Embedded Selforganizing Systems

Issue Topic: "Embedded Systems and Its Application in Medical and Biomedical Field"

The Human Innate Nature Based Medical Expert System for Self-Diagnosis

Michael Gennadievich Grif

Faculty of Automation and Computer Engineering
Novosibirsk State Technical University
Novosibirsk, Russian Federation
E-mail: grifmg@mail.ru

Yumchmaa Ayush

Faculty of Automation and Computer Engineering
Novosibirsk State Technical University
Novosibirsk, Russian Federation
Power Engineering School
Mongolian University of Science and Technology
Ulaanbaatar, Mongolia
E-mail:yumchmaa@must.edu.mn

Abstract— This article considers the some results of the diagnostic text in the medical expert system, which based on integration of the "european" and "eastern" medicines. One of the proposed work directions is the study of the patient's questionnaire informative value, where the questions of a prototype with binary answers "YES-NO" are performed as corpuscular diagnostic testing (CTD). Then, in determining the real disease (diseases), there arises an actual task of selecting the most preferred CDT at each step of the diagnostic processes. CDT is an indivisible diagnostic procedure, which estimates one or more indicators of the condition of the body. In addition, using by real determination of human inborn innate nature gives opportunity to understand internal causes of chronic diseases. Also, for selection more probabilistic diseases have been applied the Bayesian approach and the conditional entropy.

Keywords— *medical expert system (MES); european and eastern medicines; bayesian approach; conditional entropy; a priori probability; a posteriori probability; symptoms and diseases*

I. INTRODUCTION

Medical decision support system is intelligent information aid system for physicians, including data of diseases and their related symptoms. On the specified symptoms, system provides list of more probabilistic diseases, based on the medical statistical data of symptoms of disease on results of laboratory's analyses and clinical practices. Also such as systems contain different reference information, which helps to physicians to make the plan of medical examination and treatments. For the diagnosis, systems use to diagnose information about patients: gender and age data; subjective symptoms (patient complaints). And, some systems need results of objective examinations, laboratory analysis and instrumental methods; risk factors (data of anamnesis vitae), too. As distinct from the analogues, system doesn't require big part of information about patient data. The first patient information about gender and age, symptoms, which is inserted by physicians, allows doctor to focus their attention on the most probabilistic diseases. Decision-making problems in the complicated medical situations are considered: time deficiency, low patient information; limited availability of reference literatures [1].

Generally, doctor makes a decision on the own experiences and laboratory analyses when he is working with patients. Additionally, in the practice of European medicine, more considers to individual symptoms of diseases.

In principle, difference of the eastern medicine from the western medicine is consideration of any diagnose based on whole bodies as complex unified system (innate nature of human). Proper identification of any diagnosis should consider basics of the human's lifestyle and healthy conditions. Regarding to eastern medicine methodology such as information will give sign to doctor to estimate human diseases and to offer more correct treatment for human. Then, in determining the real disease, there arises an actual task of selecting the most preferred CDT at each step of the diagnostic processes. Model of the choice of information sources on the next step of diagnostic process is considered. The criterion of the CDT choice is supposed to be its diagnosticability (informative value), allowing a number of formalizations. For example, residual entropy for couple "the diagnosis – its denial" ($D_i, notD_i$), $i = 1, n$, where n -number of the

diagnoses considered by the doctor or the likelihood ratio for information formed by CDT.

In the proposed model of information choice diagnosticability estimation procedure (informative value) can be performed by doctors as a vector, in general, for all admitted diagnosis and for individual diseases ("YES-NO"), previously ordered by the degree of non-increase of complication of disease.

If it necessary, for specification of the probabilistic diagnoses during the intelligent dialogue with doctor - user the system recommends to receive different type of additional data from patient.

For this purpose, it forms various types of morphism, which reflect by the preferable diagnostic links between the seven basic constitutional types of the human body (terminology of the Eastern medicine) and the set of "European" diagnosis (terminology of modern medicine).

II. DIAGNOSTICITY PROBABILITIES AND TYPES OF CTD AND VALUATIONS

The diagnosticity of CTD can be estimated using the likelihood ratio [2, 3]. The assessment form will depend on the following types of CTD:

1) Binary test;

2) Verbal CTD in visual valuation;

3) Quantitative CTD.

A. Binary test

It shows how well the diagnostic test determined the presence or absence of the disease in patients with confirmed or rejected the diagnosis.

