The Personalized Approach to the Processing and Analysis of Patients' Medical Data

Nataliia Melnykova [1] [0000-0002-2114-3436], Volodymyr Melnykov [2] [0000-0002-7008-4014] and

Edgars Vasilevskis[3]

[1] Lviv Polytechnic National University, S. Bandery Str., 12, Lviv, 79013, UKRAINE
[1] melnykovanatalia@gmail.com
[2] Danylo Halytsky Lviv National Medical University, 69 Pekarska str., Lviv, 79010, Ukraine
[2] v.melnikov2013@gmail.com
[3]Riga Stradins University, Institute of Anatomy and Antropology, Latvia
[3] crasen111@gmail.com

Abstract. The application of machine learning technology and Big data to solve the problem of personalized approach in the tasks of making medical decisions and predicting states will allow to study random mechanisms of modeling and forecasting of treatment stages taking into account individual patient's characteristics, analysis of the medicaments and their key characteristics. We use this information to develop innovative approaches to risk forecasting, modeling therapies, and improving the quality of medical care by personalizing treatment schemes of patients. And it will allow you to effectively optimize data processing even when new information revenues come from different sources.

The authors proposed the development of a System for the medical decisions support for the personalized data consolidation of the patient, which were receiving from the heterogeneous sources that are related healthcare. The conceptual scheme of the system was proposed and new approaches to consolidation and analysis of patient's data and forecasting of its states are offered. The use of various processing technologies for the Big data obtained will allow the study of random mechanisms for modeling and predicting treatment stages, taking into account individual patient's characteristics, analysis of the medicaments and their key characteristics. These will help develop innovative approaches to improve the risk stratification methodology, improve the quality of medical care by personalizing treatment schemes for patients.

Keywords: medical decisions, prediction of states, consolidation of personalized data, analysis of medical data, personalization of treatment.

1 The problems personalized data processing

The problem of obtaining unified, valid and qualitative information about the information object of the studied subject area is relevant in the modern world. This is due to the rapid growth of information flows, which are characterized by different types of data that are coming from different sources. The urgency of the distinct consolidation data in medicine is growing because of the need for the rapid processing of a large patient's information amount that is characterized by its heterogeneity due to the emergence of new features of the history of the disease for each patient, namely: it is the individual features of the patient, pre-treatment, biochemical indicators, presence of complications, previous medication therapy, etc.

The rapid growth in the volume of data collected the lack of alternative methods for their effective analysis, the need for significant human resources to support the data analysis process, and the high computational complexity of existing analysis algorithms lead to a steady increase in the time even with timely updating of hardware. This necessitates the emergence of new methods and tools for processing, the consolidation of personalized data for the process of collecting heterogeneous data of large volumes, as well as supporting of the doctor's decisions [1,6]. To solve this problem it is expedient to use approaches of technology of machine learning and Big Data.

For effective work, comprehensive solutions are needed for monitoring, filtering, structuring and searching for semantic relationships between the concepts of the studied area. However, you can observe a huge variety of variables, identify global trends and conclusions about assessing the status of the object under study, and predict the transitions between them, based on the information provided and using Big Data technology. The development and analysis of such different type data is used to stimulate the development of events and situations in decision support systems.

The researchers that had started the study of this problem were von Neumann, developers of IBM, academics of the school Lebedev SO, 17 Glushkov V.M. (system analysis, the theory of conflict games, problem-oriented systems of modeling and data processing) [1, 5], which led to the development of languages of block programming, decision support systems. So, changing the class of research - from operational to analytical, the emergence of new types of data, the need for quick access to them, led to an increased interest in the problem of integration and processing of data in order to improve the quality of solutions. The greatest activity in the field of data integration research was in the 90's of the XX century. and in our time [4,10] due to the rapid development of Business Intelligence, Machine Learning and increased data warehousing capabilities (increasing the amount of stored data, analytical processing of data).

