Available online at www.sciencedirect.com

ScienceDirect

IFAC PapersOnLine 52-25 (2019) 70–74

Artificial Intelligence Technologies Application for Personal Health Management

B.A. Kobrinskii*. O.G. Grigoriev**
A.I. Molodchenkov*. I.V. Smirnov******
N.A. Blagosklonov*****

**Federal Research Center "Computer Science and Control" of Russian Academy of Science, Moscow, Russian Federation (e-mail: kba_05@mail.ru)*
***Federal Research Center "Computer Science and Control" of Russian Academy of Science, Moscow, Russian Federation (e-mail: oleggpolikvart@yandex.ru)*
****Federal Research Center "Computer Science and Control" of Russian Academy of Science, Moscow, Russian Federation (e-mail: aim@tesyan.ru)*
*****Federal Research Center "Computer Science and Control" of Russian Academy of Science, Moscow, Russian Federation (e-mail: ivs@isa.ru)*
******Federal Research Center "Computer Science and Control" of Russian Academy of Science, Moscow, Russian Federation (e-mail: nblagosklonov@gmail.com)*

Abstract: Health management, or in other words health optimization, is coming to the forefront in most of the world's countries. Almost every person has a certain predisposition to the development of chronic diseases, which is determined by corresponding risk factors. Some of these risk factors can be managed in order to minimize the risk of disease, which promotes health optimization. The intelligent system presented here performs the task of monitoring manageable risk factors, assessing their impact, and formulating recommendations to reduce this impact. The performance of the system is demonstrated on the examples of myocardial infarction, stroke and depression. The intelligent system evaluates the risks of diseases on the basis of individual risk factor values and provides recommendations to reduce these risks. A health manager who possesses additional information about specific aspects of the person's health can adjust the recommendations issued by the system.

Keywords: artificial intelligence, risk factors, health stability, intelligent health management system, automatic control, alarm health system, knowledge base, semantic networks, health optimization, myocardial infarction, stroke, depression.

1. INTRODUCTION

According to the World Health Organization, every year 41 million people around the globe die from noncommunicable diseases (NCDs) – i.e., 71% of all deaths. Cardiovascular diseases (CVD) account for the largest share, which is 17.9 million people, and the proportion of young and middle-aged people is constantly increasing. The risk factors (RF) for these diseases are largely related to life-style: a low level of physical activity, smoking, an unhealthy diet, and excessive drinking. The health optimization system presented here is aimed at eliminating such risk factors and/or at reducing their negative impact. Investigation of this problem must be systemic in terms of monitoring risk factors: for example, the development of ischemic processes in the heart and brain often correlates in patients and leads to the exacerbation of both processes. It is important at all times to control the risk factors of stress and anxiety, which adversely affect all organ systems and, in addition, contribute to the onset of depression.

Due to the large number of risk factors (predictors) and their possible combinations, doctors cannot fully take their impact into account, which would be necessary in order to develop and optimize recommendations for reducing the risk of disease. The intelligent system presented here performs risk factor dynamics monitoring, assesses RF impact on disease risks and forms personalized recommendations to reduce these risks, thus enabling a solution to the above problem. It is an alert health system for the individual and the doctor, which warns of an increase in the risk of disease and, consequently, deterioration of health.

2. HEALTH STABILITY AND DISEASE RISK FACTORS

Human health stability (optimization) depends primarily on a combination of genetic and social factors. According to S.V. Emel'yanov (Emel'yanov, 2009), in control theory it is important to identify situations in which it is possible to obtain the optimal or suboptimal stabilization solution in the presence of considerable uncertainty in the description of the control object and/or external forces. Exactly this is observed in health assessment and management.

An individual's health status is assessed here in terms of threats to his health (disease risks). His social microenvironment is determined by lifestyle, including physical activity and diet. Risk factors can be categorized as

Peer review under responsibility of International Federation of Automatic Control.
10.1016/j.ifacol.2019.12.448

manageable (for example, physical activity, body weight, drug-controlled hypertension, diabetes mellitus adjusted by a special diet) or non-manageable (for example, genetic conditions, previous myocardial infarction or stroke). Risk factors, excluding those that are non-manageable (molecular-genetic, age and some others), can be ameliorated by specific recommendations for health optimization.