TABLE I   TOTAL AMOUNT OF PATIENTS WHO GAVE ANALYSE

		Disease		
		Presence	*Absence*	
Test	Positive	a	b	a+b
	Negative	c	d	c+d
All amount of patients		a+c	b+d	

From the table I can to calculate the sensitivity (1) and specificity (2) of diagnostic test, when (a+c) is the total numbers of all sick and (b+d) is healthy patients.

$$(Se) = a/(a+c) \tag{1}$$
$$(Sp) = d/(b+d) \tag{2}$$

Sensitivity is the proportion of positive test results in the group of sick patients. Specificity is the proportion of negative test results in the group of healthy patients.

Using these values, we can calculate the likelihood ratio for a positive result:

$$LR+ = Se/(1-Sp) \tag{3}$$

Similarly, for a negative result:

$$LR- = (1-Se)/Sp \tag{4}$$

Sensitive test often gives a positive result in the presence of the disease (detect it). However, it is particularly

informative when it gives a negative result, because rarely misses patients with the disease. A specific test rarely gives a positive result in the absence of the disease. It is particularly informative if the result is positive, confirming the (assumed) diagnosis.

The likelihood ratio for a positive test result is sensitivity divided by 1 minus specificity. Thus, the likelihood ratio reflects both a sensitivity and specificity of the test.

$$LR+ = Se/(1-Sp) = \left[a/(a+c)\right]/\left[b/(b+d)\right] \tag{5}$$

$$LR- = (1-Se)/Sp = \left[c/(a+c)\right]/\left[d/(b+d)\right] \tag{6}$$

Whereas, the likelihood ratio of a positive test result was equal to 1, it means that the probability of a positive test, the patient is the same as the probability of a positive test result in healthy one.

One of the great benefits of likelihood ratio is that they can help to go beyond the rough classification of test results (norm-pathology) facing if describe the accuracy of the test only in terms of sensitivity and specificity at a single split point.

B. Verbal CTD in visual valuation

Verbal CTD in visual valuation of the likelihood is not on individual fragments of a picture or ECG (values CDT), but on the ability of the test to separate two hypotheses: the presence of the disease and its absence. Or, information I is realized not a result of CDT, but utilizations of the efficiency of results have any output data. Accordingly, the likelihood is being estimated not for a specific report of CDT, and the test as a whole. For example, $P\left(ECG/D_i\right)$ and $P\left(ECG/notD_i\right)$, $i = 1, n$, which means:

- Probability of detecting disease, where it exists
- Probability of detecting disease in the absence of the same ECG.

There are no fundamental prohibitions on making the worst decisions on the basis of the same information. For the purposes of correct determining the diagnosticity of CDT in this group is necessary to postulate its best use.

C. Quantitative CDT

It allows for the discretization. For example, values of common blood test or biochemistry involving three verbal grades:

- Normal
- Below normal
- Above normal

In this group of CDT need to move from a relationship of credibility for specific values of the test and fixed the diagnosis to the likelihood ratio (LR) for the test as a whole (see Table II).

TABLE II   LIKELIHOOD RATIO FOR ALL VALUES OF CDT

№	Table Column Head	$P\left(I_k/D_i\right)$	$P\left(I_k/notD_i\right)$
1.	I_1	$P\left(I_1/D_i\right)$	$P\left(I_1/notD_i\right)$
2.	I_2	$P\left(I_2/D_i\right)$	$P\left(I_2/notD_i\right)$

№	Table Column Head	$P\left(I_k/D_i\right)$	$P\left(I_k/notD_i\right)$
k.	I_k	$P\left(I_k/D_i\right)$	$P\left(I_k/notD_i\right)$
n.	I_n	$P\left(I_n/D_i\right)$	$P\left(I_n/notD_i\right)$

In this case, eliminate zero probabilities. Evaluate the distinguishability of hypotheses as the ratio $P\left(I_k/D_i\right)/P\left(I_k/notD_i\right)$ for each row of the matrix.

If, from this calculate the expected value of LR for each value of the indicator CDT, then it will be diagnostic ability of CDT in whole.

III. RELATED WORK WITH MATRIX OF "SYMPTOM-DIAGNOSIS"

In the table SD is consisting of a set D of diagnoses. (columns of the matrix) and S symptoms (the matrix rows). In cells of this matrix placed information on the likelihood of a specific symptom under fixed disease.

