

# Towards Education and Emotion Based Semantic Group Recommendations for Health

Haridimos Kondylakis<sup>1</sup> and Kostas Stefanidis<sup>2</sup>

<sup>1</sup> ICS-FORTH, Greece

kondylak@ics.forth.gr

<sup>2</sup> Tampere University, Finland

konstantinos.stefanidis@tuni.fi

**Abstract.** Nowadays, more and more people are using the Web to search for health information. However, it is widely accepted that it is really hard for people to determine the quality of the presented information and to accurately judge on the relevance to their own condition. The FairGRecs system recommends to small groups of persons health documents selected by caregivers. The system exploits ontologies to model patient profiles and documents content, and then it uses a notion of semantic distance between patients in order to provide useful recommendations by incorporating the notion of fairness. In this paper, we describe the next step in this direction, namely adapting recommendations considering the educational level of the end-users and their psycho-emotional status.

**Keywords:** Recommendations · Group Recommendations · Health Recommendations.

## 1 Introduction

During the last decade, the number of users who look for health and medical information online has dramatically increased. However despite the increase in those numbers, it is very hard for a patient to accurately judge the relevance of some information to his/her own case and to identify the quality of the provided information. On the other hand, existing health information services (e.g. WebMD, MayoClinic Patient Care, Medicine Plus, HONSearch, PHIR [1, 6, 7]) consider only a limited amount of personal information. An optimal solution for patients would be to be guided by healthcare providers to resources of high quality, that they can easily comprehend and understand. However, healthcare providers have less and less time to devote to their patients. As such, guiding each individual patient appropriately is a really difficult task. On the other hand, the use of group-dynamics-based principles [9, 8, 14, 11] of behavior change have been shown to be highly effective leading to enhanced discussions and social support. However, identifying information for a group of participants is really challenging.

---

Copyright ©2019 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

FairGRecs [15,16] focuses on recommending interesting health documents selected by health professionals, to groups of users, incorporating the notion of fairness, using a collaborative filtering approach. The overall approach is based on a notion of semantic distance between documents and user profiles. Our motivation for this work, is to offer a list of recommendations to a caregiver who is responsible for a group of patients. The recommended documents need to be relevant to the patients current profiles. To exploit patients profiles, we use the data stored in individual accounts of personal health-care record (PHR) after acquiring informed consent from users.

However recommendation algorithms so far ignore the fact that patients profiles are multifaceted. For example, recommending the proper document should not only focus on the patients relevant problems but also on their health literacy (namely, the ability to obtain, read, understand, and use health care information in order to make appropriate health decisions and follow instructions for treatment), educational level and psychoemotional status, as emotions can greatly affect the cognitive processes. In this paper, we explore those dimensions, as well paving the way for a new system incorporating all aforementioned aspects.

## 2 Ratings, Semantic Distance, Health Literacy and Psychoemotional Status into the Mixer

The first step in exploiting profile information, is to be able to record it. To this purpose, specific short validated questionnaires [12] have been used that are being answered by the members of a group. All information captured is then modeled and stored using an ontology [2]. After answering those questionnaires, specific values are automatically calculated and stored in patient profiles regarding those key profile areas. Among others, numerical scores (1 to 5) exists for health literacy level, educational level, cognitive closure and anxiety that we further use for providing recommendations.

Furthermore, for documents, we also need to have information regarding the target population concerning the 4 aforementioned dimensions. As such, all documents entered by the caregivers are annotated with numbers regarding target population health literacy, education level, cognitive closure and anxiety. In addition, the documents are automatically annotated using ICD-10 ontology<sup>3</sup>, and all annotations are stored into the document corpus.

Now, given a set of data items  $I$  and a set of patients  $U$ , we need to focus first on single user recommendations. A patient, or user,  $u$  might rate an item  $i$  with a score  $r(u, i)$ . The subset of items rated by a user  $u$  is denoted by  $I(u)$ . Typically, the cardinality of  $I$  is high and users rate only a few items. For the items unrated by the users, recommender systems estimate a relevance score, denoted as  $relevance(u, i)$ . As we are using the collaborative filtering approach, similar users should be located via a similarity function that will evaluate the similarity between two users. Then items relevance scores should be computed for

---

<sup>3</sup> <http://www.icd10data.com/>

users taking into account their most similar users. The novelty of our approach lies in the fact that instead of using only classical similarity notions or based only on their diseases as in [16], we consider also the dimensions above.

