

УДК: 613.96-053.6; 004.82

Кобринский Б.А., д.м.н., профессор, зав. отделом систем поддержки принятия клинических решений, Федеральный исследовательский центр «Информатика и управление» РАН, e-mail: kba_05@mail.ru

Григорьев О.Г., д.т.н., Главный научный сотрудник, ФИЦ ИУ РАН, e-mail: oleggpolikvart@yandex.ru

Смирнов И.В., Кандидат физико-математических наук, доцент. Зав. отделом «Интеллектуальный анализ информации» Института проблем искусственного интеллекта ФИЦ ИУ РАН, e-mail: ivs@isa.ru

Молодченков А.И., к.т.н. Научный сотрудник, Федеральный исследовательский центр «Информатика и управление» Российской Академии Наук, ул. Вавилова, д.44, кор.2, г.Москва, 119333, Россия; Преподаватель, Российский Университет Дружбы Народов, ул. Миклухо-Маклая, д.6, г. Москва, 117198, Россия, e-mail: aim@tesyan.ru

Благосклонов Н.А., Инженер-исследователь, Федеральный исследовательский центр «Информатика и управление» Российской академии наук, e-mail: nblagosklonov@gmail.com

Принципы создания интеллектуальной системы формирования и сопровождения плана мероприятий по здоровьесбережению

31

Аннотация

В статье описывается создание базы знаний для интеллектуальной системы управления здоровьем при риске инсульта, инфаркта миокарда и депрессии. Основная цель базы знаний состоит в том, чтобы выявить и оценить факторы риска и ситуации, которые могут привести к таким заболеваниям, сформировать план профилактических мер. База знаний реализована с использованием гетерогенной семантической сети и использует экспертные мнения о факторах риска и событиях, влияющих на здоровье человека. Данные собираются с помощью вопросников, мобильных устройств, историй болезни и информации из социальных сетей. Анализ данных и текста осуществляется с целью оценки состояния пользователя. Все полученные данные накапливаются в единой базе данных. База знаний устанавливает факторы риска, в том числе изменения этих факторов с течением времени, а также обстоятельства или события, которые могут ускорить возникновение патологии. Гипотезы генерируются об активных факторах риска и обстоятельствах, которые могут привести к увеличению или уменьшению риска. Профилактические меры по снижению этих рисков предлагаются путем анализа выдвинутых гипотез. Рекомендации относительно профилактических мер формируются с помощью базы знаний, пользовательской библиотеки случаев и методов совместной фильтрации.

Ключевые слова:

гетерогенная семантическая сеть, база знаний по прогностической медицине, экспертные знания, семантический анализ, факторы риска, рекомендации.

Проблема здоровьесбережения выходит на первый план в связи с «омоложением» болезней, старением населения, коморбидностью патологии.

Рассмотрим принципы построения системы, которая позволяет решать следующие задачи: (1) анализ персональной предрасположенности к различным заболеваниям с оценкой риска конкретного заболевания (в пилотной версии рассматриваются Инфаркт, Инсульт и Депрессия), (2) определение индивидуальных факторов риска (ФР) повышающих вероятность этих заболеваний (с учетом возможных воздействующих событий), (3) динамическая оценка изменений в состоянии здоровья индивидуума (в рамках сведений, предоставляемых человеком и наблюдающими его медицинскими работниками), (4) ранжированные рекомендации, с учетом психологического профиля индивидуумов, направленные на уменьшение влияния управляемых факторов риска хронических заболеваний, (5) мониторинг здоровья и уровня риска пользователей системы.

Рассматриваемая система основывается на комплексном подходе, сочетая экспертные оценки состояния здоровья в рамках анализируемых факторов риска, рекомендации по снижению их уровня, а в последующем, по мере накопления больших массивов информации (Big Data) из различных источников (гаджеты для мониторинга физиологических показателей, ЭМК, вопросники, открытые данные из социальных сетей), интенсивное использование многопараметрических данных (data intensive domains) для машинного обучения [1]. Импорт деперсонифицированных данных из ЭМК будет обеспечиваться на условиях согласования с разработчиками конкретных МИС и последующего семантического анализа большого объема клинических данных для проверки выделенных ФР и поиска новых предикторов заболеваний. Создаваемая система ориентирована на интеллектуальный анализ данных для выявления и оценки уровня персональных факторов риска в их динамике.

Особенностью данной системы, в отличие от подобных, является сочетание различных методов искусственного интеллекта, включая экспертные знания [2], извлечение знаний из текстов, являющееся важным методом получения данных [3], представление знаний в виде неоднородных семантических сетей (НСС) [4], где узлы базы знаний включают атрибуты факторов риска и рекомендации по их коррекции,

логические выводы на НСС по мере изменения персональных данных, подбор рекомендаций с учетом индивидуальных особенностей и отслеживание динамики изменения состояния здоровья (уровня рисков). Рекомендации будут передаваться пользователям после просмотра их лечащим врачом и необходимой коррекции с учетом не учитываемых в настоящее время системой хронических заболеваний. Возможная реакция пациента на предлагаемые ему прогнозы и планы мероприятий предполагается анализировать с участием психологов.

