Recommendation System based on CBR algorithm for the Promotion of Healthier Habits

Gineth M Cerón-Rios, Diego M Lopez-Gutierrez,
Belén Díaz-Agudo, and Juan A. Recio-García

Department of Telematic Engineering
Universidad del Cauca Popayán, Colombia
email:gceron@unicauca.edu.co, dmlopez@unicauca.edu.co
Department of Software Engineering and Artificial Intelligence
Universidad Complutense de Madrid, Spain
email: belend@ucm.es, jareciog@fdi.ucm.es

Abstract. Recommender systems are becoming very popular as they are able to predict the preferences of a user. This make recommendation based on the user profile, past ratings or/and additional knowledge such as user contextual information. Applied to the health area, they can take advantage of context information to support health promotion and disease prevention.
We present a recommender system for the promotion of physical activity called CoCARE. It recommends videos about physical activity based on a user profile, his/her context. The main challenge of CoCARE is the small set of videos to be recommended, because the selection of the videos is done manually by of health experts. Several health recommender systems have this same problem. Although today there are a large number of videos available on the Internet related to physical activity. These could not be included in the data base of CoCARE; because these do not have enough information to be categorized and profiled.
This article proposes a CBR system, this assigns a physical activity category to new video. In this way the new video will be added to the list of CoCARE recommendations. In this CBR process, basically consists on analyzing the description of the new video and compare it with the cases base of CoCARE, selecting the category of most similar cases.

1 Introduction

Recommender systems in the health area have been proven as useful tools to help patient-oriented decision making systems, promoting physical activity and disease prevention, in general, to improve health conditions through healthier habits. Health recommender systems (HRS) aim to promote health programs, to provide patients with relevant information, products or services, using knowledge about his/her personal health record systems [1].

In the literature, we find few systems that recommend health educational multimedia contents [2–6]. Users get recommendation of exercises (stretching,

Copyright © 2017 for this paper by its authors. Copying permitted for private and academic purpose. In Proceedings of the ICCBR 2017 Workshops. Trondheim, Norway

strengthening,etc.), with outdoor or indoor sessions, based on the user information taken from mobile devices, activity bracelets, sensors, and his/her personal health records and risk factors [7].

We have developed **"CoCARE"** a platform for promotion of healthy lifestyle on the basis of a context-aware recommendation system designed for mobile smart devices [8]. Advancements in technology, mobile devices, sensors, and wearable devices, provide users with self-monitoring dynamically acquired information of her physical activities. CoCARE recommends multimedia content of physical activity and healthy diet based on a user-context model. Given a user profile and a category, the system recommends some videos about convenient physical activities for this user at this moment. Our system relies on an initial database of activity videos that are labeled with information used during the recommendation process. Currently the system has a limited number of videos that have been manually acquired from experts in the health area.

CoCARE has a database with 80 videos. These have been tagged with its title, description, category and keywords (see example in Table 1). One video could be recommended to several users based on a decision model given by domain expert.

Title	Description	Category	Keywords
Physical Therapist Shows How To Walk Correctly	Rehab and Revive Physical Therapy We can and we will get better together! Orange County Physical Therapist and Certified Functional Manual Therapist, Dr. Lin talks about how the hip, the legs,and the arms correlate to proper walking and how it can help you walk more efficiently. Proper walking helps prevent pain and other chronic injuries.	Walk	advance, amble, foot it advance, amble, foot it.

Table 1: Example of CoCARE Videos Description

Concretely, the decision model is built from a dataset of 597 instances (rows), 6 attributes and 1 main class (see Table 2) created by the expert. CoCARE builds a decision tree using a supervised learning algorithm. Then, the system classifies the query with information about the user and his/her context and uses the tree to recommend contents based on the current user situation [9].

In this paper we deal with the problem of video acquisition and tagging. Internet provides with a huge amount of videos, most of free use, related with physical activities: dancing, running, fitness, GAP. Our goal is to use these videos as recommendation items in our system. To do that, we would need to annotate the videos with information about the potential users that would benefit from them. We propose a CBR system to automatically classify videos given its textual description. This CBR system also computes similarity between the CoCARE user profiles set and the new video categories, to found its categories.

