RECOMMENDATION ALGORITHM BASED ON COLLABORATIVE FILTERING AND ITS APPLICATION IN HEALTH CARE

Igor Kulev
Faculty of Computer Science and Engineering, University "Ss. Cyril and Metodius", Skopje, Macedonia

Elena Vlahu-Gjorgievska
Faculty of Administration and Information Systems Management, University "St. Kliment Ohridski", Bitola, Macedonia

Vladimir Trajkovik
Faculty of Computer Science and Engineering, University "Ss. Cyril and Metodius", Skopje, Macedonia

Saso Koceski
Faculty of Computer Science and Engineering, University "Goce Delcev", Stip, Macedonia

ABSTRACT

This paper presents a novel recommendation algorithm that generates recommendations and suggestions for preventive intervention. Presented algorithm is part of the Collaborative health care system model called COHESY.

The purpose of recommendation algorithm is to give a recommendation for performing a specific activity that will improve user's health, based on his given health condition and set of knowledge derived from the history of the user and users like him. The aim of the recommendation algorithm is to discover which activities affect change in the value of each health parameter individually. Once revealed, algorithm can use that information in situations it recognizes as same or similar to previous health conditions of a same or another user with similar medical condition. If there is evidence in users' history, that the execution of a certain physical activity has improved user health parameters and condition, it can be concluded that the activity can help him or other users with similar health issues and improve their health condition.

In this paper we also evaluate the proposed algorithm by using generic data.

I. INTRODUCTION

Advances in communication and computer technologies have revolutionized the way health information is gathered, disseminated, and used by healthcare providers, patients and citizens. This led to a new research field called "e-health" [1]. Recent trends in health care support systems are focused on developing patient-centric pervasive environments and the use of mobile devices and technologies in medical monitoring and health care systems [2]. Such systems enable healthcare personnel to be able to timely access, review, update and send patient information from wherever they are, whenever they want [3]. In that way, pervasive health care takes steps to design, develop, and evaluate computer technologies that help citizens participate more closely in their own healthcare [4], on one hand, and on the other to provide flexibility in the life of patient who lead an active everyday life with work, family and friends [5].

The presented recommendation algorithm is part of the collaborative health care system model (COHESY) [7]. This model gives a new dimension in the usage of novel technologies in the healthcare. Cohesy uses mobile, web and broadband technologies, so the citizens have ubiquity of support services where ever they may be, rather than becoming bound to their homes or health centers. Broadband mobile technology provides movements of electronic care environment easily between locations and internet-based storage of data allows moving location of support [6]. The use of a social network allows communication between users with same or similar condition and exchange of their experiences.

Cohesy has simple graphical interfaces that provide easy use and access not only for the young, but also for elderly users. It has more purpose and includes use by multiple categories of users (patients with different diagnoses). Some of its advantages are scalability and ability of data information storing when communication link fails. Cohesy is interoperable system that allows data share between different systems and databases.

The proposed recommendation algorithm, which is part of the social network in Cohesy, is based on the dependence between the values of the health parameters (e.g. heart rate, blood pressure, arrhythmias) and the users' physical activities (e.g. walking, running, biking). The basic idea is to find out which physical activities affect change (improvement) of the value of health parameters. This dependence continues to be used by the algorithm to recognize the same or similar health situations found in another user with similar characteristics. If there is information in the users' history that after performing some physical activity their health condition has improved, the algorithm accepts this knowledge and proposes the activity to other users with similar health problems. To achieve this in the proposed recommendation algorithm classification and filtering algorithms are applied in order to group users with similar characteristics. Such use of classified data provides relevant recommendations based on prior knowledge of users with similar health conditions and reference parameters. Thus, this system recommendation affects on improving the health condition, thereby obtaining better life quality of the user.

The usage of recommendation algorithm, as part of the social network, is the main component and advantage of Cohesy which differentiates it from other health care systems. These components provide a new perspective in the use of information technologies in pervasive health care by providing relevant data from a larger group of users, grouped on the basis of various indicators.

II. RECOMMENDATION SYSTEM

Exercises can have big influence on the people's health condition [8]. However, same types of physical activities do not necessarily have the same effect on different people. Our goal is to design a system that will learn which types of physical activities might improve the health condition of the user and will use that knowledge in order to make recommendations.

Inputs to the system are the physical activities performed by the users and the readings of the health parameters (body weight, blood pressure, heart rate, blood glucose level). We need to find the dependency between the change of the people's health parameters and the physical activities that they perform.

Standard recommendation algorithms can't be used because they need data about the content of the items (content-based recommender systems) or the user-item pairs (collaborative recommender systems). In our domain we either need to know how physical activities affect the health parameters (this knowledge should be supplied by human experts) or the quantitative representation of the influence of each performed physical activity on the health parameters.