Epidemiological studies in various populations have reported numerous risk factors for stroke, including age, ethnicity, blood pressure, pre-existing cardiovascular disease, and diabetes mellitus. Integrating these reported factors, several assessment models have been developed to evaluate an individual's absolute risk for stroke onset. One of the most well-regarded stroke risk appraisal tools is the Framingham Stroke Risk Score. However, despite its prominence, several widely accepted risk factors for stroke, such as body weight and shape and family history were not included in the original Framingham model. A new stroke model was later developed that combines the original Framingham algorithm with seven additional literature-derived risk factors; this new model was created using a novel modeling process, called synthesis analysis. The additional seven risk factors included African American ethnicity, physical exercise level, body mass index, waist circumference, height, HDL cholesterol and the use of hormone replacement therapy (Zhou et al., 2017).

The risk factors knowledge base of the presented system was created with expert assistance sourced from automated analysis of studies on cardiovascular pathology and depression. The total number of risk factors, including their attributes and attribute features, is more than 250 items. The attribute structure was developed on the basis of knowledge engineering, taking into account the features of the system under consideration. Attribute features (values), or their combinations, are correlated with the linguistic risk level estimated by experts (from very low to very high, 5 levels total). A quantile linear regression model with 0.05, 0.5 and 0.95 level quantiles is used to assess personalized risks of human health changes (Koenker et al., 2001). The results of the disease risk assessment for a given period of time were utilized as initial data for training the model. When training the model, the loss function was optimized:

$$L(\omega,\theta) = \sum_{t|y_t \geq f(t,\omega)} \theta|y_t - f(t,\omega)| + \sum_{t|y_t < f(t,\omega)} (1-\theta)|y_t - f(t,\omega)|,$$

where θ is the quantile level, t - time point for risk assessment modeling, $f(t, \omega)$ - result of the risk assessment modeling at time t, y_t - risk assessment of the disease at time t, ω - model parameters vector.

3. HEALTH DATA COLLECTION

To obtain data on the dynamics of indicators that are risk factors (predictors) of diseases, questionnaires, mobile devices (gadgets), social networks, and data from medical records are used. Questionnaires include information about user risk factors, potentially stressful events, and personal characteristics, that may have an impact on the development of pathological processes. The system collects information about diseases, family history, lifestyle, nutrition, social status, physiological data, laboratory and functional studies results. The system has the capability to connect to a GoogleFit personal profile and to automatically collect a variety of data: physical activity (activity duration, its intensity), sleep, pulse, weight, etc. It is also possible to establish access to a personal social network profile for automatic collection and analysis of information about the user's social network activity to assess his personality traits, predictors and risk factors of depression (Stankevich et al., 2018).

Different scales used in the system allow a comprehensive assessment of an individual's health status and a forecast of possible changes in a positive or negative direction.

A personalized approach takes into account individual characteristics, including both the risks of pathologies manifested in the course of life (for example, obesity) and potential chronic diseases.

4. KNOWLEDGE BASE FORMATION PRINCIPLES

The knowledge base is built on heterogeneous semantic network principles (Osipov, 1990). It includes objects with their properties, relationships between objects, properties of objects, and their values. Objects include both risk factors with their attributes and values (which may include, among other things, the results of observations and analyses), as well as disease risks and recommendations. The latter are divided into 3 levels: name, description and explanation. At present, the knowledge base includes three diseases - myocardial infarction, stroke, depression.

A cognitologist (knowledge engineer), in collaboration with experts, and with the help of a risk factor disease table builds a semantic network. Figure 1 – shows a network fragment illustrating the connection between risk factors and the level of disease risks and recommendations.