The paper runs as follows. Section 2 describes the recommender system of CoCARE based on decision model. Section 3 explains the CBR process to automatically annotate new videos. Section 4 evaluates the CBR system. Section 6 concludes the paper and discusses some lines of future work.

2 CoCARE

CoCARE (see Figure 1) is a context aware recommender system that recommends videos on physical activity (PA) and healthy diet (HD) to patients for promotion of her healthy habits. CoCARE incorporates a context- adaptable interface based on decision trees.

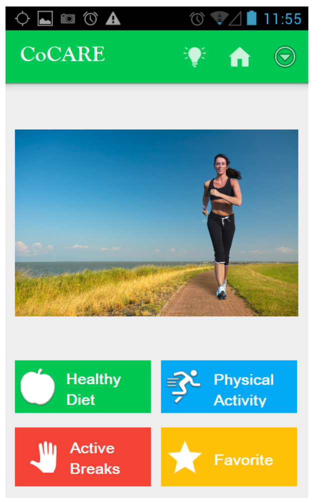

Fig. 1: Mobile CoCARE

CoCARE recommends multimedia content of physical activity and healthy diet based on user and contextual information. The basic user model includes details on the user personal profile (see Table 2). The system takes advantage of additional contextual factors to provide with personalized recommendations of multimedia content. The query includes static information like user profile, and dynamic features like geo-lacation or indoor location, date (day or season), daily schedule of the user and it can detect when the user has company. [8].

Although the CoCARE system works well as a prototype, it relies on an initial video database of 80 videos. That means different problems:

- Users get repeated contents after a while.
- Lack of novelty contents provokes user desertion.
- The task of including new videos is cumbersome.
- New videos were included without expert supervision and they were misclassified and never recommended to the right users.

We propose a CBR solution to solve these problems and assign tags (video category, keywords and user profile tags) to new videos based on the comparison to existing ones.

Attribute	Description	Value
BMI	Represents a previous inference in the user's physical condition. It is a nominal fact type.	Low weight, normal weight, overweight, obesity.
Life cycle	It is a model fact inferred from their date of birth. Despite being an integer, is taken as a nominal value in the relationships table.	Teenagers, adult.
Ethnicity	Indicates the racial group a person belongs.	It is a nominal value. Indigenous, afro, other.
Trauma	It represents a person with disability. It is a nominal fact.	Mobility, visual, auditory, without trauma.
Preference	It is the aim of the system user. It is a nominal fact.	Health, beauty, sport.
Cardiovascular disease	Indicates a user clinical condition. It is a nominal fact.	Diabetes, hypertension, without risk.
Category	It is the class of dataset. It is a nominal fact.	Dance, walk, bodily exercises, stretch, stretch eyes, limbs, personal hygiene, HIIT, labours, labours limbs, labours eyes, LISS, swim, eyes,relaxation, SCC, jog.

Table 2: User Profile Attributes

3 CBR process

Figure 2 shows the CBR process. To classify new videos, we have implemented two sequential CBR systems. The first CBR system (CBR1) receives the description of new video and returns categories from similar videos. The case base is gathered from the CoCARE video data base and contains 80 instances. Each case is described by 6 keywords and its solution is a category tag (see categories in table 2). For example:

- Description = prancing, tapping, dribbling, moving, braiding, waltz.
- Solution = dancing.
  This CBR1 module implements a k Nearest Neighbour algorithm to find the most suitable categories for a given video description. Concretely, we use a 3-NN algorithm with a keyword based similarity measure to select the three categories with highest similarity values.

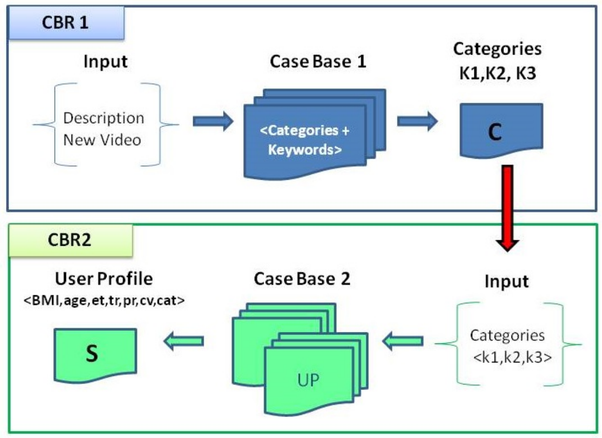

Fig. 2: CBR Process

Once the categories have been retrieved from the first case base, the second CBR module(CBR2) estimates the most suitable user profiles for the video description.