People that have some diagnosis, for example people with heart problems, should not be recommended some types of activities, for example fast running, because these activities can worsen their health condition. The system should be carefully designed in order to use efficiently all the available information.

For every person and at every moment there is a set of useful activities that can potentially improve his health condition. Our recommendation algorithm discovers and recommends these useful activities to the user. The algorithms for activity recommendation must be based on few main principles: (1) Except the physical activities, there are other factors that affect the change of the people's health condition (medicines, food and psychic condition) and the deficiency of this additional information brings to bigger inaccuracy in the recommendations; (2) People can be grouped according to their characteristics (diagnosis, place of living) and for each of these groups activities have specific effect on the parameters which is similar for people in same group, and is different for people in different groups; (3) For every person activities do not influence his/her health with the same intensity and in the same way.

We will separate the health parameters into two groups. Descriptive parameters express the factors that are important for the people's health (for example age or place of living) and control parameters express the characteristics of the people's health (for example blood pressure or blood sugar level). The goal of the proposed algorithm is to find activities that could change the control parameters towards preferable state (normal health condition).

A. Data representation and preprocessing

Primary data used by the recommendation algorithm are the physical activities performed by the users and characterized by type of activity, date, duration and intensity, and the readings of the health parameters characterized by time and value. Each health parameter needs to be one-dimensional. If the value of the health parameter is described by more than one dimension then each dimension should be seen as a distinct parameter or the parameter should be transformed into a derived one whose value is one-dimensional.

Duration and intensity of the activity are used by the recommendation algorithm to determine the amount of impact of the activity on the health parameters. The types of activities can be more general e.g. running, walking, swimming, but they can also be more specific e.g. fast running, mountain cycling. The intensity of the performed activity should not be a value specific for the activity, but an approximation of the effort given by the user in order to complete the activity (this is more important if more than one health parameter is used by the recommendation system). Probably the best estimation of the intensity is the number of calories burned while performing the activity.

The values of some health parameters can change a lot in a short amount of time. This is caused by the noise contained in the readings or by the natural behaviour of the health parameters. The health condition of the users needs to be described by stable values. This is why we propose weighted average of the readings to obtain the parameter value at a given moment.

Some of the possible parameter values indicate normal health condition, but some of them indicate worsened health condition. The information contained in the parameter values that shows how much desirable is the current state of the parameter is important in the recommendation algorithm. That is why the measurements should be mapped into new classes that will give them semantic meaning. For example, the classes could be "under normal", "normal" or "above normal" value, or if the parameter is age, they could be "young", "adult" or "old". The model which describes the mapping is done by expert. The mapping can be done by using fuzzy discretization. This technique allows each parameter value to belong to different classes with different memberships.

Every user has a health profile that can be defined as a combination of the parameter values at some moment. The current health profile is created by the combination of the most recent parameter readings. For each user a history of different health profiles is kept. These health profiles are used to compare two users if they had similar health condition at some point in the past.

B. Overview of the algorithm

The algorithm for recommendation of physical activities consists of four main phases:

Categorization of the users according to their diagnosis and filtering of all users that do not belong to the same category with the active user;

Selection of the users most similar to the active user according to the history of the health profiles by using collaborative filtering;

Calculating the usefulness of the activities to the active user and his similar users by using their health history and history of performed activities;

Generation of recommendations by using the calculated usefulness of the activities.

C. Categorization according to diagnosis

We assume that physical activities have similar effect to the people with the same diagnosis. If the conclusions for the usefulness of the activities are based only on the data from the users belonging to the same group as the active user, they would not be very accurate because the activities that have a negative effect to the group with a certain diagnosis would get much lower relevance so they would not be recommended. Categorization of the users is made to that the users having the same diagnosis and the same set of permissible activities are grouped together. In this phase we need expert knowledge to define the classes and the conditions for membership in these classes.

We need to automate this process because the expert might not be always available. All the users could be denoted by diagnosis (or other relevant characteristics) and some of the users should be assigned a class by an expert. These data could be used by some classification algorithm (i.e. decision tree, neural network) as a training set in order to assign classes to the unlabeled samples.

D. Selection of the most similar users

The second phase of the algorithm is used to select the users that belong to the same class as the active user and that are very similar to the active user regarding their health history. The selected users' data would help to get better recommendations.

In this phase we use the sets of health profiles belonging to the users with the same class as the active user. The main assumption is that if two users had the same combination of parameter values in the past, there is bigger probability that similar latent factors affect their health condition. The current health profile of the active user is analyzed and is compared with the saved health profiles of all other users. If some user has at least one health profile similar enough (according to some measure such as Euclidean distance) to the current health profile of the active user, then we declare this user as similar to the active user and his data are used in the next phase of the algorithm.

