

Specifics personalized approach in the analysis of medical information

N. Melnykova¹, U. Marikutsa²

¹ *Lviv Polytechnic National University, e-mail: melnykovanatalia@gmail.com,*

² *Lviv Polytechnic National University, e-mail: umarikutsa@gmail.com*

Received February 7 2016: accepted April 12 2016

Abstract. In this article suggest some new approaches to solving the problems of medical data analysis and their personalization. To accomplish this was proposed to create a decision support system to the execution of sequence stages of analysis of patient's data.

The main stages of development and design of decision support systems that enable to make decomposition of control process and describe the relationship between input and output control flows. Applying the theory of decision trees during construction of decision trees of decision support system is due to the formation of a sequence of questions asked by the doctor when searching an individual approach when choosing a treatment.

Decision tree creates a hierarchical structure of rules. This approach allows you to present the logic of sequence issues by doctor in solving the medical problem history and it makes possible to simulate decision making process by physician when selecting treatment scheme.

Search the target value of output of decision medical support system makes it possible to select top of graph system that is located with more probability on the best way to the target. Important step in addressing the process of personalizing treatment schemes is estimated function that is based on Bayes theorem. Weight of occurrence next event corresponds to the highest value of the posterior probability of occurrence of the next state, given the time-dependent input parameters.

Proposed improved method of decision-making for personalization standard schemes by modifying the method of decision-making based on decision trees considering relationship between the input parameters and evaluation function and result of its works is a personalized therapeutic scheme of treatment.

It analyzed the quantitative results of applying the proposed method and existing for determining personalized schemes.

Key words: decision support system, data formalization, therapeutic treatment schemes, personalized of data.

INTRODUCTION

The problem of personalizing data is one of the important problems that arise in the development, design and improvement of information systems in all fields of human activity. Personalizing makes it possible to use standardized methods of processing information of

particular objects to their individual characteristics. Medicine is a branch of human activity where personalizing plays an important role in providing valid medical care to patients with any disease. Consequently, personalized medicine a certain model of health care for people based on the selection of diagnostic, therapeutic or preventive measures that would be optimal for the chosen people because of its genetic, physiologic, biochemical or other features. The main goal of personalized medicine is optimization and personalization of the prevention and treatment to avoid undesirable side effects by identifying the individual characteristics of the human body.

The development of personalized medicine is increasingly affecting the direction of the development of medicines. This is why pharmaceutical companies to change their paradigm of research and development, in addition to changing the process of deciding on whether to use a drug for the treatment of a particular patient. Although the development of personalized medicine requires more resources and more organizational changes, in addition, most of these changes should be carried out outside the pharmaceutical industry. It promotes the study of personalized medicine approaches in the analysis of individual patient data, development of methods of treatment assignment and personal disease risk assessment, taking into account the quality of life, and forming recommendations for pharmaceutical production allowing for the patient. For realization this approach need the creation and implementation of new technologies of personalized medicine.

THE ANALYSIS OF RECENT RESEARCHES AND PUBLICATIONS

By using current mathematical techniques such as Bayesian approach, the method of cluster analysis, method of reverse output developed expert systems like Hastrohraf, Whonet, Casnet et al., which are aimed at processing the received medical data presented in the form of spectral characteristics results maintaining a database of patients. Some large medical centers are trying to develop "intelligent" programs that allow diagnosis or narrow range of diseases in the differential diagnosis or to determine the minimum set of diagnostic tests such as expert system installation diagnoses Internist, BPLab, SpeseLabs Medical, Meditech, A & D, Omron, CardioVita [1,13,18].

Application of semantic networks in decision making implemented through relationships between objects, may make it possible appearance of conflict between elements of classes of pharmaceuticals and diagnosis [8,14,15].

stages of treatment, in the decision making process for formation of recommendations of therapeutic schemes for specific patients with different pathologies and other diseases.

OBJECTIVES

The main objective is to develop new approaches concerning personalizing of medical information in the process of support medical decision for determine the

THE MAIN RESULTS OF THE RESEARCH

The process of decision making support can ensure that the priority requirements for data management that determine its practical expediency and effectiveness.

