

Fairness in Group Recommendations in the Health Domain

Maria Stratigi
University of Crete
Greece
stratigi@csd.uoc.gr

Haridimos Kondylakis
ICS-FORTH
Greece
kondylak@ics.forth.gr

Kostas Stefanidis
University of Tampere
Finland
kostas.stefanidis@uta.fi

Abstract—During the last decade, the number of users who look for health-related information has impressively increased. On the other hand, health professionals have less and less time to recommend useful sources of such information online to their patients. To this direction, we target at streamlining the process of providing useful online information to patients by their caregivers and improving as such the opportunities that patients have to inform themselves online about diseases and possible treatments. Using our system, relevant and high quality information is delivered to patients based on their profile, as represented in their personal healthcare record data, facilitating an easy interaction by minimizing the necessary manual effort. Specifically, in this paper, we propose a model for group recommendations following the collaborative filtering approach. Since in collaborative filtering is crucial to identify the correct set of similar users for a user in question, in addition to the traditional ratings, we pay particular attention on how to exploit health-related information for computing similarities between users. Our special focus is on providing valuable suggestions to a caregiver who is responsible for a group of users. We interpret valuable suggestions as suggestions that are both highly related and fair to the users of the group. In this line, we propose an algorithm for identifying the top- z most valuable recommendations, and present its implementation in MapReduce.

I. INTRODUCTION

Medicine is undergoing a revolution that is transforming the nature of healthcare from reactive to preventive. The changes are catalysed by a new systems approach to disease which focuses on integrated diagnosis, treatment and prevention of disease in individuals. This will replace our current mode of medicine over the coming years with a personalized predictive treatment. While the goal is clear, the path is fraught with challenges. One of these challenges is the problem of the quality and the amount of information that can be found online [2] since health information is one of the most frequently searched topics on the Web.

During the last decade, the number of users who look for health and medical information has dramatically increased. Already from 2002, a percentage of 80% of all adults in the United States were estimated to have looked online for health information, whereas the 23% of the Europeans were using the internet to be informed about their health needs [15]. However despite the increase in those numbers and the vast amount of information currently available online, it is very hard for a patient to accurately judge the relevance of some information to his own case and the same applies to care providers.

This paper focuses on current research activities related

to the implementation of a personal medical information recommendation system within the iManageCancer EU research project¹. The project has the objective to provide a cancer specific self-management platform designed according to the needs of patient groups, while in parallel focusing on the wellbeing of the cancer patient with special emphasis on avoiding, early detecting and managing adverse events of cancer therapy but also, importantly, on the psycho-emotional evaluation and self-motivated goals. The platform is centered in a Personal Health Record which regularly monitors the psycho-emotional status of the patient and will periodically record the everyday life experiences of the cancer patient with respect to the therapy side effects, while different groups of patients and their families share information through diaries, and clinicians are provided with clinical information.

The work presented in this paper is targeted at improving the opportunities that patients have to inform themselves in the internet about their disease and possible treatments, and providing to them personalized information and recommendations. Its goal is threefold: (1) to deliver relevant information to patients, based on their current profile as represented in their personal healthcare record (PHR) data, (2) to ensure the quality of the presented information by giving medical experts the chance to control the information that is given, and (3) to facilitate an easy uptake of the new system by minimizing the necessary manual effort.

More specifically, in this paper, we propose a model for group recommendations, an area mostly unexplored in the health domain, following the collaborative filtering approach. Since in collaborative filtering is crucial to identify the correct set of similar users for a user in question, we explore different similarity measures that take into consideration specific health-related information, in addition to traditional ratings. We use different designs for aggregating the recommendations for the group that reflect different semantics; intuitively, we either consider that strong user preferences act as a veto or alternatively, we focus on satisfying the majority of the group members. Our goal is actually to provide valuable suggestions to a caregiver who is responsible for a group of patients. We interpret valuable suggestions as suggestions that are both highly related and fair to the patients of the group. Given a constraint z on the number of returned recommendations, we propose an algorithm to locate the top- z most valuable recommendations, and present its implementation in MapReduce.