## 2.1 Similarity based on ratings

Traditionally, two users are similar if they have rated data items in a similar way, i.e., they share the same interests. For calculating their similarity, we exploit the Pearson correlation metric:

$$RatS(u, u') = \frac{\sum_{i \in X} (r(u, i) - \mu_u)(r(u', i) - \mu_{u'})}{\sqrt{\sum_{i \in X} (r(u, i) - \mu_u)^2} \sqrt{\sum_{i \in X} (r(u', i) - \mu_{u'})^2}},$$

where  $X = I(u) \cap I(u')$ ,  $\mu_u$  is the mean of the ratings in  $I(u)$ .

Pearson correlation actually measures the linear dependence between two users  $u$  and  $u'$ : it has a value between +1 and 1, where +1 is total positive linear correlation, 0 is no linear correlation and 1 is total negative linear correlation.

Alternatively, in a content-like approach, users interests, or profiles, can be represented as structured, unstructured or semi-structured data. In structured profiles, there is a small number of attributes, each profile is described by the same set of attributes, and there is a known set of values that the attributes may have. Unlike structured profiles, in unstructured profiles, there are no attribute names with well-defined values. In between, in semi-structured profiles, there are some attributes with a set of restricted values and some free-text fields. A common approach to deal with free text (fields) is to convert the text to a structured representation, in which each token may be viewed as an attribute with an integer value indicating the number of times the token appears in the text. In a more sophisticated approach, each token can be associated with a tf-idf value,  $v(t, d)$ , that is, for a token  $t$  in a text  $d$ , a function of the frequency of  $t$  in  $d$ , the number of texts containing  $t$ , and the total number of texts. The intuition behind tf-idf is that the tokens with the highest values occur more often in that text than in other texts, and therefore are more important. In this scenario,  $RatS(u, u')$  can be evaluated as the cosine similarity of the vectors representing the profiles of  $u$  and  $u'$ .

## 2.2 Similarity based on semantic distance

In the health domain, usually people have similar interest in health documents if they have similar health problems. To identify similarities between health problems and eventually between users, we exploit the ICD10 ontology. We represent ICD10 as a tree, with health problems as its nodes. For a node  $A$  in the tree,  $weight(A) = w * 2^{maxLevel - level(A)}$ , where  $maxLevel$  is the maximum level of the tree,  $level(A)$  returns the level of each node and  $w$  is a constant ([16] shows that  $w = 0.1$  returns optimal results). Weights will help us differentiate between siblings nodes in various levels; we want sibling nodes in the higher levels to share greater similarity than those in the lower ones.

For computing the semantic distance between two nodes A and B, we compute their distance from the lowest common ancestor C. The distance between A and C is calculated by accumulating the weight of each node in the path, as  $dist(A, C) = \sum_{n \in path(A, C)} weight(n)$ . In overall, the similarity between A and B is:

$$simN(A, B) = 1 - \frac{dist(A, C) + dist(B, C)}{maxLevel * 2}.$$

Then, given two users  $u$  and  $u'$ , we calculate their overall similarity by taking into consideration all possible pairs of health problems between them. Specifically, we take one by one all health problems of  $u$ ,  $Problems(u)$ , and calculate the similarity with all the problems of  $u'$ ,  $Problems(u')$ , as follows:

$$SemS(u, u') = \frac{\sum_{i \in Problems(u)} ps(i, u')}{|Problems(u)|},$$

where  $ps(i, u') = max(\forall_{j \in Problems(u')} \{simN(i, j)\})$ .

### 2.3 Similarity based on education & health literacy level

For documents, regarding the same information, people have similar interest in health documents that require the same educational and health literacy level to be comprehended. As such, the similarity between two users is calculated by the Euclidean distance between the corresponding values:

$$EducStatusS(u, u') = \sqrt{(HLiteracy(u) - HLiteracy(u'))^2 + (EducLevel(u) - EducLevel(u'))^2}.$$

### 2.4 Similarity based on psycho-emotional status

Finally, anxiety and cognitive closure highly affect the documents preferred by people in specific periods of time - as anxiety and cognitive closure can fluctuate over time. As such, we use the Euclidean distance between the values of those two properties. As psychoemotional questionnaires are being answered periodically, we consider each time only the latest measurements on these:

$$PsychStatusS(u, u') = \sqrt{(Anxiety(u) - Anxiety(u'))^2 + (CognClosure(u) - CognClosure(u'))^2}.$$