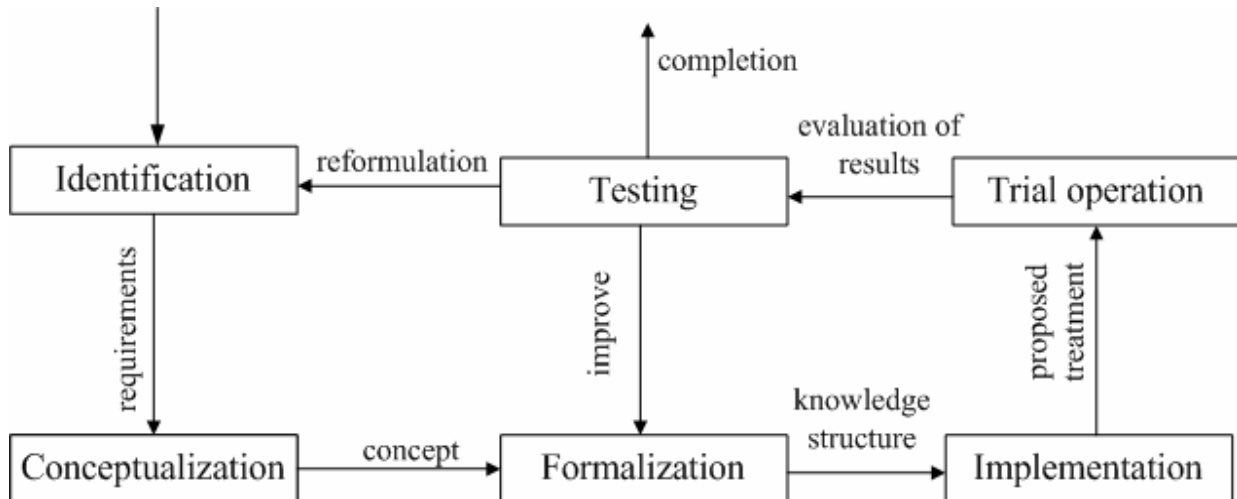


Fig. 1. Stages of development and design DMSS

Data integrity is achieved by preserving the correctness and accuracy of the knowledge base. To balance the conflicting demands knowing the general requirements of all facilities required to structure the knowledge base so that the service was generally better. To this purpose apply public medical ontology (Galen to determine the clinical condition, UMLS for the National Library of Medicine USA, ON9 for certification of health of the general parameters as well as engineering, chemical ontology, Suggested Ontology for Pharmacogenomics, Symptom Ontology, NCI Thesaurus, Ontology of Medically Related Social Entities, Ontology for General Medical Science, etc.) containing information about the hierarchical diagnosis and treatment of diseases, represent an area of knowledge in health care, pharmaceutical industry, etc. [2,9,10,12]

Providing universal access to data makes it possible to access to existing applications and databases to develop new applications to work with these same data. That is, new applications can access to most of these data, and there is no need to create new data [7].

The decision medical support system (DMSS). One approach solving the problem of personalization is to create DMSS that comes down to the execution sequence of stages analysis of patient's data:

- Identification of the patient as an object system,
- Conceptualization of data that determine the time-dependent and time-independent object parameters,
- Data formalization, converting data into quantitative indicators,
- Implementation process of processing the entire data set.

The main stages of development and design of decision support systems are presented in the form of transitions state diagram that enable to make decomposition of control process and describe the relationship between input and output control flows.

When developing DMSS often need to reconcile the results from previous phases and evaluate decisions taken there. The result of the development stages DMSS is the answer to the question what should be done and what resources are necessary involve, namely [3,11,15]:

- Identification tasks,
- Determine the participants of the design process and their role,
- Identify resources and target,
- Definition ways of presenting of all knowledge types,
- Formalization of basic concepts,
- Definition of search strategy decision,
- Modeling of the system work,
- Evaluating the adequacy of the system of targets selected methods, concepts and tools,
- Filling the knowledge base by expert,
- Verification the validity and competence DMSS,
- Formulation of requirements to DMSS.

The scheme of processing the input patient's data characterize by main stages providing the following steps in the DMSS (Figure 2).

It is known that the main source of knowledge for DMSS is an expert. Therefore, to create the medical DSS great importance is the modeling logical doctor reasoning. As a first approximation, logical doctor reasoning can be considered as process of operating with "medical

memory" input information for which is the patient's characteristics. This approach "medical Memory" is the primary and logical reasoning - secondary. However, the considerations can have a significant impact on the structure and nature of "memory care."

Due to the lack single algorithm description of the patient's state, diagnostic and therapeutic processes should consider traditional methods of information processing and decision making. It should be noted that until last time most widely used so-called tabulated algorithms, based on calculations using relationship.

Decision tree is one of the most popular approaches to solving problems of Data Mining. Objectives arising from the objective needs have hierarchic character [6,16]. Top-level objectives cannot be achieved until the goal of next lower level will be solved. The objectives are

specified in the move down the hierarchy of levels. In the process of building an expert system by means of decision trees should seek clear and specific formulation of objectives, provide for a quantitative or ordinal assessment of the degree of their achievement.

Construction of DMSS decision trees when using this approach is caused by the formation of a sequence of questions being asked the doctor when seeking an individual approach when choosing a treatment.