¹<http://imanagecancer.eu/>

The rest of this paper is structured as follows. Section 2 presents the high-level architecture of our system, Section 3 introduces our fairness-aware group recommendation model, while Section 4 describes its implementation in Map-Reduce. Section 5 elaborates on how to compute similarities between users, and Section 6 presents a preliminary evaluation of our approach. Section 7 presents related work and, finally, Section 8 concludes the paper and discusses future directions.

II. ARCHITECTURE

The architecture of the system is shown in Fig. 1. As we can see, there are many patients using daily a Personal Health Record (PHR) system, named iPHR². Within the system, users can record and manage their problems, medication, allergies, procedures, laboratory results etc. As soon as a new problem is selected, behind the scenes, the corresponding SNOMED-CT term is saved at the database to enable interoperability and further usage. Via the available app, users can use a search engine to find useful documents selected by the experts and then, can rate the individual results. Our recommendation engine accesses the patient profiles and their document ratings and automatically recommends a set of documents that a caregiver can suggest to a set of patients. To achieve this functionality, MapReduce tasks are executed exploiting a set of similarity measures to identify similarities among users. As we will see in the sequel, we target at identifying results that are highly-related to the profiles of the patients that the caregiver is responsible for, and fair for those particular patients.

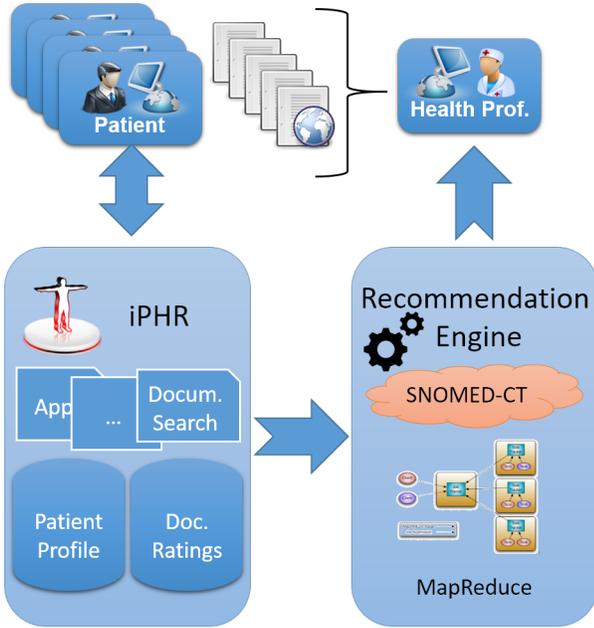


Fig. 1. The high-level architecture of our system.

III. MODEL

A. Single User Rating Model

Assume a recommender system in the health domain, where I is a set of data items to be rated and U is the set

of patients in the system. A patient, or user, $u \in U$ might rate an item $i \in I$ with a score $rating(u, i)$ in $[1, 5]$. Typically, the cardinality of the item set I is high and users rate only a few items. The subset of users that rated an item $i \in I$ is denoted by $U(i)$, while the subset of items rated by a user $u \in U$ is denoted by $I(u)$.

For the items unrated by the users, recommender systems estimate a relevance score, denoted as $relevance(u, i)$, $u \in U$, $i \in I$. There are different ways to estimate the relevance score of an item for a user. In the content-based approach (e.g., [16]), the estimation of the rating of an item is based on the ratings that the user has assigned to similar items, whereas in collaborative filtering systems (e.g., [22]), this rating is predicted using previous ratings of the item by similar users. In this work, we follow the collaborative filtering approach. Similar users are located via a *similarity function* $simU(u, u')$ that evaluates the proximity between $u, u' \in U$ by considering their shared dimensions. We use P_u to denote the set of the most similar users to u , hereafter, referred to as the *peers* of u .

Definition 1 (Peers): Let U be a set of users. The peers P_u of a user $u \in U$ consists of all those users $u' \in U$ which are similar to u w.r.t. a similarity function $simU(u, u')$ and a threshold δ , i.e., $P_u = \{u' \in U : simU(u, u') \geq \delta\}$.

Given a user u and his peers P_u , if u has expressed no preference for an item i , the relevance of i for u is estimated as:

$$relevance(u, i) = \frac{\sum_{u' \in (P_u \cap U(i))} simU(u, u') rating(u', i)}{\sum_{u' \in (P_u \cap U(i))} simU(u, u')}. \quad (1)$$

After estimating the relevance scores of all unrated user items for a user u , the items A_u with the top- k relevance scores can be suggested to u .