### 2.5 Single User Recommendations

To compute the similarity between two users  $u$  and  $u'$ , we use the function:

$$S(u, u') = AVG(RatS(u, u'), SemS(u, u'), EducStatusS(u, u'), PsychStatusS(u, u')).$$

Then, let  $P_u$  denote the most similar users to  $u$ . The overall relevance of  $i$  for  $u$  is estimated as:

$$relevance(u, i) = \frac{\sum_{u' \in (P_u \cap U(i))} S(u, u') r(u', i)}{\sum_{u' \in (P_u \cap U(i))} S(u, u')}.$$

After estimating the relevance scores of all unrated items for  $u$ , the items  $A_u$  with the top- $k$  relevance scores are suggested to  $u$ .

## 2.6 Group recommendations

Since recommendations are typically personalized, different users are presented with different suggestions. However, there are cases where a group of people participates in a single activity. For this reason, recently, there are methods for group recommendations, trying to satisfy the preferences of all the group members. These methods can be classified into two approaches [3]. The first approach creates a joint profile for all users in the group and provides the group with recommendations computed with respect to this joint profile (e.g., [17]). The second approach aggregates the recommendations of all users in the group into a single recommendation list (e.g., [8, 13]). Our work on group recommendations follows the second approach, since it is more flexible [3, 10] and, typically, offers opportunities for improvements in terms of efficiency.

This way, our goal is to first estimate the relevance scores of the unrated items for each user in the group, and then, aggregate these predictions to compute the suggestions for the group. That is, the relevance of an item  $i$  for a group of users  $G$  is:

$$relevanceG(G, i) = Aggr_{u \in G}(relevance(u, i)).$$

As in [16], we employ 3 different designs regarding the aggregation method *Aggr*. Firstly, we consider that strong user preferences act as a veto; this way, the predicted relevance of an item for the group is equal to the minimum relevance of the item scores of the members of the group:

$$relevanceG(G, i) = \min_{u \in G}(relevance(u, i)).$$

Alternatively, we focus on satisfying the majority of the group members and return the average relevance for each item:

$$relevanceG(G, i) = \sum_{u \in G} relevance(u, i) / |G|.$$

Targeting at increasing the fairness of the resulting set of recommendations, we also use the *Fair* method. Here, we consider pairs of users in the group, in order to identify what to suggest. In particular, a data item  $i$  belongs to the top- $k$  suggestions for a group  $G$ , if, for a pair of users  $u_1, u_2 \in G$ ,  $i \in A_{u_1} \cap A_{u_2}$ , and  $i$  is the item with the maximum rank in  $A_{u_2}$ . For locating fair suggestions, initially, we consider an empty set  $D$ . Then, we incrementally construct  $D$  by

selecting, for each pair of users  $u_x$  and  $u_y$ , the item in  $A_{u_x}$  with the maximum relevance score for  $u_y$ . If  $k$  is greater than the items we found using the above method, then we construct the rest of  $D$ , by serially iterating the  $A_u$  lists of the group members and adding the item with the maximum rank that does not exist in  $D$ .

### 3 Conclusions

In this paper, we argue that common problems and ratings are not enough for capturing similarity between users, and additional properties should be considered as well, such as educational and health literacy level, anxiety and cognitive closure. All these factors highly affect the people's interest and understanding of information and especially in situations, where they are really stressed because of significant health problems.

The next step is to pilot and evaluate the system within the cancer domain. We already have a corpus available for cancer patients through the iManage-Cancer EU project [4] and also a PHR system where individual patients register and use the system. After signing the appropriate consent [5], our intention is to make available the FairGRecs mechanism to the patients, through the PHR system, offering useful recommendations to them and evaluating eventually the recommendations proposed. This will shed light to the advantages of our solution and will allow us for further refinements.