Decision tree creates a hierarchical structure of rules, classified on a "if - then" that looks like a tree [4,17,20]. Each vertex of this tree is a specific question that to put to a patient and a branch of tree that comes from the top responsible on the alternative answers and lead in its turn to new issues.

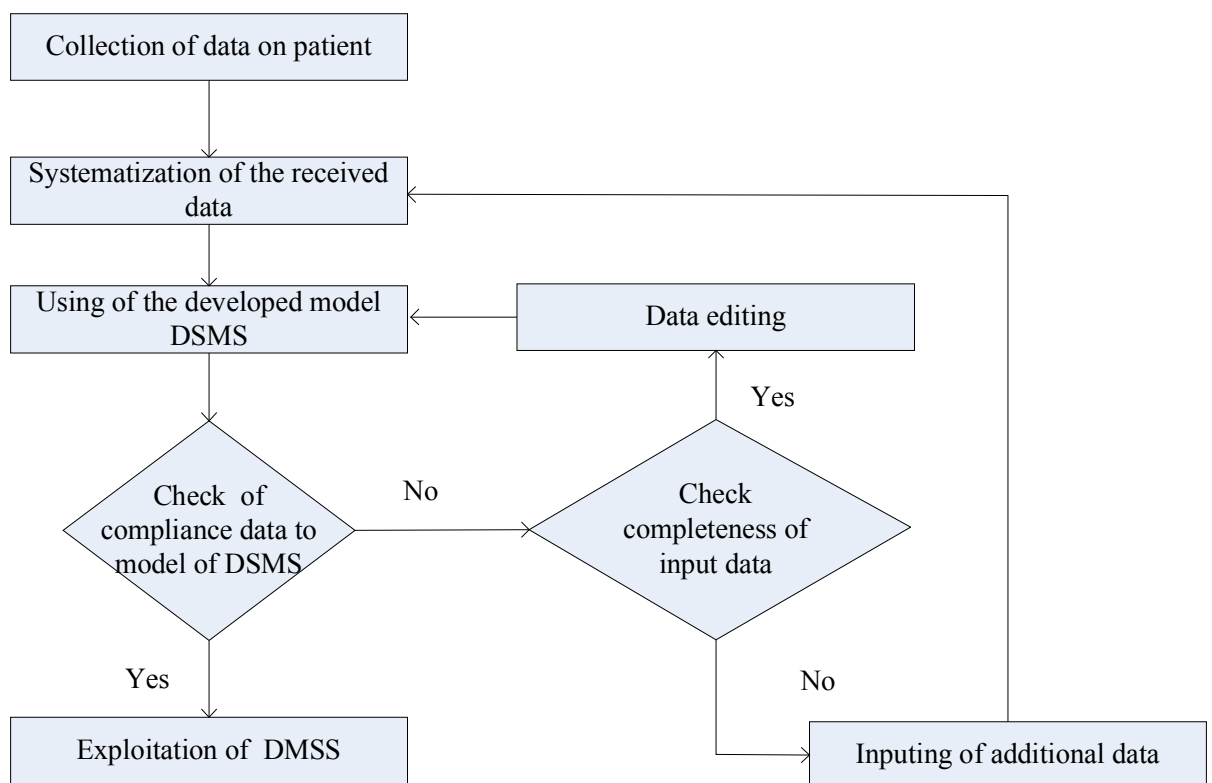


Fig. 2. Scheme of processing the input data of patient

The question may be, for example, the following: "Is applied prior antibiotic therapy?" If answer is yes, there is a shift to the right node of the next level, if it is negative - to the left node. Then again put a question related to the corresponding node. The program moves from question to question as long as no solution is found or been exhausted possible transitions.

Main stages of medical decision making shown at the transition of states diagram of decision medical support system (Figure 3).

The disadvantage of this approach is as follows:

- when trying to build a similar tree to solve complex the medical problems the number of vertices of

branching becomes so great that the analysis of the logical tree detected extremely difficult,

- slightest changes made to the application logic, leading to the need to re-build tree and reprogram the entire task.

At the same time, this approach is very convenient because it allows you to present the program logic assembly sequence issues by doctor in solving the medical problem history taking in clinical conditions. This approach makes it possible to simulate decision making process by physician when selecting treatment regimens.

Medical decision-making by using of statistical data processing is the employ methods of mathematical

statistics. It is based on processing large volumes of information collected on diseases that are subject to machine diagnostics.

Processed information can be used in different ways.

Some the medical expert systems based on the use of pattern recognition theory. It is necessary to have specific case histories with known diagnoses.