B. Group Rating Model

Most previous works focus on recommending items to individual users. Recently, group recommendations that make recommendations to groups of users instead of single users (e.g., [21], [17]), have received considerable attention. Commonly, a method for computing group recommendations first estimates the relevance scores of the unrated items for each user in the group, and then, aggregates these predictions to compute the suggestions for the group. Formally, the relevance of an item for a group is computed as follows:

Definition 2: Let U be a set of users and I be a set of items. Given a group of users G , $G \subseteq U$, the group relevance of an item $i \in I$ for G , such that, $\forall u \in G, \nexists rating(u, i)$, is:

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

Similar to [17], we employ two different designs regarding the aggregation method *Aggr*, each one carrying different semantics. In both cases, the prediction for an item is computed taking into account the relevance of the item for the group members without considering the whole set of recommendations returned to the group. Firstly, we consider that strong user preferences act as a veto; this way, the predicted relevance of

²iph.ics.forth.gr

an item for the group is equal to the minimum relevance of the item scores of the members of the group:

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

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

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

As in single user recommendations, the items with the top- k relevance scores for the group are recommended to the group.

C. Fairness in Group Recommendations

In this work, we exploit the concept of group recommendations in order to provide valuable suggestions to a caregiver responsible for a group of patients. Our goal is to locate suggestions that include data items that are both highly related and fair to the patients of the group.

This way, given a particular set of recommendations for a caregiver, it is possible to have a user u that is the least satisfied user in the group for all items in the recommendations list, that is, all items are not related to u . Therefore, although the caregiver may like as a whole the set of recommendations, the package selection is not fair to u . In actual life, where the caregiver is concerned for the needs of all patients in his group, we should recommend items that are both strongly relevant and fair to the majority of the group members.

Motivated by this observation, we target at having insights into the properties of the produced recommendations in order to help making the algorithmic process transparent, non-discriminative and accountable [26]. Specifically, to increase the quality of the recommendations for the caregiver, we consider a fairness measure [19] that evaluates the goodness of the recommendations as a set. This way, given a user u and a set of recommendations D , we define that D is fair to u , if D contains at least one data item that belongs to the set of items with the top- k relevance scores for u . Intuitively, the fact that the group recommendations contain a highly relevant item to u , makes both u and his caregiver tolerant to the existence of other items that are not highly related to the user, considering that there are other members in the group who may be related to these items.

Definition 3 (Fairness): Given a group G and a set of recommendations D , the fairness of D for G is defined as:

$$\text{fairness}(G, D) = \frac{|G_D|}{|G|},$$

where G_D denotes the users for whom D is fair.

In overall, we define the fairness-aware value of D for G as follows:

$$\text{value}(G, D) = \text{fairness}(G, D) \cdot \sum_{i \in D} \text{relevance}G(G, i).$$

Algorithm 1 Fairness-aware Group Recommendations Algorithm

Input: A group of users $G = \{u_1, \dots, u_n\}$, and the sets of recommendations A_{u_x} for each user $u_x \in G$.

Output: The fairness-aware set of the z recommendations D for G .

```

1: begin
2:  $D = \emptyset$ ;
3: while  $|D| < z$  do
4:   for  $x = 0; x < n; x++$  do
5:     for  $y = 0; y < n; y++$  do
6:       if  $x \neq y$  then
7:         Find the item  $i \in A_{u_y}$  with the maximum
            $\text{relevance}(u_x, i)$ ;
8:          $D = D \cup i$ ;
9:       end if
10:    end for
11:  end for
12: end while
13: end

```

D. Problem Statement

Given a caregiver who is responsible for a group of patients G , and a restriction z on the number of the recommended data items, we would like to provide z suggestions for items that are highly relevant to the preferences and problems of all the group members, and also, exhibit high fairness.