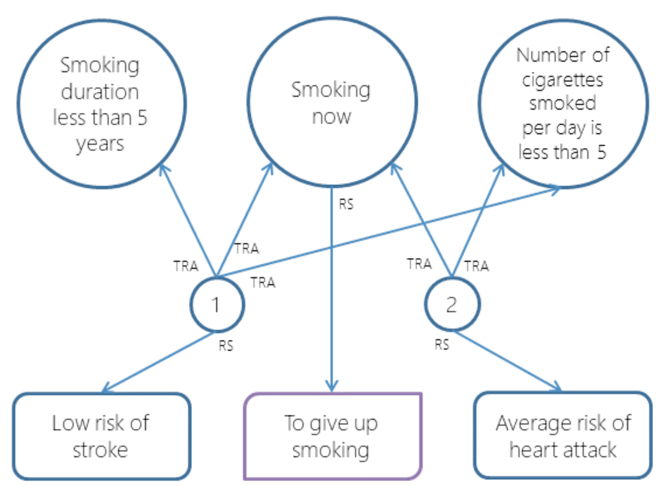

Fig. 1. Knowledge base fragment (on the example of RF "smoking")

1 and 2 - Hidden nodes "AND"

TRA is a link indicating that a node with an inbound link is always observed if a corresponding outgoing link node exists.

RS is a link indicating that a node with an inbound link can be observed if a corresponding outgoing link node exists.

The intelligent health optimization system problem solver uses an adapted argumentation algorithm (Osipov, 2016), which allows not only to confirm or reject a hypothesis, but also to minimize the set of generated hypotheses. This speeds up the system and enables rapid formation of a final set of hypotheses in order to activate the recommendation node. The input to the algorithm is a subset of the semantic network nodes set, which is formed on the basis of system user health data. For each user the system generates a number of hypotheses about his risks of non-communicable diseases. Subsequently, on the basis of his disease risk level, a number of recommendations are issued as hypothetical solutions (for example, regarding motor activity, nutrition, etc.).

5. RECOMMENDATIONS AND THEIR MODIFICATION

The system performs automatic selection of recommendations for lifestyle changes with the help of the knowledge base and on the basis of such user data as health status, lifestyle, and individual characteristics. The logical conclusion implemented in the health optimization system offers those recommendations that most fully apply to the risk factors of a particular user, taking into account all the information received about him. The recommendations are intended to modify lifestyle and are ranked based on the disease risk level and the number of RFs that affect this disease risk. Recommendations may suggest changing modes of work, rest, activity, type of food, etc. Recommendations are based on expert opinions, obtained with the help of automated semantic analysis of research conducted in different countries, and other methods of artificial intelligence.

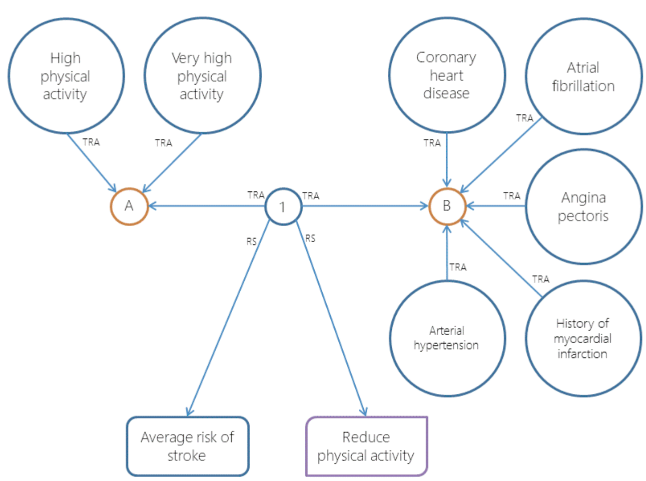

Fig. 2. Fragment of the stroke knowledge base (with heart disease in anamnesis)

A, B - hidden nodes "OR"

1 - hidden node "AND"

TRA is a link indicating that a node with an inbound link is always observed if a corresponding outgoing link node exists.