This second module has a case base with 597 instances where every case is described by several categories and 6 user profile attributes as the solution (see table 2) . By this way, the solution will be a new user profile **UP**. The algorithm compares locally the similarity value for every attribute of the user profile (see table 2) using majority voting or weighted majority voting to select the fit attribute to the solution.

$$\mathcal{CB}^1 = < C_1^1, C_2^1, \ldots, C_m^1 > \quad (1)$$
$$\mathcal{C}_i^1 = < keywords, categories > \quad (2)$$
$$\mathcal{CB}^2 = < C_1^2, C_2^2, \ldots, C_n^2 > \quad (3)$$
$$\mathcal{C}_i^2 = < categories, UP > \quad (4)$$
$$\mathcal{UP} = < BMI, age, et, tr, pr, mc, category > \quad (5)$$

4 Evaluation

We evaluate our system using leave-1-out cross-validation.

We used the video description of each case on **CB** as a query. We proposed the category and profile for UP_r, that is compared to the stored solution UP_q. We compute the similarity between attributes of UP_q and UP_r, using a binary function [0,1]. We compared if the attributes of the retrieved profile UP_r are equal to attributes of test case UP_q. So we calculated the α value (see the equation ec. 6)

$$
\begin{aligned}
Sim(UP_q, UP_r) &= \alpha \\
&\text{where} \\
a &= Sim(BMI_q, BMI_r)\epsilon[0,1] \\
b &= Sim(age_q, age_r)\epsilon[0,1] \\
c &= Sim(et_q, et_r)\epsilon[0,1] \\
d &= Sim(tr_q, tr_r)\epsilon[0,1] \\
e &= Sim(pr_q, pr_r)\epsilon[0,1] \\
f &= Sim(mc_q, mc_r)\epsilon[0,1] \\
g &= Sim(category_q, category_r)\epsilon[0,1] \\
\alpha &= 0.1 * (a + b + c + d + e + f) + 0.4 * g \qquad (6) \\
& \qquad (7)
\end{aligned}
$$

The table 3 shows an example. Our query in this example is UP_q and **D**= “Steve and Jackie take you through how to get the most out of power walking and show you how beneficial it truly is. Yes it is an Olympic sport!”.

First the CBR1 module got 3NN categories as: Walk, Exercises and HIIT. So the CBR2 module compared the UP associated to these categories and found the most similarity UP_r. Next the system uses cross validation and retrieves a success solution only if the similarity measure $\alpha \geq 0.7$. This process is the comparison between attributes UP_q and UP_r. For example the table 3 shows a test case, we obtained a score of α greater than 0.7, so UP_r was added to **CB**.

Test Case	BMI	age	et	tr	pr	cv	V
UP_q	normal weight	Adult	other	without trauma	beauty	diabetes	Walk
UP_r	overweight	Young	Other	without trauma	beauty	diabetes	Walk
Test Case	a	b	c	d	e	f	g
Test Case	0	0	1	1	1	1	1

Table 3: Example cross validation

In each leave-1-out step, we obtained 3 values: the similarity of the best cases returned by the CBR1 module (1-NN), the 2 best cases (2-NN) and the 3 best cases (3-NN). Next we made 2 tests with Majority Voting (MV) and Weighted Majority Voting (WMV) in CBR2 module. Figure 3 shows results from our experiment as:

- Case 1, it is represented by the blue bar. We found the user profile with majority voting (MV) for 1NN, 2NN and 3NN.
- Case 2, it is represented by the red bar. We found the user profile with weighted majority voting (WMV) only for 2NN and 3NN.