The peculiarity of the current researching is to analyze not only the types of data (descriptions), but semantics. Especially this is active development of means for operative collection of various data, loading them into a data warehouse, analysis and forecasting, which is observed in the spheres of energy and administrative management [5,11]. But at the present time, this problem is relevant in medicine. Specifics of processing large volumes of medical data sets need to develop new methods of analysis, consolidation and forecasting to support medical decisions during diagnosis, treatment and rehabilitation.

The process of analyzing medical data is characterized by a number of definite problems that arise in solving a class of problems, namely [3,8,13]:

- Fuzzy data presented;
- classification of data;
- consolidation of data;

— determination of the general patient's condition;
— identification of personalized treatment decisions;
— assessment of the reliability of the resulting conclusions;
— assessment of risks;
— prediction of the patient's conditions under the influence of the applied therapy.

As a consequence, there are problems in data processing, namely: the absence of methods of analysis suitable for use due to their variety (for medicine - time-dependent data of the general condition of the patient, poorly structured data of laboratory studies, etc.), the need for significant human resources to support the analysis process data, high computational complexity of existing analysis algorithms and rapid growth of data collection. This in turn leads to a steady increase in the time spent on data analysis, even with the regular updating of computer tools..

Thus, there is the task of developing an effective unified method for analyzing and consolidating personalized data, which will allow it to be used not only for medicine, but also for other subject areas.

The main tasks of the analysis of medical data, which are relevant both to the diagnosis and treatment, Fig. 1

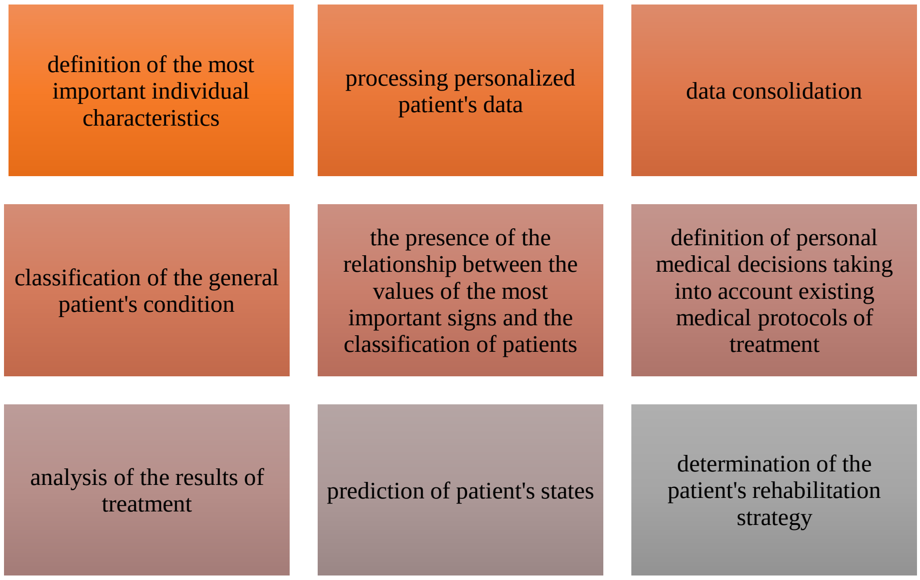

Fig. 1. The main tasks of the analysis of medical data

2 Formalized representation of medical data

The availability of effective storage, access and modification of information about the state of the object (patient), as well as combining with actual flow data about the investigated situational problem will allow to work out the structure of medical data. To do this, you need to define the structure of personalized data.

The personalized data PD is a set of data and its elements are subsets of time-independent data (A) and time-dependent data (S) of the object under study and which characterize its general state.

$$PD = \{A, S\}. \tag{1}$$

The elements of a subset of time-dependent data for example in the treatment of purulent surgical infection are: anthropometric data, functional parameters, laboratory data, diagnosis, tumors, anatomical localization, localization of the inflammatory process, patient's weight, etc.

$$A = \{a_1, a_2, a_3, \dots, a_n\}. \tag{2}$$

The elements of the set of time-independent data are: allergy, concomitant pathology, competing drugs, related drugs, previous drug therapy, patient age, cost-effectiveness factor, etc.

$$S = \{s_1, s_2, s_3, \dots, s_m\}, \tag{3}$$

where

$$A \cup S \rightarrow PD. \tag{4}$$

By analyzing data sets that are characterized by diverse information and data representation in different models, one can use the data consolidation approach by creating associations between data objects from different participants; improving access to sources with limited own means of access; ensuring the ability to execute queries without access to a real data source; data consolidation as a result of a user's request; maintaining a high level of accessibility and recovery.