Namely, our goal is to locate the set D^* of z data items for which:

$$D^* = \underset{|D|=z}{\operatorname{argmax}} \text{value}(G, D).$$

A brute-force method to locate the z most fair recommendations of D , for a group of patients G is to first produce all $\binom{m}{z}$, $m = |D|$, possible combinations of recommendations and then pick the one with the maximum $\text{value}(G, D)$. The complexity of this process is exponential and therefore, the computational cost is too high even for low values of m and z . A number of lower-complexity heuristics have been proposed to locate subsets of elements (e.g., in [6]). In this paper, we use the following variation: we construct a fairness-aware subset of recommendations based on the *value* of D for G .

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 . The above process is shown in Algorithm 1.

Using the *Fairness-aware Group Recommendations Algorithm*, when the number z of the returned recommendations D computed for a group G is greater than or equally to the group size $|G|$, the fairness of D for G is equal to 1.

Proposition 1: Let G be a group of users and D be the set of z recommendations computed by Algorithm 1 for G . If $z \geq |G|$, then $\text{fairness}(G, D) = 1$.

IV. IMPLEMENTATION IN MAPREDUCE

In this section, we elaborate on the adaptation of our recommender system in MapReduce. The MapReduce programming model consists of two consecutive procedures that

can be grouped into jobs. First, the Map phase receives a set of (key, value) pairs and transforms it into a new output set of pairs. Second, the Reduce phase receives a set of (key, value) pairs that share the same key and are sorted according to their value; it performs a summary operation on them to produce a new, usually smaller set of pairs.

The description of the implementation assumes that our input consists of a set of user rating triples $R = \{(u, i, rating(u, i)) | u \in U, i \in I\}$, and a set of user ids $G \subseteq U$, composing the group related to a caregiver. Specifically, we compute the list of peers for each member of the group (Definition 1), the relevance of every item to each member of the group (Equation 1) and, finally, aggregate those scores to get the final relevance of each item for the group (Definition 2). In order to identify the peers of each member, we compute the similarity between each member of the group and every other user, outside the group. The implementation of this approach consists of three MapReduce jobs.

- **Job 1 - Find partial users similarity score and the unrated items.** In this first MapReduce job, we have two different outputs. Given our input (as described above), we map it with i as the key and $(u, rating(u, i))$ as the value, where u is a user and $rating(u, i)$ is the rate given to the item by u . The reducer checks if any user in the group has rated that item. If not, then this item will be considered as a recommendation, and the output will be the same as the one given by the map phase. If one (or more) member(s) of the group G is (are) among the users who have rated the item i , then, for all different pairs of users between member and a non-member in the group, we calculate a partial score needed to eventually evaluate the $simU$ function. The output in this case has as key the user u_G , that belongs in the group and as value a pair that consists of his potential peer and the calculated score.
- **Job 2 - Calculate $simU$.** In this job, we finish the calculations of $simU(u_G, u)$. We take as input the second output from Job 1, with the partial scores. The mapper output has as key the pair $\langle u_G, u \rangle$, where u_G is the user in the group and u is a likely peer, and value the similarity score between u_G and u . The reducer then aggregate the values and checks if the total similarity score is above the threshold δ . The output of the reducer has the same key as the one from the mapper and value, the calculated $simU$, given that is above the threshold.
- **Job 3 - Calculate user and group relevance.** Having already calculated the similarities of the users from Job 2 and the list of potential recommendations from Job 1, with this third job, we find the relevance of the item i to the user u using Equation 1. In addition, given an aggregation method, defined either with minimum or the average scoring function, we calculate the relevance of the item i for the entire group. The input for this job, is the first output from Job 1, i.e., the list of items that no user in the group has rated. The map output has as key the item i and as value a pair consisting of a user and a rating. The reducer then calculates the two relevance scores and gives them both as output. The final sorting and top-k selection of those relevance values is trivial when k elements are small enough to fit in memory. When this is not the case, we can use the

top-k MapReduce algorithm suggested in [5].

The functionality of these three MapReduce jobs are outlined in Figure 2. After these jobs have completed, and the majority of the computations done, we perform Algorithm 1 in a centralized manner.

V. SIMILARITY FUNCTIONS

To produce recommendations for a group of users, we follow the collaborative filtering approach. A crucial step in this line of work is to find similar users to the ones that belong in the group in question. To find such similarities, we exploit data from the health domain. Specifically, we take into consideration the ratings that users have given to particular documents, and the users' profiles. The knowledge stored in user profiles vary. There are demographic information, such as age and gender, medical measurements or procedures that the user may have undertaken, and the health problems that they have faced in the past or still suffer from.