RS is a link indicating that a node with an inbound link can be observed if a corresponding outgoing link node exists.

Recommendations are represented by nodes (Fig. 2) associated with risk factors or their combinations, the effects of which they should minimize. The algorithm ranks the recommendations in a sequence that takes into account the number of factors covered and the levels of risk these factors produce. Initially, the recommendations are sorted by risk level from highest to lowest. Next they are sorted by the number of risk factors: those are selected that impact more factors, the influence of which can be reduced.

Let us consider in greater detail the types of knowledge base nodes used in the system. The system has visible (isVisible = true) and hidden (isVisible = false) nodes of different types. Visible nodes are: (a) risk factors, including, as stated above, their attributes and values, (b) disease risk levels, (c) recommendations for reducing the negative impact of controlled risk factors. If necessary, visible nodes include comments as well. These comments are designated only for doctors and contain additional information about detected deviations of physiological parameters or other indicators, which are risk factors. Logical linking of visible nodes is accomplished using hidden nodes. This allows integrating risk factors and treating them as a comprehensive indicator. Nodes can be of three types: normal (nodeType = NoneType), AND (nodeType = AND), OR (nodeType = OR). A node of the usual type corresponds to visible nodes. AND/OR nodes correspond to hidden nodes. They are used to implement integrating logic operations.

Consider the implementation of links between nodes. Three different types of links define the relationship between any nodes in the knowledge base. A positive moderate (soft) link between nodes, referred to as RS, corresponds to the situation when a node with an incoming link can be observed if a node with outgoing link is also observed. A positive strong link, referred to as a TRA, corresponds to the situation where a node with an incoming link always exists when a node with outgoing link is also observed. A negative strong link between nodes, referred to as S, is used in cases where a node with an incoming link is never observed when a node with outgoing link is observed.

TRA connection is always directed from AND/OR general nodes to other nodes. In contrast, nodes of type OR are characterized by the presence of incoming TRA connections.

The RS connection type is used for communication with hypotheses formed in the knowledge base.

The S connection type is used only for relations between hidden nodes in cases where a node with an incoming link is never observed when a node with outgoing link is observed.

A logical conclusion from the knowledge base can be derived on the basis of an "arbitrary" set of the currently available user data, i.e. in terms of limited information. Recommendations are modified when there are changes in arguments (RF), and the level of the disease risk depends on the user fulfilling the recommendations issued for disease prevention. Various argumentation tools in medical subject area applications (Kobrinskii, 2014) allow to take into account various situations and risk factor changes. Various hypothetical recommendations for preventive measures are obtained as a result.

A decision on the choice of hypothesis is based on a superior number of risk factors, which are arguments for making a decision on whether to confirm or reject one of the compared hypotheses.

In the heterogeneous semantic network that is formed, recommendations, as well as risk levels, are hypotheses. Links to recommendations can be made from both visible nodes and hidden nodes (AND/OR). The sequence of inferences that leads to the recommendation is affected by the representation of knowledge by experts. For example, due to the risk factor associated with smoking, the recommendation to "quit smoking" can be activated directly from the "currently smoking" node. However, the same recommendation will be present in the list of hypotheses if the corresponding node with some level of risk is activated.

In order to assess the risk of stroke, in addition to directly ascertaining whether the patient smokes or not, information on the number of cigarettes smoked per day and on the duration of smoking is required. To assess heart attack risk, less information is needed - according to experts, duration of smoking does not affect the risk level, so it's enough to know only the number of cigarettes smoked per day.

An example of a hidden node connection is the recommendation "to reduce motor and/or physical activity" while simultaneously activating the nodes "high or very high motor activity" AND "heart disease in anamnesis". In this example, there is also a direct link from the hidden node to the node with the stroke risk.

The algorithm for selecting recommendations to reduce the influence of risk factors consists of the following steps:

Input: an array of nodes of the heterogeneous semantic network, with values set to (YES), an array of nodes of the heterogeneous semantic network, with values set to (NO). Node values are based on user data.