Due to the increase of patient information during diagnostic and treatment processes, and depending on the implementation of the data directory (Dd), which contains a model management environment (Mm), which allows you to create new connections and manage the links between them.

The relationship between a directory, a model management environment and a consolidated data repository (CDR) can be represented as a mapping [3, 19]:

$$Mm(\,Dd\,) \rightarrow CDR \tag{5}$$

The more models can be identified, the more accurate the information will be in the CDR, and will allow the integration, search and processing of data in the data space.

3 Formalization of the process of analysis of personalized medical data

In accordance with defined data sets to ensure the personalization of solutions, a conceptual model for personalizing decisions on the definition of treatment is proposed, Fig. 2.

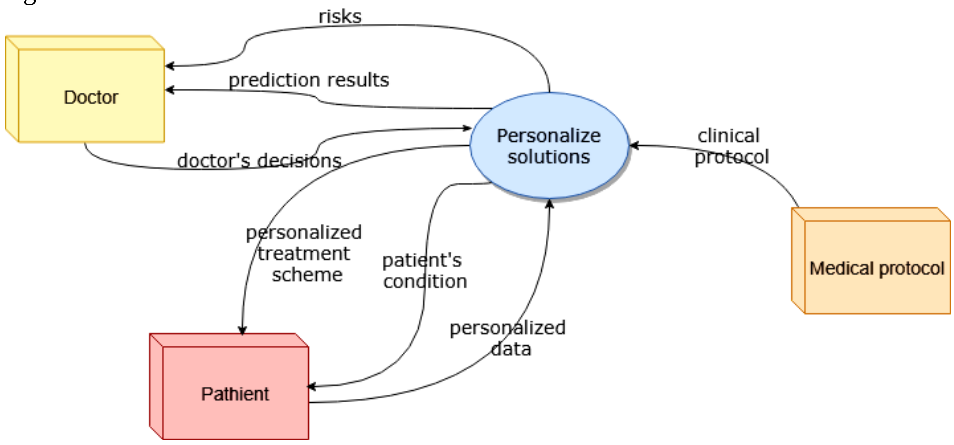

Fig. 2. The conceptual model of personalization solutions for determining treatment

The process of processing and analyzing personalized data for the search for medical solutions is presented in pairs:

$$APD = \{GS, PD \cup D\}. \tag{6}$$

GS is a set of patient states; PD is a collection of personalized data, where rank (p) is the total number of personalized data; V_p is a set of values of the attribute p_i.

$$PD = \{p_1, p_2, \ldots, p_{rank(p)}\} \tag{7}$$

and $p_i: GS \to V_p$;.

It is assumed that the set *D* - the set of personalized solutions has the finite size of *rank (d)*, V_d - the set of attribute values according to the protocol of treatment.

$$D = \{d_1, d_2, \ldots, d_{rank(d)}\} \tag{8}$$

and $d_i: PD \to V_d$.

PD called attributes set conditions, and D - solutions. O_i (the set of objects) is i-th class of solutions, the d_i-value and its solution obtained from the set of solutions D.

$$O_i = \{g_{i_k} \in GS, : D\left(g_{i_k}\right) = d_{i_k}\}; \tag{9}$$

Thus, the rule of decision is a formula of the form

$$d_k = \left(p_{i_1} \to v_{j_1}\right) \wedge \left(p_{i_2} \to v_{j_2}\right) \wedge \ldots \wedge \left(p_{i_{rank(p)}} \to v_{j_{rank(d)}}\right). \tag{10}$$

4 The application of personalized data analysis for decision-making tasks

There is a large number of special algorithms for constructing classification rules [6]. These algorithms are NP-computable from a computational point of view, and their application requires the use of special heuristics to reduce the total amount of computations. In this case, the total number of rules received can be significant, which will require additional efforts to select the best ones. Among the well-known approaches to the formation of such heuristics is the method of associative rules, reverse output, Bays, Boolean considerations [2, 10, 11]. Solving the decision making problem in determining the treatment in terms of applying the rules received should be based on the experience of previous experience.