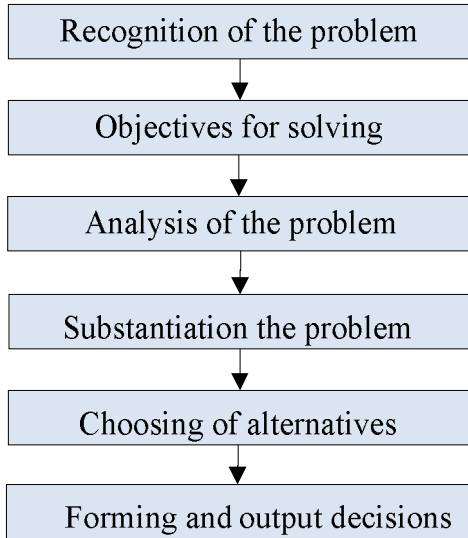


Fig. 3. Main stages of the decision-making medicines

The data are analyzed to determine the statistically "typical" treatment schemes for each disease. Appointed

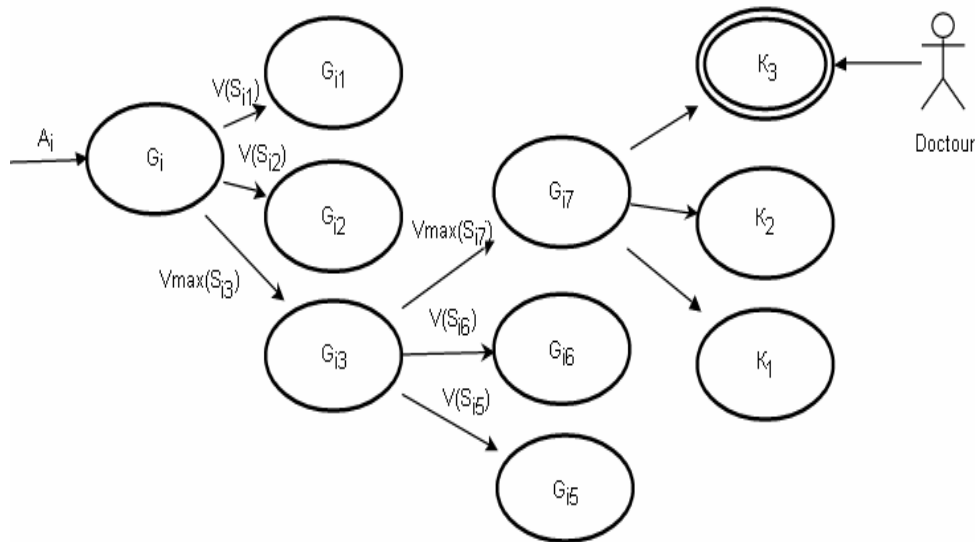


Fig. 4. Search target value of outputs system of decisions medical support.

However, the decision tree (tree classification) cannot be used when the classification parameters are dependent. When analyzing the patient's state cannot be said about the independence of the patient's parameters, even if there

are no obvious signs of this. drugs that are more characteristic of him, based on the information collected, processed by statistically. Case of diseases used in the analysis of the history of each patient to determine how "similar" case considered the "typical". In assessing the difference between comparable cases, the program generates a decision about treatment.

$$A \subseteq B \Rightarrow Z, \tag{1}$$

where: A - set of time-dependent data of the patient's state, B - set of characteristics of the drug research, Z - set of source data that is relevant scheme of treatment.

We propose visualization of decision-making process when selecting patient treatment schemes where set productions and output data are organized in some system of conservative treatment presented as a decision tree. A fragment of the tree selection schemes from the tops - drugs z_1, z_2, \dots, z_{12} and ribs - the rules of $p_{12}, p_{13}, \dots, p_{8,15}$, which performed the selection, shown in Figure 5.

Definition of conservative treatment schemes in the analysis set time-dependent and time-independent parameters $\Psi = \{S, A\}$, where $A = \{a1, a2, a3\}$, $S = \{s1, s2, s3\}$ and related productions.

In applying this approach for processing large arrays of information does not result in a balanced decision tree that is height is obtained very different, it slows down the process of obtaining valid information about the treatment.

are no obvious signs of this.

So for selection of personalized treatment schemes should be modified algorithm for constructing decision tree and its properties.