Output: a subset of hypotheses relating to preventive recommendations.

Step 1. The inference rules are defined in the software module.

Step 2. Size of the observable nodes array is stored in the variable Size, variable previousSize = 0.

Step 3. Cycle 1 while Size is not equal to previousSize.

Step 3.1. Cycle 2 on all nodes from previousSize to Size. Denote each such node as $e \in O$.

Step 3.1.1. If the node e is not a hypothesis, then apply rule 1 and rule 2, otherwise apply rule 1.

Step 3.2. Update the value of previousSize, which will be equal to the size of O, before applying rules 1 and 2.

Step 3.3. Get the set of hypotheses formed in steps 3.1.1 and apply rule 4 to this set.

Step 3.4. Update Size.

Step 4. Get a set of hypotheses - H.

Step 5. For each hypothesis from the set of hypotheses H, apply rule 3.

Step 6: Update a set of hypotheses - H.

Step 7. For each hypothesis from the set of hypotheses H, apply rule 5.

Step 8. Apply the algorithm to minimize the explanatory set.

Step 9. Return a subset of hypotheses related to preventive measures.

The recommendations produced by the health optimization system first go to the doctor (health manager). The latter has the right to make changes in accordance with the individual characteristics of the patient known to him. After that, the system user receives them, and the adjusted version of the recommendations is sent to a special storage module. In future, the nodes and links of the knowledge base can be adjusted on the basis of these changes.

6. HEALTH MANAGEMENT

Management as a human activity has always existed. Usually, the term "management" implies the presence of a certain structure or object, such as a state, corporation, system, automobile, aircraft, etc. At its core, management is any activity aimed at changing the parameters of a structure or an object. In this sense, management can be attributed to the assessment of health and prevention of detected disease risks ("health management").

The main terms and phrases used in control theory are: control object, object state parameters (characteristics), measurement of object state parameters, measuring devices, controller, control loop, devices affecting the object, direct and feedback chains, control goal (criterion) and management effectiveness. The widespread use of automation, currently observed, leads to greater complexity of both the objects themselves and the tasks of automation, which has already led to the rapid development of the science of control (Pugachev, 1974; Emel'yanov, 2009).

Managing an object is to change any of its characteristics. Thus, health, as a characteristic of the object "person", can be considered as the goal of management. Health is considered as any other value that can be observed, measured, data about which are available in Health Medical Records (HMR) and which can be learned from the person himself. As a rule, specialists in providing medical care consider the object only as a patient, client, consumer, whose opinion is important only for the correction of preventive or therapeutic actions. From the standpoint of management theory, in most situations related to the optimization of health, the person himself is the primary and main figure, that is, the regulator, indirectly influencing his own risk factors. That is, health management takes place (Dartau, 2003). The health manager is also the regulator at the stage of issuing preventive recommendations, which are formed by the intelligent system in accordance with the individual risk factors identified.

The natural desire of a person to be healthy in management theory is classified as a goal, the process of achieving which is called control. Consequently, as noted above, the object of management is the person, and the goal is formulated in relation to one of its most complex characteristics - health (health optimization).

In relation to a person as an object of management, an intelligent system of health optimization can be an integral part of the regulating unit, focused on developing recommendations and building a plan of preventive measures. The system receives data on the object of management from a variety of sources: HMR, social networking, handheld gadgets, etc. With the help of the knowledge base the system assesses the diseases risk level (based on dynamically changing data) and then forms a set of recommendations and preventive measures. The results of the knowledge base are indirectly transmitted to the control object.

The developed intelligent system assumes that between it and the end user there is a health manager (general practitioner, family doctor) who is a full participant in the process. The health manager is also a possible adjuster of recommendations issued.

At the same time, it is necessary to remember that a living organism is an open system that communicates with the external environment, is subject to external influence, and has its own internal control loops. A person is a full participant in the health management process. He participates in the assessment and correction of his state of health together with the health manager, who has the necessary knowledge about health management. The person carries out the action on the object (himself) and also evaluates the results. At the same time, he periodically refers to the health manager, who, in turn, turns to the intelligent system to form a plan of preventive measures. This equates to participatory P4 Medicine.