We conclude that the better result from our CBR system was 3NN with WMV. In this case, we obtained a value of α greater than 0.85 surpassing the results achieved of the other tests. Our experiment shows that greater than 90% of the cases are correctly classified.

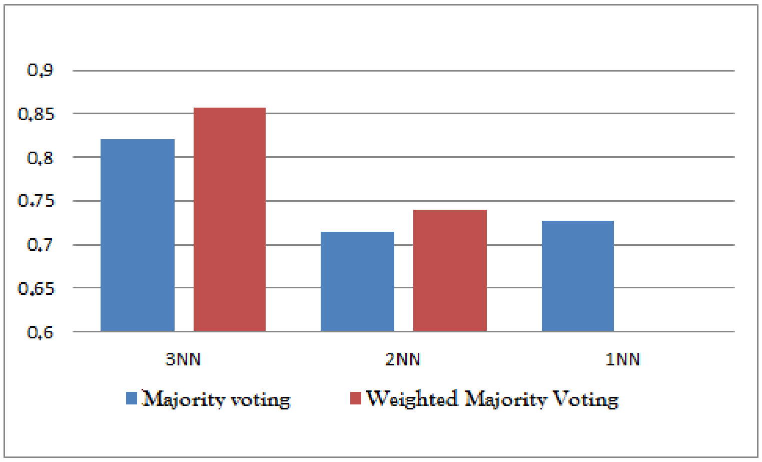

Fig. 3: Average of similarity α

5 Related works

The literature to continuation are about design and implementation of CoCARE. We approached works of health recommendation systems.

Related works such as [3, 4, 6, 10, 11] are mobile context platforms that integrates sensor technology, cognitive tutoring and evidence-based social design for health promotion. User selects group activities (jogging, walking, fitness, and yoga) to make recommendations about stretching exercises, outdoor strengthening or others, based on gender, age, weight, height, location. Diabeticlink [5], is a mobile recommender of videos and articles about exercise and healthy diabetic diet, based on user data and data of sensors. It uses the collaborative filtering recommendation technique. Finally, it generates progress reports based on the user blood glucose, his/her lifestyle, body mass index and time of physical activity. Kalico [12] is a mobile recommender system of healthy food restaurants, the user suggestion are based in his/her profile, location, budget and preferences. It provides a list of nearby restaurants in alphabetical order and presents a list of healthy menus recommendation in each it. Kalico is a system that only promotes healthy eating and for people who want to eating out.

The previously works mention the use of user models, data modeling and use of recommendation techniques, but they do not describe the selection of recommendation techniques and how performance the validation them. On the other hand, these works didn't make mention about how many recommendations have their systems, apparently the systems has few contents to recommender.

The next related works relate with CBR topic. [13, 14], these works use a CBR algorithm to recommend diabetes care videos to adults, however there is no evidence that their systems can retrieves additional information from the videos description. There are other systems that retrieve textual information [10, 15, 16], recover the sentences that are necessary and complete the sentence, others recover symptoms of some disease with the user profile. However, there is no evidence a system that finds user profiles with just the description of an item (video).

Our CBR systems retrieves a user profile from the video description. This could be useful in others areas such as education, commerce and / or advertising. For example one recommendation systems could find a user profile fit for learning content or advertising videos with just the description of item.

6 Discussion and Conclusions

We have described our CoCARE recommender system. CoCARE recommends videos of physical activity categorized by health experts. But the problem is that they are very few, to include a new video must be properly categorized for a user profile. Our CBR allows you to categorize the video and find an appropriate profile from the description of a video. In this work we proposed a system composed of 2 CBR system, the first categorizes the new video and the second delivers the appropriate profile. We have evaluated that the CBR system delivers a better response if the first CBR is 3NN and CBR2 is with similarity.

Our CBR system uses little input knowledge to get an adequate solution. It offers a simpler alternative to associate videos to the needs and preferences of different users.

Our system benefits the user and the health expert, with the possibility of having new recommendations that help the adherence of the physical activity program.

The work presented in this paper opens several lines of future work.

When you have very short video descriptions the CBR system loses precision in finding the right category, although the results we obtained are very promising we have considered that they can be improved if we extend the description from synonyms using an ontology of synonyms and algorithms matching of learning.