Next, we present three ways for measuring the similarity between two users. The first one makes use of the document ratings that a user has given in the system. The second measure compares the users profiles, and lastly, the third measure exploits semantic information achieved from an ontology related to health problems.

A. Similarity based on Ratings

A user can rate a document with a score in the range of 1 to 5. Our supposition here is that, if two users have rated documents in a similar way, then we can say that they are similar, since they share the same interests.

In order to calculate the similarity between two users, based on their ratings, we use the Pearson correlation (Equation 2), where $I(u)$ denotes the items that user u has assigned a rating, $rating(u, i)$ is the score that user u gave for an item i (for simplicity, $r(u, i)$), and μ_u is the mean of the ratings in $I(u)$, i.e., the mean of the ratings of u .

$$RS(u, u') = \frac{\sum_{i \in I(u) \cap I(u')} (r(u, i) - \mu_u)(r(u', i) - \mu_{u'})}{\sqrt{\sum_{i \in I(u) \cap I(u')} (r(u, i) - \mu_u)^2} \sqrt{\sum_{i \in I(u) \cap I(u')} (r(u', i) - \mu_{u'})^2}} \quad (2)$$

B. Similarity based on User Profiles

Towards exploiting user profiles, we consider all the information contained in a profile as a single document. This will enable us to compute the term frequency (tf) and inverse document frequency (idf) scores for the words contained in the user profiles. The tf score measures the occurrences of a given word within a document, and that score is then normalized by the idf score. The idf score is the log of the ratio of the total number of documents to the number of documents containing that word.