The management system can be made more complex by taking into account other control loops and by dividing the functions of the regulator between several subjects, depending on the specific task and specific situation.

7. CONCLUSIONS

The proposed intelligent system is focused on maintaining a stable state of personal health, i.e. on health optimization through preventive measures. For this, multiple risk factors are taken into account, the disease risk level is determined, and personally oriented recommendations are issued. The theoretical basis of the system is patient-centered eHealth and P4 medicine, including predictive, preventive, personalized and participatory directions (Flores et al., 2013; Elborn, 2013). This implies the active participation of patients in maintaining their health. Thus, the health optimization system can provide patient-focused, patient-activated and patient–entrusted health care (Dawson et al., 2009). As a result, personalized medicine on the basis of intelligent technologies can finally shift the focus to early prevention based on–scientifically proven prediction of possible diseases for each individual.

The described intelligent system of health optimization enables dynamic control of the state of health of any individual. Recommendations using the knowledge base are updated as new information becomes available. The management system automatically controls the regularity of new input from patients from all sources, including Health Medical Records. The system is scalable and open (**it** allows you to automatically supplement it with new diseases and adjust the knowledge base nodes and connections accordingly).

The system is currently being tested in two different ways. The first of these is a retrospective study of stroke patients. The system analyzes data (mainly from medical records) several years before a stroke and, based on these data, provides stroke risk assessment for patients in dynamics, up to the onset of an event. The second, prospective, includes an assessment of stroke risk, myocardial infarction and depression in "healthy" patients, followed by validation of the results by specialists.

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

Dartau L.A. (2003). Theoretical aspects of health management and the possibilities of its implementation in the Russian Federation. *Problemy upravleniya*. (2), pp.43-52. (In Russ.).

Dawson, J., Tulu, B. and T.A. Horan (2009). Towards Patient-Centered Care: The Role of E-Health in Enabling Patient Access to Health Information. In: *Patient-Centered E-Health* (Wilson E.V. (Ed)), pp.1-9. Information Science Reference, Hershey, New York.

Elborn, J.S. (2013). The impact of personalised therapies on respiratory medicine. *European Respiratory Review*. Vol.22 (127), pp.72-74.

Emel'yanov, S.V. (2009). *Selected works S.V. Emelyanova: in 2 volumes* (Korovin S.K. (Ed)), 560 p. Izdatelstvo Moskovskogo universiteta, Moscow. (In Russ.).

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

Kobrinskii, B.A. (2014). Argumentation Systems: Medical Applications. *Automatic Documentation and Mathematical Linguistics*, Vol.48 (2), pp. 78–80.

Koenker, R. and K.F. Hallock (2001). Quantile regression, *Journal of economic perspectives*, Vol.15 (4), pp.143-156.

Osipov, G.S. (1990). Formulation of subject domain models: Part 1. Heterogeneous semantic nets. *Soviet Journal Comput and Systems Sci*. Vol.30 (2), pp.1-12.

Osipov, G.S. (2016). *Artificial Intelligence Methods*, 296 p. FIZMATLIT, Moscow. (In Russ.).

Pugachev, V.S. (Eds.) (1974). *Automatic Control Fundamentals*. Nauka, Moscow. (In Russ.).

Stankevich, M., Smirnov, I., Ignatiev, N., Grigoriev O. and N. Kiselnikova (2018). Analysis of Big Five Personality Traits by Processing of Social Media Users Activity Features. *Proceedings of the XX International Conference DAMDID/RCDL 2018 (Moscow, Russia, October 9–12, 2018)*. pp.235-239.

Zhou, X.-H., Wang, X., Duncan, A., Hu, G. and J. Zheng (2017). Statistical evaluation of adding multiple risk factors improves Framingham stroke risk score. *BMC Medical Research Methodology*. Vol.17 (58), pp.1-13.