Each diagnosis with its denial is a group of hypotheses,

where $\{S_j\}, j = \overline{1,m}$ is a set of symptoms.

For any S_j we assume, that the symptom is there,

or it is not - $\overline{S_j}$.

Introduce the following notation:

$P(D_i), i = \overline{1,n}$, is the prior probability of the diagnosis established before any symptoms. $P(\overline{D_i}) = 1 - P(D_i)$.

$P\left(\dfrac{S_j}{D_i}\right)$ – the conditional probability of implication

symptom if many symptoms are observed at the patient;

$P\left(\dfrac{D_i}{S_j}\right)$ - a posterior probability of diagnosis for existing

symptoms.

Two options were considered, when diagnoses:
- alternative;
- compatible.

The conditional probabilities reflecting credibility of receiving at the patient of the answer to a question of existing symptoms (absence) on condition of existing (absence) diagnosis, is defined only for those symptoms which are informative at probability recalculation of the diagnosis in Bayesian formula. For example, S_j ="ear pain", D_i ="hemorrhoids" is uninformative symptom for diagnosis. Evaluation of a prior probability occurs:
- based on the statistical probability of the diagnosis (can be taken from the published medical statistics);

- using the knowledge base to open for free access;
- the values of the likelihood of symptomatic information to calculate posterior probabilities of diagnoses given the following four.

$$\left\{P\left(S_j/D_i\right)\right\}; \left\{P\left(\overline{S_j}/D_i\right)\right\}; \left\{P\left(S_j/\overline{D_i}\right)\right\}; \left\{P\left(\overline{S_j}/\overline{D_i}\right)\right\}$$

Posterior probabilities for all pairs of "diagnosis-symptom" with any answer to the questions are estimated using Bayes rule. If the symptom used as a diagnostic criterion, with the first question, the Shannon amount of information contained in the answer, then we have two particular conditional entropies (residual) [5]:

$$H\left(\frac{D_i}{S_j}\right) = -\left[P_{apost}\left(\frac{D_i}{S_j}\right)\cdot\log P_{apost}\left(\frac{D_i}{S_j}\right) + P_{apost}\left(\frac{\overline{D_i}}{S_j}\right)\cdot\log P_{apost}\left(\frac{\overline{D_i}}{S_j}\right)\right] \quad (5)$$

Similar to:

$$H\left(\frac{D_i}{S_j}\right) = -\left[P_{apost}\left(\frac{D_i}{S_j}\right)\cdot\log P_{apost}\left(\frac{D_i}{S_j}\right) + P_{apost}\left(\frac{\overline{D_i}}{S_j}\right)\cdot\log P_{apost}\left(\frac{\overline{D_i}}{S_j}\right)\right] \quad (6)$$

It is necessary to know $P\left(S_j\right)$ and $P\left(\overline{S_j}\right)$ to find the average conditional entropy.

The algorithm of choice the question by the criterion of maximal informative symptom is following.

1. For all D_i estimates the average conditional entropy for the answers yes and no, this will be the rating of diagnosticity of symptoms (response) at the first question.
2. In the next step, take the average for each row of the Table II. These are the estimates of the diagnosticity (informative value) the answers for the entire set of diagnoses;
3. Choose a question with maximum informative value
4. Estimate posterior probabilities of diagnoses based on responses in real time and proceed to item 1, excluding used question. As the a priori probabilities are previous step of an a posteriori.

After examining the human constitution by Eastern medicine test (Tibetan medicine), it's distinguished among a number of diagnoses a group of diseases, and increased the coefficient of initial probabilities [4].

IV. THE ARCHITECTURE OF MES

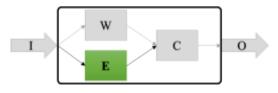MES architecture consists of the main parts (see Fig. 1).

Fig .1   Architecture of MES

Here we donate, I - input of application, W - western medicine diagnosis, E - eastern medicine diagnosis, C - comparison unit and O - output of application.

The MES has both: database and knowledge database. Knowledge database stores rules based on eastern medicine methodology. Rules related to innate nature of human for making decision for diagnose. Innate nature of human idea discussed in open source of many books about eastern medicine (Tibetan medicine). The fundamental concepts of the Tibetan medicine are to keep the three bodily humors in balance (wind, bile, phlegm). Main architecture bases in one fundamental structure of request processing part and switches off some obligatory parts. In literature on expert system, three key components described. All these components are in each expert system: knowledge database (KDB), inference engine (IE) and user interface (UI).