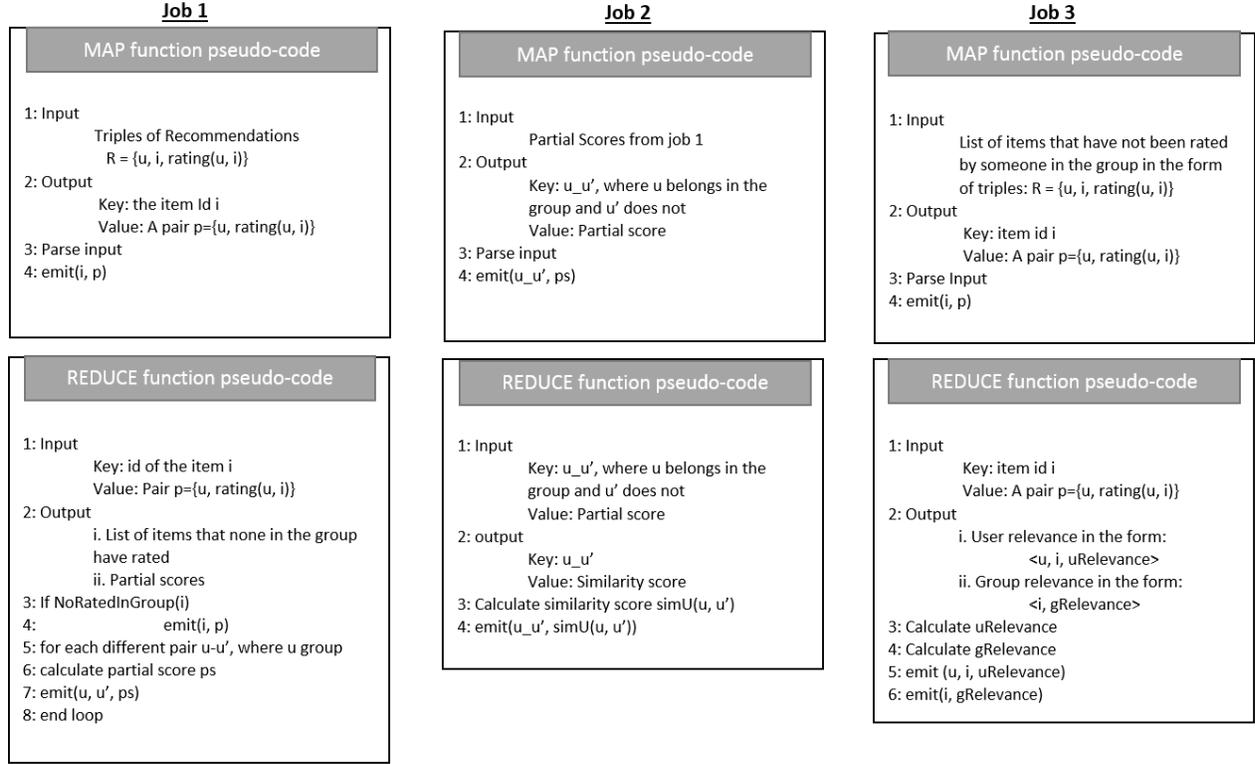


Fig. 2. The MapReduce pseudo-codes for the three jobs.

Definition 4 (IDF): Let N be the total number of documents in a set of documents D , and $|\{d \in D : t \in d\}|$ be the number of documents, where the term t appears. Then:

$$idf(t, D) = \log \frac{N}{|\{d \in D : t \in d\}|}.$$

By multiplying the tf and idf scores, we can determine how common a word is in our documents. Words that are present numerous times in a single document, but are rarely encountered in the rest, will achieve high tf-idf score. In addition, we can also filter out the common words. As a term appears in more documents, the ratio inside the logarithm approaches 1, bringing the idf and tf-idf closer to 0.

After we have computed the tf-idf scores, we have essentially converted our documents, i.e., our user profiles, into numbers. This way, each document can be represented as a vector. In the final stage, we want to find the similarity score between our transformed documents, which in turn reflect the similarity score between our users. This is easily achieved by calculating their cosine similarity.

$$CS(A, B) = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (3)$$

where the vectors A and B represent the profiles of two users u and u' .

C. Similarity based on Semantic Information

In the health domain, for two people to be considered similar, a major area where they need to have correlation to each other, are their health problems. In our case, we represent the health problems of users, by utilizing the SNOMED CT ontology³. This gives us the advantage, that we can make use of the hierarchy that is provided by the ontology. To compute the similarity between two users based on their diseases, we have two phases. First, we have to find the similarity between the problems of the two users. Secondly, we have to calculate the overall similarity of the users based on their health problems.

1) *Similarity between health problems:* Each health problem is represented by a node in the tree that is generated by the class hierarchy of the SNOMED CT ontology. To calculate the similarity between two health problems, we will identify the shortest path that connects those two nodes in the tree. Longer path means a smaller similarity.

For example, in Table I, we have 3 different patients. First, we will compare the health problems of the first two patients, namely Patient 1 and Patient 2. In the problem category, there are the values: 'acute bronchitis' and 'chest pain'. Based on the SNOMED-CT ontology tree, we can calculate that the shortest path between those two nodes is 5. In contrast, if we calculate the shortest path for Patient 1 and Patient 3, then the shortest path between his problem 'Tracheobronchitis' and 'Acute Bronchitis' is only 2. So the similarity based on the health problems between patients 1 and 3 is greater than the

³<https://bioportal.bioontology.org/ontologies/SNOMEDCT>

TABLE I. THE INFORMATION CONTAINED IN THE FIELDS OF 3 DIFFERENT PATIENTS PROFILE

Patient	Field	Value
Patient 1	Problem	Acute bronchitis
	medication	Ramipril 10 MG Oral Capsule
	gender	Female
	procedure	
	age	40
Patient 2	Problem	Chest pains
	medication	Niacin 500 MG Extended Release Tablet
	gender	Male
	procedure	
	age	53
Patient 3	Problem	Tracheobronchitis Broken arm
	medication	Ramipril 10 MG Oral Capsule
	gender	Male
	procedure	
	age	34

one between patients 1 and 2.

2) *Overall similarity based on health problems:* Using the approach describe above, we can calculate the similarity between two health problems. But a user will not have only one disease in his or her profile. So we need a procedure to calculate the overall similarity between two users based on their health problems. One method to deal with this problem is, to find the similarities of all the pairs of problems that are included in the profile of the two users. After that, we can calculate their harmonic mean (Equation 4). That score will be the overall similarity score of the two users, based on their health problems.

$$SS(u, u') = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}} \quad (4)$$

In the equation above, we denote with x_i the various similarities calculated by comparing the health problems of the users u and u' , and with n the total number of pairs of problems that are included in the profiles of u and u' .

