Programming Department of National Technical University 'KPI', Kharkov, Ukraine; e-mail: koa@kture.kharkov.ua

Nataliya Bilous – Ph.D., head of ITLCVS laboratory, professor of Kharkov National University of Radio Electronics, Kharkov, Ukraine; e-mail: belous@kture.kharkov.ua

Olga Kozina – Ph.D., lecturer of Computers and Programming Department of National Technical University 'KPI', Kharkov, Ukraine; e-mail: koa@kture.kharkov.ua