Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 145 (2018) 238–241

Procedia Computer Science

www.elsevier.com/locate/procedia

Postproceedings of the 9th Annual International Conference on Biologically Inspired Cognitive Architectures, BICA 2018 (Ninth Annual Meeting of the BICA Society)

# Health Management System Knowledge Base for Formation and Support of a Preventive Measures Plan

Oleg G. Grigoriev[a], Boris A. Kobrinskii[a], Gennadiy S. Osipov[a], Alexey I. Molodchenkov[a], Ivan V. Smirnov[a] *

[a]*Federal Research Center "Computer Science and Control" of RAS, Vavilova str., dom 44, kor.2, Moscow, 119333, Russian Federation*

**Abstract**

This paper describes the creation of a knowledge base for an intelligent healthcare management system dealing with cases of stroke, myocardial infarction and depression. The main purpose of the knowledge base is to discover and evaluate risk factors and situations which can lead to such diseases, and to enable the formation and support of a preventative measures plan. The present version of the knowledge base is implemented using a heterogeneous semantic network approach and utilizes expert opinions about risk factors and events influencing an individual's health. Data includes genetic predisposition, lifestyle, and external environment. Data is compiled with the aid of questionnaires, mobile devices, case histories and information from social media. Information from social media is analyzed using data and text mining methods with the goal of evaluating the user's condition. All of the data obtained is accumulated in a single database. The knowledge base establishes risk factors, including changes in those factors over the course of time, and circumstances or events which might precipitate the emergence of pathology. Hypotheses are generated about the current state of the user's health, the active risk factors which created conditions for the onset of disease, and circumstances which might produce an increase or decrease in risk factors. Prophylactic measures to reduce those risks are suggested through analysis of the hypotheses generated. Recommendations regarding prophylactic measures are formed with the aid of the knowledge base, the user case-library, and collaborative filtering methods. Recommendations are based on 4P medicine, which re-quires mandatory participation of system users in maintaining their health.



* Alexey I. Molodchenkov. Tel.: +7-499-135-7310;.
*E-mail address:* aim@tesyan.ru

1877-0509 
10.1016/j.procs.2018.11.050

*Keywords:* Heterogeneous semantic network, Knowledge base of predictive medicine, Risk factors, P4 medicine.

## 1. Introduction

P4 medicine, which is currently undergoing active development around the world, considers prevention and a personalized approach to be the two pillars of healthcare [1]. Therefore, a personalized prevention plan built on P4 principles will aim to reduce the impact of individual risk factors and situations that could contribute to the onset of disease. The high number and multidimensionality of risk factors and adverse situations make it impossible for a doctor to take more than a fraction of them into account. Therefore, to address problems of personalized health maintenance it is necessary to use methods of artificial intelligence and big data analysis.

## 2. Formulating the Problem

This study investigates methods that enable detection of prenosological conditions. Combinations of dynamically changing risk factors and possible adverse situations are analyzed toward this end. These combinations determine a personalized plan of preventive measures.

## 3. System knowledge base

The knowledge base is a computer model of a heterogeneous semantic network [2]. It consists of classes and objects, their properties, relationships between objects, and properties of the objects. The knowledge base model looks like:

$$KB = <CL, O, C, CT, P>,$$

where CL is the set of classes; O is the set of objects; C is the set of relationships; CT - the set of types of relationships; P - set of procedures.

The class cl is an arbitrary pair <N, Ch>, where N is the name of the class, and Ch is its characteristic (the set of properties of the class). Objects are the main components of the knowledge base. They can be various entities of the subject area: statements, facts, results of observations, analysis, recommendations. The object

o = <oN, oCh>, where oN is the name of the object, oCh is its characteristic, which is specified by a non-empty set of attributes and their values. Each attribute Ai, regard-less of the object under consideration, is associated with the domain Di = D (Ai). Each object belongs to a certain class, the properties of the class are inherited by the objects belonging to it. The class name, as a rule, plays a large informative role. Objects that form the knowledge domain are divided into two groups. The first group includes the original objects that determine the inference in the system (they are called risk factors), and the second one - the target objects, in order to achieve certain characteristics of which hypotheses are developed. The knowledge base generates hypotheses for a given user on the basis of risk factors and various potentially adverse situations. Risks of diseases, preventive measures aimed at reducing the risks of diseases, can serve as hypotheses at different stages. Various relationships between objects are used for generating hypotheses and inferences in the system. Relationships are executed by calling corresponding procedures. The system solver is responsible for generation of hypotheses, their selection, and issuance of recommendations.

## 4. Pathways to solving the problem of health maintenance

Automated health assessment involves regular monitoring of the character of changes in the health status of a user based on the dynamics of individual risk factors and possible adverse situations. Data for these are received from: 1) monitored parameters (mobile devices), 2) questionnaires filled by users, 3) information from social networks, 4) medical records. Intelligent analysis of data and texts is used for data mining. In the future, it is anticipated that intensive processing of large data will be performed as information is accumulated.

Health amelioration assumes the ability to influence manageable personal risk factors and possible situations. To accomplish this, algorithms are used which calculate the risk of diseases and the functional state of specific individuals, along with their temporal dynamics (determined with the use of mobile devices, analysis of texts in social networks, and changes in indicators in electronic medical records). Preventive measures plan recommendations are formulated in accordance with similar cases in the closest user groups.