E. Calculating the usefulness of the activities

The most important phase of the algorithm is the phase where we calculate the usefulness of the activities. We are given the current health condition of the user and we want to find activities that would improve the values of the parameters that are not in the normal range.

In this phase we use the control parameters and the history of readings and performed activities by each user from the set of similar users. For each performed activity we check its influence on the change of the parameter value. Our main assumption is that the activities that happened in the interval between two measurements influenced the parameter change with intensity that depends on the moment of occurrence and that is proportional to the effort given to complete the activity.

We create a model that will represent the effect of the activity on the parameter value after its completion. Very short period of time (few days) and very long period of time (few weeks) after the completion of the activity we should not expect that the activity influenced the parameter change a lot. We should find the most suitable measurement after the activity that reflects its influence. In this phase we calculate the usefulness of each type of physical activity on the parameter change for each similar user. This is given by the formula (1):

$$V_{u,a,p} = \frac{imp_p \cdot \sum_{a_u} \left(\frac{next_p(a_u) - prev_p(a_u)}{timeSpan(duraion(a_u))} \right) \cdot v(a_u) \cdot d(u*, p) \cdot int(a_u)}{num(a_u)} \quad (1)$$

where imp_p is a coefficient that gives the importance that we give to the parameter, $prev_p(a_u)$ and $next_p(a_u)$ are the measurements that have the biggest validity before and after the activity, a_u is a variable that alters all instances of the activity type, $num(a_u)$ is the number of instances, $v(a_u)$ is the validity of the selected measurements before and after the activity a_u, $d(u*, p)$ is the direction of the desirable parameter change, this value can be -1, 0 or 1 (the parameter value of the active user should decrease, remain the same or increase), $int(a_u)$ is the magnitude of the given effort by the user (for example the distance travelled) and $timeSpan(duraion(a_u))$ is a function of the duration.

The usefulness of the activity for user u is calculated by the sum of the usefulness of the activity on each health parameter (2):

$$V_{u,a} = \sum_{p \in P} V_{u,a,p} \quad (2)$$

F. Generation of recommendations

There are few different ways to utilize the results produced in the previous phase in order to make recommendation. The simplest method is to find the most useful activity according to the formula (3):

$$V_a = \sum_{u \in U} V_{u,a} \quad (3)$$

This method is not very accurate in the case when some user has much bigger usefulness of the activity on the change of his health parameter than the other users because of the specific influence of the activity on his health condition. This value would dominate in the sum, although that activity might not have the same influence on the active user. Other method is to recommend the activity which is considered as the most useful by the most number of similar users. We used this method in our implementation.

III. EVALUATION BY USING GENERIC DATA

Our recommendation algorithm tries to find the usefulness of each type of activity on the bio-medical parameter change. Activity is considered useful if it changes the global parameter value towards the desired one. The change of a parameter value might be influenced by many factors. It is impossible to make a mathematical model that takes into account all these factors, so we tried to make a model for the parameter change, under the influence of the activities performed, that is simple and as closer to the reality as possible. We assume that each performed activity has some influence on the parameter change and that the parameter change is influenced only by the effect of the activities (pharmacological influence is neglected). We model a single activity influence to the global parameter change by a function whose shape is similar to a Poisson probability mass function.

In our experiment we consider two different types of generic activities and one parameter (for example weight). Both of them influence the parameter with the same intensity, but with different direction. The first of them has stimulatory and the second has inhibitory effect on the parameter. In our experiment we generate random activities and the time interval between consecutive activities comes from uniform probability distribution. The change of the parameter value when two activities with different directions occur in one-month interval is shown on Fig. 1.

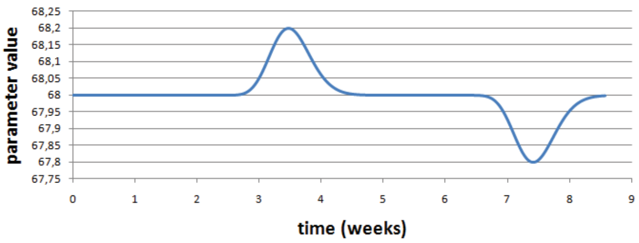

Figure 1: Change of the parameter value caused by the execution of two activities. The first activity has a stimulatory effect on the parameter and occurs between the third and the fourth week and the second activity has an inhibitory effect on the parameter and occurs between the sixth and the seventh week.

The main goal of our experiment is to test the accuracy of the proposed algorithm and the way it changes when we increase the number of activities and measurements in the training set. We generated random activities with interval between consecutive activities which comes from a uniform distribution $U(0,4)$ where these values represent days. These data are shown on Fig. 2. We also generated measurements with interval between consecutive measurements which comes from a uniform distribution $U(0,16)$ where these values represent days.