VI. PRELIMINARY EVALUATION

In our preliminary evaluation, we experiment with the performance of our underlying fairness-aware heuristic by comparing it against the brute-force algorithm in terms of the time. The complexity of both methods depends on the number m of candidate recommendations to choose from and on the required number z of recommendations to select.

We experiment with a number of different values for m and z . However, the exponential complexity of the brute-force algorithm prevents us from using large values for these two parameters. Therefore, we limit our study to $m = 10, 20, 30$ and $z = 4, 8, 12, 16, 20$. In Table II, we show the results for the brute-force method and our heuristic. We observe that the brute-force method consumes much more time than the heuristic. Note that the fairness of the produced results are identical in both cases verifying Proposition 1.

VII. RELATED WORK

Traditionally, recommendation approaches are distinguished between content-based and collaborative filtering.

TABLE II. BRUTE-FORCE VS. HEURISTIC FAIRNESS.

m	z	Brute-force - Time (msec)	Heuristic - Time (msec)
10	4	37	10
	8	41	13
20	4	712	19
	8	72254	23
	12	171414	34
	16	13340	46
30	4	3981	23
	8	3425266	33
	12	116735821	45
	16	322371457	65
	20	124219934	83

Content-based approaches recommend items similar to those the user previously preferred (e.g., [16]), while collaborative filtering approaches recommend items that users with similar preferences liked (e.g., [12]). Several extensions have been proposed, such as time-aware recommendations (e.g., [28], [25]) and group recommendations (e.g., [1], [17]). Lately, there are also approaches on extending database queries with recommendations [13], [24].

To facilitate the selection of similar users to a query user, clustering has been employed to pre-partition users into clusters of similar users and rely on cluster members for recommendations. For example, [17] employ full-dimensional clustering; as explained though, full dimensional clustering is not the best option due to the high dimensionality and sparsity of data. Dimensionality reduction techniques, like PCA, could be applied to reduce dimensionality, however clusters existing in subspaces rather than in the original (or reduced) feature space will be missed. More recently, several works (e.g., [14], [18]) augment users through subspace clustering algorithms to improve the quality of recommendations.

In addition, there are already several approaches trying to provide to the patients and healthcare providers search engines with high quality medical information such as WebMD⁴, MayoClinic Patient Care⁵ and Medicine Plus⁶. One of the most well-known search engine is HONSearch⁷, which aims to improve the quality of the information intended to both patients and medical professionals by facilitating quick access to the most relevant and up-to-date medical discoveries. Each indexed web document has to fulfill specific criteria in order to be included in the search engine and uses a multi-stakeholder approach to include the relevant web documents. Although the approach seems to be promising, there has been also some concerns on the quality [8] on the indexed web sites. To this direction, but targeting medical experts, AskHermes [3] is an online question answering system, trying to answer specialized clinical questions. However, these engines provide a rather limited set of information and they are neither dynamically adapted according to individual preferences nor consider concepts like group recommendations and fairness.

On the other hand, there are already several modules that exploit the profiles stored in PHR systems to automatically present useful information to the patients. For example, the

⁴<http://www.webmd.com/>

⁵<http://www.mayoclinic.org/patient-care-and-health-information>

⁶<http://www.nlm.nih.gov/medlineplus/>

⁷<http://www.hon.ch/HONsearch>

ADE (adverse drug effects) [20] is an alerting system to inform patients about potential risks and the adverse effects of the medications they receive. Interesting approaches to our direction are STEPPS [4] and the discontinued MyDailyApple [20]. STEPPS tries to personalize the automatic retrieval of health information using profiling information from an electronic patient record, whereas MyDailyApple provided simple personalized recommendations based on patients Google Health Profiles. However, on both cases, the lack of semantics leads to poor results, and the lack of a search engine, to allow patients interaction, limits the patients options.

To our knowledge, PMIR [10], [9], [11] is the only system exploiting patients profiles to provide both automatic and non-automatic high quality information to patients employing semantics, reasoning and exploiting also user preferences. PMIR focuses on delivering relevant information to patients, based on their current profile as represented in their personal healthcare record (PHR) data. To ensure the quality of the presented information the documents indexed are carefully selected by medical experts, focusing specific on the cancer domain, trying to control the provided information.