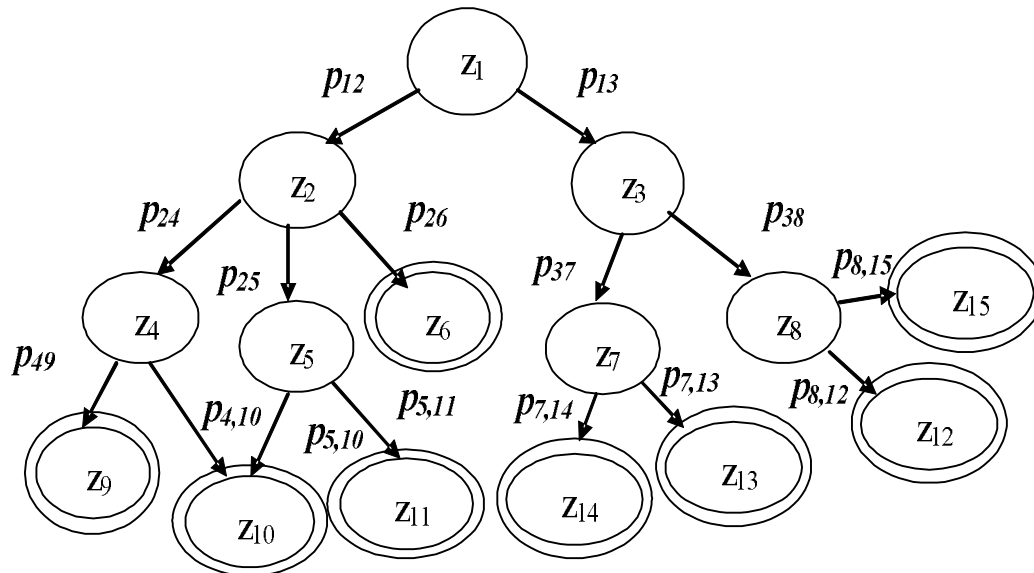


Fig. 5. The decision tree in finding therapeutic treatment schemes

Search the target value of output of decision medical support system. To assess the occurrence of the next event in the DMSS we suggest a proprietary method that will provide orientation when choosing target schemes of treatment.

This makes it possible to select top of graph system that is located more likely on the best way to the target. So placing the vertices should be in order of increasing values of $V(\Psi)$.

Another important step in addressing the process of personalizing treatment schemes is estimated function $V(\Psi)$, which is based on Bayes theorem. Weight of occurrence next event corresponds to the highest value of the a posterior probability of occurrence of the next state, given the time-dependent input parameters (S), ie,

$$V(S) = \max(p(G|S)) \tag{2}$$

In determining the outputs function taken into account the probability of all states according to the input parameters, it gives the doctor-expert whole range of possible therapeutic schemes on which it makes the conclusion of an optimal treatment scheme of the patient, taking into account his personal data.

As a result of applying the theory ordered search algorithm and evaluation function definition, characterized by probability of occurrence next state in the presence of specific system parameters, gives orientation in choosing the appropriate targeted treatment schemes.

In Figure 4 shows a piece of finding tree of targeted solutions by using evaluation function.

So, using the Bayes formula, we obtain the posterior probability of an event G of an event at S :

$$P(G|S) = \frac{P(S|G)P(G)}{P(S)} = \frac{P(S|G)P(G)}{P(S|G)P(G) + P(S|\bar{G})P(\bar{G})} \tag{3}$$

where: all values are known.

So, we have a priori the probability $P(Z)$, contained in the knowledge base. But getting input parameter S and the probability listing on Bayes formula, we can write it in place of $P(G)$. When there is a new incoming parameter this leads to a new update (increase or

decrease) of this probability. Each time the current value of the probability will be considered as a priori when applying Bayes formula.

Arises the problem of management logical conclusion, namely in the Knowledge Base has a set of hypotheses that system states

$$Z_1, Z_2, \dots, Z_m,$$

and the final set of indicators (input time-dependent parameters):

$$S_1, S_2, \dots, S_m.$$

Each hypothesis Z , has a corresponding subset associated with its parameters (input parameters). According to the strategy ordered search algorithms in the space of states system of medical decision-making can consistently processing the full list of possible parameters. In the end, counting all the hypotheses that take into account the emergence of all possible input parameters, the system comes to the end result, but it is more probability to assigned to the value of evaluation function $V(S)$, which is a defining for determining the next state of the system.

According to this we offer a practical example of finding the next state of the system.

Example: Suppose there is some state G and the patient has a diagnosis of furunculosis. It is believed that in accordance with the statistical data is known a priori probability $P(G)$ that the patient has a diagnosis furunculosis. Let S means that the patient who has a defined diagnosis furunculosis, has renal failure.

So evaluation function $V(S)$, defined using the Bayes formula, enables to get the probability that when diagnosis furunculosis the patient has renal failure of specified input parameters. In order to use the Bayes formula, you need to know probabilities:

$P(S|G)$ - probability that the a patient has renal failure at diagnosis furunculosis specified input parameters,

$P(S|\bar{G})$ - probability that a patient hasn't renal

failure at diagnosis furunculosis specified by input parameters.