Overall, we target at a general processing model that puts humans in the core, in order to produce recommendations for health-related documents that take into consideration additional perspectives like transparency and fairness. Transparency facilitates the understanding of data through, typically, exploration and explanation, used for assisting users identify the what, where, when and how of a data item. For example, exploration can support users by offering sophisticated discovery capabilities. Differently, explanations target at telling the story that the data has to say, by providing the reasons behind specific recommendations. Fairness in data processing can be expressed as the lack of bias, where bias can come from data processing methods that reflect the preferences of the data scientists designing them. Regarding fairness in group recommendations, the goal is to locate, when possible or helpful, suggestions that include data items fair to the members of the group. That is, we should be able to recommend items that are both strongly related and fair to the majority of the group members

### Acknowledgement

This work has been partially supported by the Virpa D project funded by Business Finland and by the BOUNCE project that has received funding from the European Unions Horizon 2020 Research and Innovation Programme.

## References

1. Galatia, I., Haridimos, K., Lefteris, K., Maria, C., Eleni, K., Kostas, M., Manolis, T.: Personal health information recommender: implementing a tool for the empowerment of cancer patients. *ecancer* **12**, 851 (2018)
2. Genitsaridi, I., Marias, K., Tsiknakis, M.: An ontological approach towards psychological profiling of breast cancer patients in pervasive computing environments. In: *PETRA* (2015)
3. Jameson, A., Smyth, B.: Recommendation to groups. In: *The Adaptive Web, Methods and Strategies of Web Personalization* (2007)
4. Kondylakis, H., Bucur, A.I.D., Dong, F., Renzi, C., Manfrinati, A., Graf, N.M., Hoffman, S., Koumakis, L., Pravettoni, G., Marias, K., Tsiknakis, M., Kiefer, S.: imanagecancer: Developing a platform for empowering patients and strengthening self-management in cancer diseases. In: *30th IEEE International Symposium on Computer-Based Medical Systems, CBMS 2017, Thessaloniki, Greece, June 22-24, 2017*. pp. 755–760 (2017)
5. Kondylakis, H., Koumakis, L., Hänold, S., Nwankwo, I., Forgó, N., Marias, K., Tsiknakis, M., Graf, N.M.: Donor’s support tool: Enabling informed secondary use of patient’s biomaterial and personal data. *I. J. Medical Informatics* **97**, 282–292 (2017)
6. Kondylakis, H., Koumakis, L., Kazantzaki, E., et al.: Patient empowerment through personal medical recommendations. In: *MEDINFO* (2015)
7. Kondylakis, H., Koumakis, L., Psaraki, M., Troullinou, G., Chatzimina, M., Kazantzaki, E., Marias, K., Tsiknakis, M.: Semantically-enabled personal medical information recommender. In: *Proceedings of the ISWC 2015 Posters & Demonstrations Track co-located with the 14th International Semantic Web Conference (ISWC-2015), Bethlehem, PA, USA, October 11, 2015*. (2015)
8. Ntoutsi, E., Stefanidis, K., Nørvåg, K., Kriegel, H.: Fast group recommendations by applying user clustering. In: *ER* (2012)
9. Ntoutsi, E., Stefanidis, K., Rausch, K., Kriegel, H.: Strength lies in differences: Diversifying friends for recommendations through subspace clustering. In: *CIKM* (2014)
10. O’Connor, M., Cosley, D., Konstan, J.A., Riedl, J.: Polylens: A recommender system for groups of user. In: *Proceedings of the Seventh European Conference on Computer Supported Cooperative Work* (2001)
11. Rausch, K., Ntoutsi, E., Stefanidis, K., Kriegel, H.: Exploring subspace clustering for recommendations. In: *SSDBM* (2014)
12. Renzi, C., Fioretti, C., Mazzocco, K., et al.: Development of psycho-emotional monitoring tools within an ehealth platform to improve patient empowerment and self-management abilities. In: *Psycho-Oncology* (2016)
13. Roy, S.B., Amer-Yahia, S., Chawla, A., Das, G., Yu, C.: Space efficiency in group recommendation. *VLDB J.* **19**(6), 877–900 (2010)
14. Stefanidis, K., Shabib, N., Nørvåg, K., Krogstie, J.: Contextual recommendations for groups. In: *ER Workshops* (2012)
15. Stratigi, M., Kondylakis, H., Stefanidis, K.: Fairness in group recommendations in the health domain. In: *ICDE* (2017)
16. Stratigi, M., Kondylakis, H., Stefanidis, K.: Fairgreco: Fair group recommendations by exploiting personal health information. In: *DEXA* (2018)
17. Yu, Z., Zhou, X., Hao, Y., Gu, J.: TV program recommendation for multiple viewers based on user profile merging. *User Model. User-Adapt. Interact.* **16**(1), 63–82 (2006)