In the last years the algorithm ID3 and its modifications C4.5, See5 are actively used to construct decision trees [14, 17]. All these algorithms build trees and generate rules based on examples.

The practical application of the classic ID3 algorithm is due to a number of problems that are typical of learning-based models and decision trees, in particular. One of the drawbacks of the ID3 algorithm is that it works incorrectly with attributes that have unique values for all objects in the training sample. The information entropy is zero for such objects and no new data can be obtained from a constructed tree and with a given dependent variable, so the subsets obtained after the partition will contain one object. To effectively overcome the shortcomings of the ID3 has been refined, resulting in its expansion, called C4.5.

C4.5 algorithm solves this problem by introducing normalization. There is evaluated not the number of objects of a class after the partition, but the number of subsets and their power (number of elements) is evaluated. However, the problem of processing exclusively independent parameters remains.

The evaluation function V (PD) is crucial for the process of personalizing the treatment scheme. This function is obtained as a result of application of the method of unification of personalized schemes and is formed on the basis of the Bayesian theorem. The weight of the appearance of the next event corresponds to the largest value of the a posteriori probability of the appearance of the next state, taking into account the time-dependent input parameters.

$$V(S) = max\left(p\left(G \mid S\right)\right). \tag{11}$$

There is a prototype of the decision support system for the analysis of personalized patient data, Fig. 4, as a result of application of a search tree using the method of unification of personal treatment schemes for targeted solutions.

The work of many methods was analyzed for processing personalized data and the processing speed of the query when looking for an individual treatment scheme. Comparison of the time complexity of making medical decisions of the methods used is represented in Table.1:

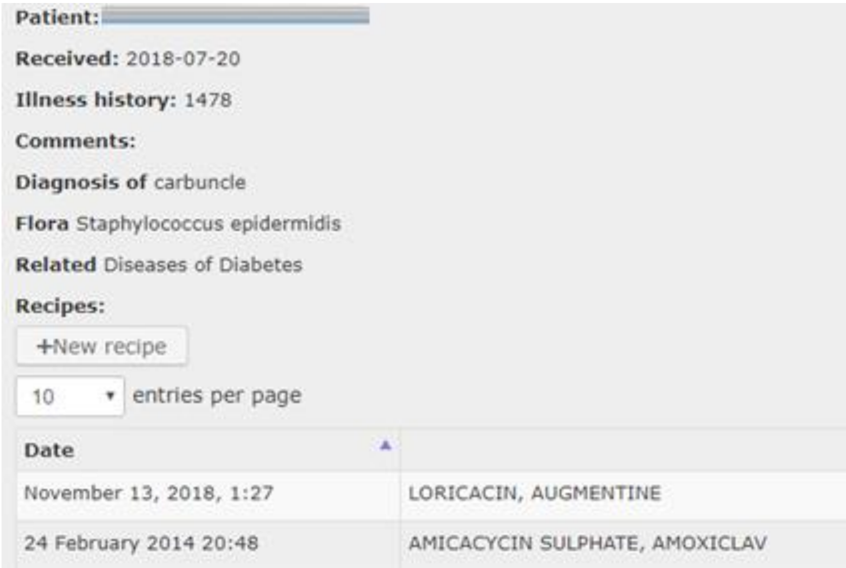

Fig. 3. The result of the processing of personalized data in the system "Antibioticus".