So all three numbers $P(G), P(S|G), P(S|\bar{G})$ - received in advance based on previous statistical research and have a universal character. So $P(\bar{G}) = 1 - P(G)$.

Can use Bayes formula for determining $V(\Psi)$, for which all parameters are known.

Suppose

$$P(G) = 0,001, P(\bar{G}) = 1 - P(G) = 0,999;$$

$$P(S|G) = 0,9; P(S|\bar{G}) = 0,01.$$

Then using the formula, we get:

$$P(G|S) \cong 0,083.$$

So $P(G|S)$ - the probability that when patient has a diagnosis furunculosis at renal failure, increased relative to initial priori probability $P(G)$.

The method of decision-making for finding personalized the standard scheme. In recent years actively using ID3 algorithm and its modifications C4.5, See5 for construct decision trees [5,19,20]. These algorithms build a tree and generate rules based on examples.

ID3 algorithm builds a decision tree from the root. The roots of the tree and each of its domestic top will be treated in accordance with certain test set checks. For building inspections will use the fact that the value of each conditional attribute allows you to split the set into subsets of examples in which all examples have the same value for this attribute. If recursively apply this fact,

putting in line the inside top of each tree that build a received as a result of breaking the set, for each of the subsets obtained will build a new subtree. The process building of subtrees will continue until the division will not go subset of elements of the same class with the same values attribute decision making. Each class meets decision tree leaf. Since the order in which the tops of trees ascribe attributes to test, is important for building a decision tree, the result will depend on the classification of attributes assigned to root and to inside the tree tops.

ID3 algorithm is using growth criterion information (information gain), or a decrease in entropy (entropy reduction) to select attribute. The degree of increase information calculated as

$$Gain(S) = Info(T) - InfoS(T), \quad (4)$$

where: $Info(T)$ - entropy of set T to breaking, $InfoS(T)$ - entropy of after breaking S.

Increase of information quantity obtained by dividing set into subsets T_1, T_2, \dots, T_k by means of to splitting S. As best attributes for use in a partition of set (S) selected the one attribute that provides greater gain information ($Gain(S)$).

Criterion Gain is calculated for all the independent variables, and then selects the variable with a maximum value of Gain. Selects the variable to partition on it at one of the classes had a greater probability of occurrence. This is possible when the entropy of $InfoS$ a minimum value and therefore the criterion Gain (S) reaches its maximum.

Practical application of classical ID3 algorithm associated with a number of problems specific to models

based on learning and decision trees in particular. One disadvantage of ID3 algorithm is that it is incorrect attributes with unique values for all objects of the training set. For objects such information entropy is zero and haven't new data from the built tree with this dependent variable will not be able to receive as obtained after splitting subset will contain one object. To effectively overcome the shortcomings of ID3 finalized, resulting in its expansion was named C4.5.

C4.5 algorithm solves this problem by introducing normalization. Evaluated number of objects is not one or another class after splitting, and the number of subsets and their power (number of items). However, the problem of processing also entirely independent parameters remains [4,14].

We propose improved method of decision-making for personalization standard schemes by modifying the method of decision-making based on decision trees considering relationship between the input parameters and evaluation function. The result of the method works is a personalized scheme.

At first formed a set of pre-drug therapy Z_{prev} on the basis of pre-processed items set of output signals Z mealy machine, that characterize standard schemes of treatment, where g_0 - initial state:

$$Z_{prev} = \beta(g_0) \quad (5)$$

I. If the top of making medical decisions tree delivered in line with the set drugs Z and each determines $a \in A$ and $s \in S$ then this top is a piece of tree. At the end the function definition f selects personalized drug scheme with set Z of time-independent a_i and time-dependent s_j parameters of the patient's state.

II. If the top stage I is not satisfied, it is considered a set of stated time-dependent parameters S_l . The current top - G_i . If the pre-set drug therapy $Z_{prev} \neq \emptyset$, where Z_{cur} - the current set of drugs is defined as the difference between the sets Z and Z_{prev} .

III. If $Z_{prev} = \emptyset$ to step II, then to the top, which should be a leaf applied P production rules which takes into account the elements of Z drugs set, a subset of time-independent and time-dependent patient's parameters and identifies the personalized treatment scheme K.

$$P = \langle Z_{prev}, a_\psi, Z_{cur} \rightarrow K, K \rangle, \quad (6)$$

$$Z_{prev} = Z_{prev} \cup K. \quad (7)$$

where: a_ψ - real time-independent parameters of the patient considering its time-dependent parameter s, which is a condition for revitalization of production, Z - set of drugs that describes a class situation, $Z_{cur} \rightarrow K$ - core products, K - personalized scheme of treatment, which is the result of the production rules.