We plan to take information about the most viewed videos on the Internet (YouTube) and use their description to classify them, assign to the new video an appropriate user profile and add to CoCARE case base **CB** automatically using collaborative filtering and CBR.

7 Acknowledgments

This work was performed under the doctoral thesis "Context-Aware Recommender System to Physical Activity Promotion" financed by Colciencias, under call "Convocatoria No 6172 (Doctorados Nacionales)". Supported by UCM (Group 921330) and Spanish Committee of Economy and Competitiveness (TIN2014-55006-R).

References

1. Wiesner, M., Pfeifer, D.: Health recommender systems: concepts, requirements, technical basics and challenges. International journal of environmental research and public health **11**(3) (2014) 2580–2607
2. Adomavicius, G., Mobasher, B., Ricci, F., Tuzhilin, A.: Context-Aware Recommender Systems. AI MAGAZINE **32**(4) (2011) 217–253
3. F, B., L., C.: Mopet: a context-aware and user-adaptive wearable system for fitness training.. Artificial Intelligence in Medicine **42**(2) (2008) 153–163
4. Lamprinos, I., Papadaki, C.: Mobile personal health application for empowering diabetic patients: The case within empower project. In: Proceedings of The International eHealth Telemedicine and Health ICT Forum. Luxembourgh. (2013)
5. Chen, H., Compton, S., Hsiao, O.: Diabeticlink: A health big data system for patient empowerment and personalized healthcare. In: Smart Health, Springer (2013) 71–83
6. Du, H., Youngblood, G.M., Pirolli, P.: Efficacy of a smartphone system to support groups in behavior change programs. In: WH '14 Proceedings of the Wireless Health 2014 on National Institutes of Health, ACM (2014) 1–8
7. Mettouris, C., Papadopoulos, G.A.: Contextual Modelling in Context-Aware Recommender Systems: A Generic Approach. Web Information Systems Engineering WISE 2011 (2013) 41–53
8. González, K.X., Carvajal, M.Á., Cerón, G.M., López, D.M.: Modelado de un sistema consciente del contexto para soportar intervenciones en actividad física y nutrición saludable. Revista Ingenierías Universidad de Medellín **15**(28) (2016)
9. Girón, E., Cerón, G.M., Rico, C., López, D.M.: Framework for data model to personalized health systems. In: 2016 8th Euro American Conference on Telematics and Information Systems (EATIS). (2016) 1–5
10. El-Sappagh, S.H., El-Masri, S., Elmogy, M., Riad, A.M., Saddik, B.: An ontological case base engineering methodology for diabetes management. Journal of Medical Systems **38**(8) (2014) 67
11. Cochrane, K., Schiphorst, T.: Developing design considerations for mobile and wearable technology: M-health applications that can support recovery in mental health disorders. In: Proceedings of the 9th International Conference on Pervasive Computing Technologies for Healthcare. PervasiveHealth '15, ICST, Brussels, Belgium, Belgium, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering) (2015) 29–36
12. Anwar, M., Hill, E., Skujins, J., Huynh, K., Doss, C.: Kalico: A smartphone application for health-smart menu selection within a budget. In: International Conference on Smart Health, Springer (2013) 113–121

13. Bermingham, A., O'Rourke, J., Gurrin, C., Collins, R., Irving, K., Smeaton, A.F.: Automatically recommending multimedia content for use in group reminiscence therap. In: Proceedings of the 1st ACM International Workshop on Multimedia Indexing and Information Retrieval for Healthcare. MIIRH '13, New York, NY, USA, ACM (2013) 49–58
14. Polignano, M., de Gemmis, M.: Recommender systems supporting decision making through analysis of user emotions and personality. In: DC@ AI* IA. (2015) 43–48
15. Linckels, S., Meinel, C.: Applications of description logics to improve multimedia information retrieval for efficient educational tools. In: Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval. MIR '08, New York, NY, USA, ACM (2008) 321–328
16. Sandhu, R., Kaur, J., Thapar, V.: An effective framework for finding similar cases of dengue from audio and text data using domain thesaurus and case base reasoning. Enterprise Information Systems **0**(0) (0) 1–18