Table 1. The table of definition of the query processing time indicator

Number of queries	**Method of unification of personal treatment schemes**	**Bays Network**	**Method of associative rules**	**Method of logical output**
20	0,056	0,059	0,057	0,06
40	0,059	0,06	0,059	0,06
60	0,06	0,0616	0,061	0,062
80	0,061	0,0625	0,062	0,063

The high speed of the unified selection method is explained by the fact that the method processes only the personalized data given in the input data set due to the balance of the search tree of the treatment scheme. As a result, the increase in selection criteria (patient's parameters) affects is inversely proportional to the list of proposed therapeutic schemes.

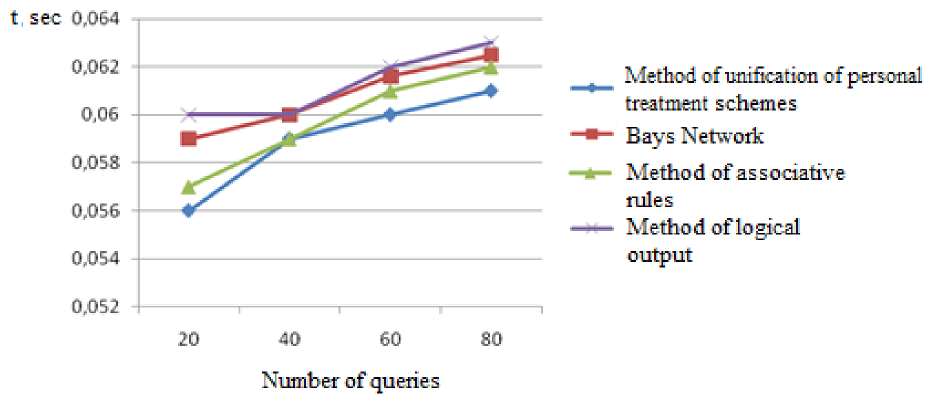

Fig. 4. Diagram of the estimation of the processing time of the request

Conclusions

So, the personalized approach to the processing of medical information is characterized by a number of problems, namely: the uncertainty of the data presented, the classification of data, data consolidation, the definition of the general patient's condition, the definition of personalized treatment decisions, the assessment of the reliability of the resulting conclusions, assessment of the emergence of risks, prediction of the conditions of the patient under the effect of the applied therapy. As a result, there are problems in data processing, the lack of methods of analysis, suitable for use due to their variety.

The approach is proposed for the unification and formalization of personalized data to simplify the process of processing and solving a number of identified problems, which will allow optimize the process of analyzing medical data and making medical decisions.

The methods of personalized processing of medical data are analyzed, namely: the method of unification of personalized treatment schemes, the Bayesian network, the method of associative rules and the method of logical deduction. This formed the vision of the effectiveness of their application for this kind of task. As a result of an increase in selection criteria (patient's parameters), the inverse proportional effect on the list of proposed therapies in the method of unification of personalized treatment schemes. This allows to increase the search speed by balancing the tree search and processing only the personalized data that arrives in the input data set.

References

1. Mulesa, P., Perova, I.: Fuzzy Spacial Extrapolation Method Using Manhattan Metrics for Tasks of Medical Data Mining. Computer Science and Information Technologies CSIT 2015, Lviv, Ukraine, pp. 104-106, (2015)