So, the each internal top is the root of subtree, the examples of the same value of one of the attributes and the attribute various values of decision-making to him correspond. The examples that have the same value of one of the attributes of decision-making correspond to the each sheet of the tree.

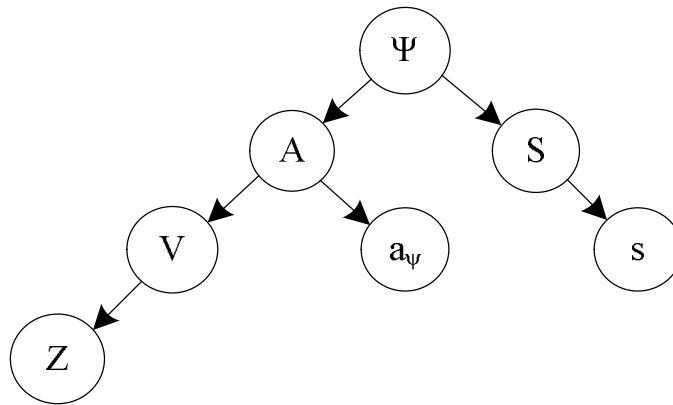


Fig. 6. The tree of output personalized treatment schemes

One question that arises in the decision tree algorithm - is the optimal size of the final tree. So, a small tree cannot cover this or that important information about the selective space. However, it is difficult to say when the algorithm has to stop, because it is impossible to predict whether the addition of a node will reduce the error. This problem is known as the "horizon effect". However, the general strategy of limiting tree remains, ie removal of nodes implemented in case they do not give more information.

It should be noted that the adjusting the depth of tree

has to reduce the size of the training models of tree without reducing the accuracy of prediction or through the cross-checking.

Reducing of tree can be top-down and bottom-up. When using a top-down approach - trimming starts from the root, from the bottom-up - reducing the sheets number of tree. One of the simplest methods of management - reducing the error of tree limits. Starting from leaves, each node is replaced by the popular class and if the accuracy of the prediction don't affect then the change is stored.

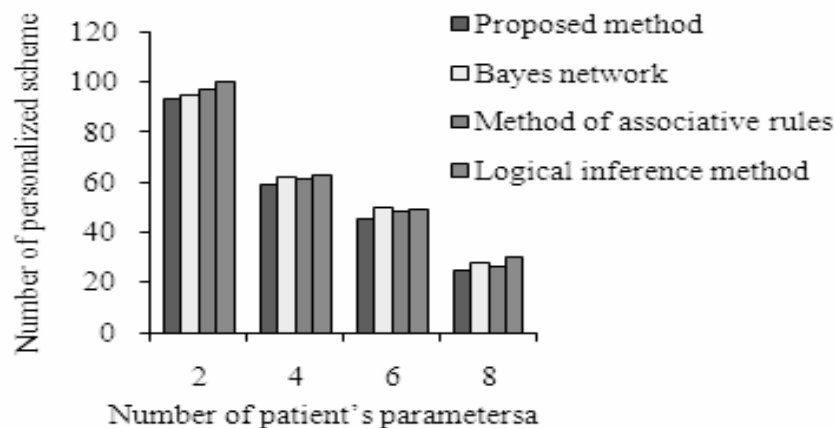


Fig. 7. The comparative diagram of the number results.

To compare the efficacy of the proposed method of data analysis was chosen three methods are similar in purpose: Bayes network, the method of associative rules and logical inference method.

High-performance unified selection method because the method by balancing search of tree treatment schemes processes only the personalized data specified in the input data set. As a result, increase in the selection criteria (the patient's parameters) inverse proportion to affect the list of proposed therapeutic schemes.

According to the analysis derived therapeutic schemes by using the submitted methods occurs the proportional reducing of the results number obtained by increasing the input patient's parameters. A result of analysis a unified selection method gives the most personalized list of therapeutic schemes.

CONCLUSIONS

The main approaches to personalized medical data envisage:

1. Design and development of decision medical support system;
2. Application of the theory of decision-making tree;
3. Developing a finding method of target value to exit decision medical support system;
4. Developing a method for decisions-making about personalization standard therapeutic schemes.

To resolve these tasks offered:

- main stages of development and design of a decision support system that enable the decomposition control process and describe the relationship between input and output control flows,

– the development DMSS need to reconcile the results from previous states and evaluate decisions making there,

– outlined the main stages of the process of making medical decisions,

– improved the method of decision-making for personalized standard schemes by introducing evaluation function the emergence of the next state in the presence time-dependent and dependent between them data a system that unlike the ordered search algorithm provides orientation when making decisions on the choice of targeted treatment regimens.