Архитектурно система состоит из следующих подсистем:

- Подсистема сбора и хранения информации. Содержит модуль сбора информации о состоянии здоровья и образе жизни человека из медицинских систем, профилей социальных сетей, опросников, гаджетов а также модуль хранения информации о состоянии здоровья, образе жизни человека и социальных факторах.
- Подсистема оценки состояния здоровья, оценки рисков и выявления проблемных зон. Содержит модуль оценки состояния здоровья и выявления проблемных зон человека на основе анализа информации о состоянии его здоровья и образа жизни, а также модуль оценки персонализированных рисков ухудшения здоровья человека на основе анализа динамики изменения его состояния здоровья и образа жизни.
- Подсистема выработки рекомендаций и формирования плана профилактических мероприятий. Содержит модуль подбора рекомендаций по изменению образа жизни в зависимости от текущих показателей состояния здоровья, образа жизни, проблемных зон и индивидуальных особенностей человека, а также модуль формирования персонализированного плана профилактических мероприятий.

Подсистемы взаимодействуют через программные интерфейсы, что позволяет использовать функции системы как отдельные сервисы и интегрировать их в другие системы. Пилотный образец системы имеет веб-интерфейс для пользователя и менеджера здоровья.

Ключевой компонентой системы является База знаний. Она состоит из разделов, узлов НСС, свойств узлов и связей между узлами. Узлы НСС – основные составляющие базы знаний. Узлами могут быть конкретные: утверждения,

события, результаты наблюдений, характеристики состояния пользователя, факторы риска заболеваний, данные о предрасположенности к заболеваниям в семье, факты об образе жизни, личностные факторы. В базе знаний системы представлены следующие разделы: «Факторы риска», «Скрытые характеристики», «Риски заболеваний», «Профилактические мероприятия», «События в жизни пользователя» и др. Скрытые характеристики – это скрытые узлы, которые отвечают за реализацию операций «И» и «ИЛИ». Для решения поставленных в данной работе задач адаптирован алгоритм аргументационных рассуждений. Предусмотрен учет как непосредственных, так и опосредованных отношений между ФР и болезнями [5].

Особенностью системы является возможность определения характеристик состояния здоровья пользователя на основе интеллектуального анализа его цифровых следов, получаемых с гаджетов, из анкет-опросников и анализа его активности в социальных сетях. В настоящее время в системе реализована возможность автоматической загрузки информации из профиля пользователя в социальной сети ВКонтакте (с его согласия), включая его посты, репосты,

комментарии, лайки, фото, информацию о друзьях, сообществах и т.п. Эта информация анализируется с помощью методов семантического анализа текстов и интеллектуального анализа данных с целью выявления, например, признаков психологического неблагополучия, депрессии, а также личностных черт, являющихся предикторами психологических заболеваний. Кроме того, анализ информации из социальных сетей может использоваться для выявления особенностей образа жизни человека, например, степени интернет-зависимости или идентификации событий в жизни пользователя, являющихся факторами риска различных заболеваний.

Заключение. Пилотный образец системы включает в себя интеллектуальный анализ факторов риска и событий с построением ранжированной последовательности рекомендаций для 3-х заболеваний (инфаркт, инсульт, депрессия). Информация о здоровье людей и предлагаемых рекомендациях будет накапливаться в базе данных системы. Безопасность персональных данных на этапе тестирования первой версии реализована с использованием личного кода, исключения адреса и анкетных данных.

СПИСОК ЛИТЕРАТУРЫ:

1. Skvortsov N.A., Kalinichenko L.A., Kovalev D.Yu. Conceptualization of Methods and Experiments in Data Intensive Research Domains // XVIII International Conference, DAMDID/RCDL 2016, Ershovo, Moscow, Russia, October 11-14, 2016. P.3-17.
2. Гаврилова Т.А., Кудрявцев Д.В., Муромцев Д.И. Инженерия знаний. Модели и методы: Учебник. СПб.: Издательство «Лань», 2016. 324 с.
3. Смирнов И.В., Шелманов А.О. Семантико-синтаксический анализ естественных языков. Часть I. Обзор методов синтаксического и семантического анализа текстов // Искусственный интеллект и принятие решений. 2013. №1. С.41-54.
4. Osipov G.S. Formulation of subject domain models. 1. Heterogeneous semantic nets // Soviet J Comput and Systems Sci. 1990. Vol.30, No.2. P.1-12.
5. Кобринский Б.А. Логика аргументации в принятии решений в медицине НТИ, сер.2. 2001. №9. С.1-8.
6. Osipov G.S. Formulation of subject domain models. 1. Heterogeneous semantic nets // Soviet J Comput and Systems Sci. 1990. Vol.30, No.2. P.1-12.
7. Кобринский Б.А. Логика аргументации в принятии решений в медицине НТИ, сер.2. 2001. №9. С.1-8.

Работа выполнена при финансовой поддержке Министерства образования и науки РФ. Соглашение №14.604.21.0194 (Уникальный идентификатор проекта RFMEFI60417X0194).

UDC: 613.96-053.6; 004.82

Kobriniskii B.A., Grigoriev O.G., Smirnov I.V., Molodchenkov A.I., Blagosklonov N.A.

Principles of creating an intellectual system of forming and maintaining a plan of measures for health saving

Abstract. The article describes the creation of a knowledge base for an intellectual health management system for risk 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, form a prevention plan. The knowledge base is implemented using a heterogeneous semantic network and utilizes expert opinions about risk factors and events influencing an individual's health. Data is compiled with the aid of questionnaires, mobile devices, case histories and information from social media. Analysis of data and text is carried out 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 active risk factors and circumstances which might produce an increase or decrease in risk. Recommendations regarding prophylactic measures are formed with the aid of the knowledge base, the user case-library, and collaborative filtering methods.

Keywords: heterogeneous semantic network, knowledge base of predictive medicine, expert knowledge, semantic analysis, risk factors, recommendations.