In the first stage of the system, the risks of such diseases as myocardial infarction, stroke and depression are investigated. Formation and support of a preventive measures plan requires analysis of the dynamics of risk factors and pre-pathological conditions.

## 5. Intelligent analysis of data and texts from social media for assessing health status

Social media - blogs, forums, social networks, are an integral part of the way of life of modern man. By analyzing the behavior of a user of social media, his activity and text messages, one can obtain information about the lifestyle of a person and the condition of his mental and physical health. Users' messages in social networks can help to determine the level of their physical activity, food preferences, bad habits, the state of obesity, etc [3, 4].

Prevention and early detection of mental disorders, such as depression, among users of social networks is a problem of special interest and importance. It can be solved with the help of analysis of a user's activity and text messages. In our system, to determine the existence of depression, a combination of Text Mining for analyzing text messages and machine learning for analyzing user activity in social media is used. We use original psycholinguistic markers [5], morphological features and word embedding for advanced analysis of texts in Russian and English. We explored different sets of features for the task of depression detection in social media on CLEF/eRisk 2017 data [6] with SVM and Random Forest algorithms and achieved best results on the test data with 63% F1-score [7].

The derived estimates of the health status of a social media user are stored in the database and are used to determine the risks of disease with the help of knowledge base inferencing.

## 6. System inference principles

Differentiating between various conditions in a user's health depends on both direct and indirect relationships between risk factors and diseases [8]. The argument "for" ensures generation of hypotheses that indicate the risks of diseases. This makes it possible to take changes in multiple risk factors and various situation-based characteristics into account. At the same time, protective factors, both biological and external, are counter-arguments that reduce the risk of diseases.

## 7. System functioning

Users of the system on their first visit to the site fill out a preliminary questionnaire (in the future, it is proposed to manually fill in only those fields that show changes in dynamics and that cannot be filled out automatically). Then the system switches to other information sources (electronic medical cards, social networks, etc.), which may contain information about a particular individual. In addition, mobile devices collect information about the user's physical condition (pulse rate, blood pressure, etc.). All the information received is fed into the knowledge base and analyzed for identification and evaluation of risk factors and health abnormalities. The presence of counterarguments, characterized by negative links to the risk of pathology, lowers the risk level. A plan of preventive measures is formed in accordance with the hypothetical disease risk, taking into account the user's individual characteristics. The plan also includes lifestyle changes. The described procedure is repeated when necessary.

## 8. Conclusion

The paper describes the knowledge base for the formation, management and continuous support of a preventive measures plan for a user's health maintenance. In order to accomplish this, the knowledge base system provides identification of those combinations of risk factors for diseases and situations that are unfavorable with regard to the

development of pathological processes. At the present stage, the knowledge base focuses on detecting and assessing those personal risks of diseases such as stroke, myocardial infarction and depression, which characterize pathological conditions of various body systems. The knowledge base is a heterogeneous semantic net-work. This makes it possible not only to describe weakly structured subject areas, but also to explain the formulated solutions.

The application of methods of artificial intelligence makes it possible to solve the problem of health maintenance using the personalized approach of P4 medicine, thus ensuring the user's involvement in the care of their own health.

## Acknowledgements

This work was financially supported by the Ministry of Education and Science of the Russian Federation. Grant No. 14.604.21.0194 (Unique Project Identifier RFMEFI60417X0194).

## References

[1] Flores, M., Glusman, G., Brogaard, K., Price, N.D., Hood, L.: P4 medicine: how systems medicine will transform the healthcare sector and society. Per Med 10(6) 565-576, (2013).

[2] Osipov, G.S.: Formulation of subject domain models. 1. Heterogeneous semantic nets. So-viet J Comput and Systems Sci 30(2) 1-12 (1990).

[3] Nyoon, S., Elhadad, N., Bakken, S.: A practical approach for content mining of Tweets. Amer J Preventive Med 45(1) 122-129 (2013).

[4] Hingle, M., Yoon, D., Fowler, J., Kobourov, S., Schneider, M.L., Falk, D., Burd, R.: Col-lection and visualization of dietary behavior and reasons for eating using Twitter. J Med In-ternet Res 15(6) e125 (2013).

[5] Vybornova, O., Smirnov, I., Sochenkov, I., Kiselyov, A., Tikhomirov, I., Chudova, N. et al. Social Tension Detection and Intention Recognition Using Natural Language Semantic Analysis (on the material of Russian-speaking social networks and web forums). In: Proceedings of the European Intelligence and Security Informatics Conference (EISIC), Sep-tember 12-14 2011, pp. 277-281. Athens, Greece (2011)

[6] Losada, D.E., Crestani, F., Parapar, J.: eRISK 2017: CLEF Lab on Early Risk Prediction on the Internet: Experimental foundations. In: International Conference of the Cross-Language Evaluation Forum for European Languages, pp.346-360. Springer, Cham (2017).

[7] Stankevich, M., Isakov, V., Devyatkin, D., Smirnov I.: Feature Engineering for Depression Detection in Social Media. In: Proceedings of the 7th International Conference on Pattern Recognition Applications and Methods, pp.426-431 (2018).

[8] Kobrinskii, B.A.: Logic of the argument in decision-making in medicine. Automatic Docu-ment and Mathematic Linguistic 35(5) 1-8 (2001).