According to the results analyzed the quantitative results of applying the proposed method and existing for determining personalized schemes. It is established that the amount of received personalized schemes by increasing the input patient's parameters decreased by 7% compared to other methods.

REFERENCES

1. **Nord E.** QALYs: some challenges. *Value Health/ E.Nord, N.Daniels, M.Kamlet.* - 2009. - 12(Suppl. 1). - Pp. 5-10.
2. **Christopher-Paul Milne.** Lack of Clinically Useful Diagnostics Hinder Growth in Personalized Medicines [Электронний ресурс] / Christopher-Paul Milne // Tufts Center for the Study of Drug Development (CSDD). —July 19, 2011.
3. **Lytvyn V.** Designing Intelligent agents based on adaptive ontologies / V. Lytvyn, N. Shakhovska, A. Melnyk, Yu. Ryshkovets // Materials of VI International Scientific conference "Intelligent decision Support systems and computational intelligence problems», Yevpatoria, May 17-21, 2010. - Vol 2 - Kherson: KNTU, 2010. - Pp. 401-404.
4. **Moshkov M.** Decision trees. Theory and Applications / M. Moshkov - Nizhny Novgorod. Publishing house Nizhegorod. University Press, 1994.
5. **Shakhovska N.** Analysis methods of decision support systems for medical / N. Shakhovska, N. Melnikova // Mathematical Machines and Systems / IPMMiS NAS of Ukraine. - Kyiv, 2011. - № 2. - Pp. 62-72.
6. **Dunham M.** Data Mining Introductory and Advanced Topics / M. Dunham, Pearson Education. — Inc., 2003.
7. **Shakhovska, N., Bolubash, Y. 2015.** "Dataspaces architecture and manage its components class projection". *Econtechmod. An international quarterly journal.* Poland, Lublin – Rzeszow. Vol. 4, No 1. Pp.87-95.
8. **Lytvyn, V., Semotyuk, O. Moroz, O. 2013.** "Definition of the semantic metrics on the basis of thesaurus of subject area". *Econtechmod. An international quarterly journal.* Poland, Lublin – Rzeszow. Vol. 2, No 4. Pp.47-51.
9. **Berko A., Vysotska V., Pasichnyk V. 2009.** Electronic content commerce systems. Lviv: NULP, p.612. (in Ukrainian).
10. **Vysotska V., Sherbyna Y., Pasichnyk B., Shestakevich T. 2012.** *Mathematical linguistics.* Lviv: "Novy Svit - 2000".
11. **Lytvyn V. 2013.** Design of intelligent decision support systems using ontological approach. *Econtechmod: Lublin, Rzeszow, Vol. II, No 1, Pp.31-38.* (in Poland).
12. **Vysochina M. 2014.** The innovative approach to the study of decision-making in the context of the specific character of a product of managerial work. *Econtechmod: Lublin, Rzeszow, Vol. III, No 2, Pp.87-92.* (in Poland).
13. **Fedasyuk D., Yakovyna V., Serdyuk P., Nytrebych O. 2014.** Variables state-based software usage model. *Econtechmod: Lublin, Rzeszow, Vol. III, No 2, Pp.15-20.* (in Poland).
14. **Franklin, Hrausser. 1996.** Ahent abo prohrama? Taksonomiya avtonomnykh ahentiv. Springer-Verlag.
15. **Harold E. 2014.** Popplewell Agents & Applicability
16. **Hlybovets N.N. 2002.** Yspol'zovanye ahentnykh tekhnolohyy v systemakh dystantsyonnoho obrazovanyya.// *Upravlyayushchye systemy y mashyny, №6, Pp.69-76.*
17. **Rashkevych Y., Tkachenko R., Tsmots I., Peleshko D. 2014.** Neural-like methods, algorithms and structures of signal and image processing in real time: Monograph /. – Lviv: Lviv Polytechnic National University Publishing House – p.256. (In Ukrainian).
18. **Medykovskyy M., Tkachenko R., Tsmots I., Tsymbal Y., Doroshenko A., Skorokhoda A. 2015.** Intelligent components of integrated automated control systems: Monograph / – Lviv: Lviv Polytechnic National University Publishing House. –p.280. (In Ukrainian).
19. **Tsmots I. 2005.** Information technologies and specialized tools for signal and image processing in real time. - Lviv: UAD. – p.227. (In Ukrainian).
20. **Lytvyn V., Oborska O., Vovnjanka R. 2015.** Approach to decision support Intelligent Systems development based on Ontologies. - *ECONTECHMOD. An international quarterly journal.* Vol. 4. No. 4, Pp.29 – 35.