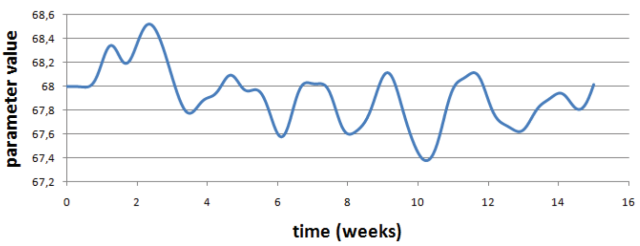

Figure 2: Change of the parameter value caused by the generated activities. The average interval between consecutive activities is two days.

After we generated the data, we tested the recommendation algorithm. In each experiment we chose a different size of the data set and we applied the recommendation algorithm trying to find recommendation about activity that will have stimulatory effect. If the algorithm recommended the type of the activity which we modelled as stimulatory we marked this trial as positive. We made 1.000 trials in each experiment in order to find the percentage of correct predictions. The results of our experiments are shown on Table 1.

Table 1: Accuracy of the recommendation algorithm as a function of the data set size.

Weeks	number of activities	number of measurements	accuracy
1.44	4	2	61.2%
2.16	7	2	66.5%
2.89	9	2	71.3%
3.61	12	2	74.2%
4.33	14	3	80.6%
5.05	17	4	87.6%
5.77	19	4	92.6%
6.49	22	5	94.6%
7.21	24	6	96.1%
7.94	27	6	97.3%
8.66	30	7	98.1%
9.38	32	7	99.1%
10.10	35	8	99.2%
10.82	37	9	99.6%
11.54	39	9	99.6%
12.26	42	10	99.7%
12.99	45	11	100.0%
13.71	47	11	100.0%
14.43	44	11	99.9%

The results show that the accuracy of the recommendation algorithm increases as we increase the number of activities and measurements. Using our model and the proposed recommendation algorithm we conclude that 20 activities and 4 measurements are enough to obtain 90% accuracy.

IV. Conclusion and Future Work

In this paper we present a recommendation algorithm as a part of collaborative health care system model - COHESY. The main purpose of this algorithm is to find the dependency of the users' health condition and physical activities he/she performs. To achieve this we consider datasets from the health and physical activities history of users and use classification algorithm on these datasets for grouping the users based on their similarity.

After we select the most similar users, we analyze the current health condition of the active user and we decide which parameters are not in the normal state and need to be improved. Then we measure the usefulness of each type of activity on the similar users assuming that they have the same current health condition as the active user. In the last phase we utilize the obtained usefulness in order to find the activity which will be recommended to the active user.

To evaluate the proposed recommendation algorithm a model of an activity influence to the global parameter change was used. The experiments conducted with generic data show that the accuracy of the algorithm increased with the duration of the observation i.e. with the number of activities and measurements observed.

ACKNOWLEDGEMENT

This work was partially financed by the Faculty of Computer Science and Engineering at the "Ss. Cyril and Methodius" University.

REFERENCES

[1] Eysenbach G. What is e-Health?. Journal of Medical Internet Research, vol 3, no 2, pp: e20. 2001.

[2] Koufi V., Malamateniou F., Vassilacopoulos G. A medical diagnostic and treatment advice system for the provision of home care. In Proceedings of 1st international conference on PErvasive Technologies Related to Assistive Environments, USA: ACM, pp:1-7. 2008.

[3] Chittaro L. Visualization of patient data at different temporal granularities on mobile devices. In Proceedings of Working conference on Advanced visual interfaces, USA: ACM, pp:484-487. 2006.

[4] Ahamed S.I., Haque M.M., Khan A.J. Wellness assistant: a virtual wellness assistant using pervasive computing. In Proceedings of Symposium on Applied Computing, USA: ACM. 2007.

[5] Ballegaard S.A., Hansen T.R., Kyng M. Healthcare in everyday life: designing healthcare services for daily life. In Proceedings of Conference on Human Factors in Computing Systems, USA: ACM, pp:1807-1816. 2008.

[6] Vlahu-Gjorgievska E., Trajkovik V. Towards Collaborative Health Care System Model – COHESY. In Proceedings of the 2011 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, IEEE Computer Society, Washington, DC, USA, pp:1-6. 2011.

[7] Taylor, C., Dajani, L., "The future of homecare systems in the context of the ubiquitous web and its related mobile technologies", In Proceedings of the 1st international conference on PErvasive Technologies Related to Assistive Environments, Athens, Greece, pp.1-4, 2008

[8] King, A. C., Taylor, C. B., Haskell, W. L., & DeBusk, R. F. (1989). Influence of regular aerobic exercise on psychological health: A randomized, controlled trial of healthy middle-aged adults. Health Psychology, 8(3), 305.