The MES can work in two different modes:

- Mode A (see Fig. 2). Here we used specific questionnaire (3x26) which we developed by innate native of human. We are donating questions as row and innate nature of human as column. Main decision based on the number of selected column answers.
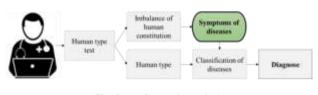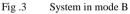

Fig .2   System in mode A

- Mode B (see Fig. 3). Here user (target groups) would be used as the questions, and the answer would be a diagnosis of the user's disease.

Fig .3   System in mode B

The questionnaire for determination of human constitution by 26 questions is affording informative results about constitutions, which based on own special characteristics of human, then likelihood data about balance of three basic humors (see Fig. 4) [5].
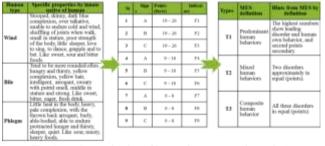

Fig .4   Determination of human innate nature in mode A

Suppose that in our example, the information received will give the following directions. For example:

- User selected simultaneously facts F1 AND F8 OR F9 then it should be find out type T1.

$$F1 \cdot [F8 + F9] \rightarrow T1$$

- User selected simultaneously F2 AND F7 OR F9 without F1 AND F3 then it should be make evident type T2.

$$F2 \cdot [F7 + F9] \cdot \overline{F_1} \cdot \overline{F_3} \rightarrow T2$$

- If all 3 facts in equal, that's mean type T3.

$$[F7 \cdot F8 \cdot F9] \rightarrow T3$$

In many ancient books, there are numbers of scripts about the diseases, their cause and treatments, which divided according to these three basic constitutions. Example, human with constitution wind much more suffers from the diseases of nervous and cardiovascular systems, and human with constitution bile suffers from diseases of digestive system. In addition, human with constitution phlegm suffers from the diseases of the lymphatic systems etc [6-8].

CONCLUSION

In the decision-making process in the field of medicine, particularly to receive with incredulity of diagnosis may be quite often requires additional examination from doctors who subjectively evaluates the symptoms of illnesses based on his knowledge and experiments. The iterative methodologies of European and eastern medical diagnosis probability will be increase. Moreover, the result of treatment can be more positive, too. Additionally it gives opportunity to doctors to realize internal causes of some chronic diseases of patient. For this goal of analysis in MES, we use Bayesian formula to determine various probabilities over 4 different situations. Such as test should help people to avoid early panic relating to the some strong diagnoses or it will be help to doctors and medical students to learn more about eastern medicine to make decision for diagnose.

REFERENCES

1. S.F. Bagnenko, "Modern principles of organization of emergency medical care: The report of the meeting of the RF Ministry of health on the issues ambulance", Saint-Peterburg, Russia, 2002, unpublished.
2. B. R. Shegal, "The information choice model in emergency medical diagnostics", in The collection of scientific wokrs, Novosibirsk: Press of NSTU, Russia, 2013, vol. 2, pp. 68-77.
3. T. A. Vlasenko, and B. R. Shegal, "Diagnostic criterion for the model of emergency medical diagnostics", in The collection of scientific wokrs, Novosibirsk: Press of NSTU, Russia, 2014, vol. 2 pp. 48-57.

4. M.G. Grif, B.R. Shegal, Yu. Ayush, and S.V. Yastrebova, "Diagnostic test design for the medical decision support system "Aimedica" based on integration of the "european" and the "eastern" medicines", IEEE R8 Russia Siberia Section, pp. 209-214, September 2017 [International Multi-Conference on Engineering, Computer and Information Sciences, SIBIRCON-2017, p.209-214, 2017].

5. Yu. Ayush, U. Tudevdagva, and M.G. Grif, "Expert system based diagnostic application for medical training", in The proceedings of international conference "Creativity in Intelligent Technologies and Data Science, CIT&DS 2017", Volgograd, Russia, 12-14 September 2017, pp.750-762.

6. S.G. Choijinnimaeva, Diseases of nervous people, or where blows wind. Moscow: Astral Press, Russia, 2010.

7. S.G. Choijinnimaeva, Diseases of big people, or what is the phlegm. Saint Peterburg: Piter Press, Russia, 2016.

8. S.G. Choijinnimaeva, Diseases of strong people, or how bridle bile. Moscow: Astral Press, Russia, 2010.