2. Melnykova, N., Markiv, O.: Semantic approach to personalization of medical data. Computer Sciences and Information Technologies. Proceedings of the 11th International Scientific and Technical Conference, CSIT 2016, pp. 59-61, (2016)
3. Melnykova, N., Marikutsa, U.: Specifics personalized approach in the analysis of medical information. ECONTECHMOD. An International Quarterly Journal on Economics of Technology and Modelling Processes, 2016, Vol. 5, No 2, pp. 113-120, (2016)
4. Perova, I., Pliss I., Churyumov, G., Franklin, M., Eze, Samer, Mohamed, Kanaan, Mahmoud: Neo-Fuzzy Approach for Medical Diagnostics Tasks in Online-Mode. 2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP), Lviv, Ukraine, pp. 34-38, (2016)
5. Melnykova, N., Shakhovska, N., Sviridova, T.: The personalized approach in a medical decentralized diagnostic and treatment. 14th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics, CADSM 2017, pp. 295-297, (2017)
6. Melnykova, N.: Semantic search personalized data as special method of processing medical information. Intelligent Systems and Computing, pp. 315-325, (2017)
7. Bodyanskiy, Ye., Perova, I., Vynokurova, O., Izonin, I.: Adaptive Wavelet Diagnostic Neuro-Fuzzy System for Biomedical Tasks. Proc. of 14th International Conference on Advanced Trends in Radioelectronics, Telecommunications and Computer Engineering (TCSET), Lviv-Slavske, Ukraine, February 20 – 24, pp. 711- 715, (2018)
8. Izonin, I., Trostianchyn, A., Duriagina, Z., Tepla, T., Lotoshynska, N.: The combined use of the wiener polynomial and SVM for material classification task in medical implants production. International Journal of Intelligent Systems and Applications. 9, pp. 40-47, (2018)
9. Boyko, N., Sviridova, T., Shakhovska, N.: Use of machine learning in the forecast of clinical consequences of cancer diseases. 7th Mediterranean Conference on Embedded Computing. MECO 2018 - Including ECYPS 2018, Budva, pp. 1-6, 2018,
10. Calvaneseand, D., Calvanese, D., De Giacomo, G., Lembo, D., Lenzerini, M., Poggi, A., Rosati, R.: Ontology-based database access. Proc. of SEBD – 2007, pp. 324–331, (2007)
11. Mintser, O., Stryzhak1, O., Denysenko, S.: Use of principles of medical ontology for construction of scenario models of post-graduate education of doctors and pharmacists. National Medical Academy of Postgraduate Education Shupyk, Medical Informatics and Engineering, Kyiv, No 2, pp. 18-23, (2013)
12. Lytvyn, V.: Design of intelligent decision support systems using ontological approach. The international quarterly journal on economics in technology, new technologies and modelling processes, Vol. II, No 1, pp. 31-38, (2013)
13. Lytvyn, V., Dosyn, D., Shkutiak, N.: Mathematical software development base domain ontology. Proceedings of II international scientific-practical conference. Modern information technology and other innovation transport, Kherson, 25-27 May 2010, HDMI, pp. 345-348, (2010)
14. Slyunyayev, A.: Methods of finding solutions in intelligent multi-agent information management system airport. Proceedings of the Military Institute of Kyiv National University Taras Shevchenko, №28, Kyiv-2010, pp. 110-114, (2010)
15. Dunham, M.: Data Mining Introductory and Advanced Topics. Pearson Education, Inc., (2003)
16. Donald, K.: Generating all trees. The history of combinatorial generation. The Art of Computer Programming, Williams, T. (4), pp. 160, (2007)
17. Perova, I., Litovchenko, O., Bodyanskiy, Ye., Brazhnykova, Ye., Zavgorodnii, I., Mulesa. P.: Medical Data-Stream Mining in the Area of Electromagnetic Radiation and Low Temperature Influence on Biological Objects. Proc. 2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP), August 21-25, (2018), Lviv, Ukraine, pp. 3-6, (2018)

18. Perova, I., Bodyanskiy, Ye., Brazhnykova, Ye., Mulesa P.: Neural Network for Online Principal Component Analysis in Medical Data Mining Tasks. IEEE First International Conference on System Analysis & Intelligent Computing (SAIC) 8-12 October 2018, Kyiv, Ukraine, pp.150-154, (2018)
19. Perova, I., Bodyanskiy, Ye.: Fast medical diagnostics using autoassociative neuro-fuzzy memory. International Journal of Computing, 16 (1), pp. 34-40, (2017)
20. Melnykova, N., Marikutsa, U., Slych, A.: The Intelligent SystemArchitecture of Personalized Management. XIV Ukrainian-Polish Conference on CAD in Machinery Design. Implementation and Educational Issues, CADMD-2016, pp. 21-22, (2016)