However, in all the aforementioned systems the notions of group recommendations and fairness are completely missing. To the best of our knowledge, our approach is the only one in the area of exploring those two notions in a medical setting.

VIII. DISCUSSION AND CONCLUSIONS

The changing nature of information distribution due to the evolution of the Web has important implications for health care. Given the wide use of the Web in providing medical information, feeding patients with appropriate content might further enhance the patients education and experience. The validity and the quality of the available healthcare information on the internet is an area of major concern mainly because these have not been well documented [7]. Although healthcare professionals should continue to strive to be the main source of information for patients, we should also be aware that most will continue to use the internet to gather information [23]. A recent publication [2] concluded that the optimal solution for patients is to be guided by healthcare providers to more optimal resources over the Web. Delivering accurate sources to patient increases his knowledge and changes the way of thinking which is usually referred as patient empowerment. As a result, the patients dependency for information from the doctor is reduced. Moreover, patients feel autonomous and more confident about the management of their disease [27].

Our platform focuses on making the available information timelier and more relevant with respect to dynamic influences in a set of patients. The idea is that even if all patients suffer from the same disease and they are in the same phase of the treatment, their interests on available information may differ based on various factors. For example, patients might have different medical background that is not directly related to the treatment, they might receive additional drugs due to other, independent treatments, or they might be affected by other external factors such as the weather in case of allergies. More subjective factors include, for example, patient preferences for more simple or more complex information. As such, the notion of fairness in recommendations is crucial. To the best of our

knowledge no other system, providing medical information, is able to be dynamically adapted in such a diverse environment.

For future work, we consider to include machine learning approaches for recommending to patients automatically useful information and to include a reasoning engine to identify correspondences in patient profiles. In addition, we would like to be able to present a semantically enhanced summary of the indexed document to the patient to augment his understanding. We expect that the evaluation of the whole platform will lead us to useful observations that might require the modification of our initial algorithms. Then, a second more thorough evaluation will follow. It becomes obvious that delivering accurate, personalized and high quality information to patients is an important topic and several challenging issues remain to be investigated in near future.

IX. ACKNOWLEDGEMENTS

This work was partially supported by the EU project iManageCancer (H2020, #643529).

REFERENCES

- [1] S. Amer-Yahia, S. B. Roy, A. Chawla, G. Das, and C. Yu. Group recommendation: Semantics and efficiency. *PVLDB*, 2(1):754–765, 2009.
- [2] G. M. Berg, A. M. Hervey, D. Atterbury, R. Cook, M. Mosley, R. Grundmeyer, and D. Acuna. Evaluating the quality of online information about concussions. *Journal of the American Academy of Physician Assistants*, 27(2):1–8, 2014.
- [3] Y. Cao, F. Liu, P. Simpson, L. Antieau, A. Bennett, J. J. Cimino, J. Ely, and H. Yu. Askhermes: An online question answering system for complex clinical questions. *Journal of Biomedical Informatics*, 44(2):277 – 288, 2011.
- [4] P. Doupi, M. A. van Wijk, J. T. van Wyk, and J. van der Lei. Personalizing patient education using internet health resources and EPR data: Pilot evaluation of the STEPPS prototype system. In *IEEE International Symposium on Computer-Based Medical Systems*, 2008.
- [5] V. Efthymiou, K. Stefanidis, and E. Ntoutsis. Top-k computations in mapreduce: A case study on recommendations. In *IEEE International Conference on Big Data*, 2015.
- [6] E. Erkut, Y. Ülküsal, and O. Yeniçerioglu. A comparison of p -dispersion heuristics. *Computers & OR*, 21(10):1103–1113, 1994.
- [7] A. Gagliardi and A. R. Jadad. Examination of instruments used to rate quality of health information on the internet: chronicle of a voyage with an unclear destination. *BMJ*, 324(7337):569–573, 2002.
- [8] Y. Khazaal, A. Chatton, D. Zullino, and R. Khan. Hon label and discern as content quality indicators of health-related websites. *Psychiatric Quarterly*, 83(1):15–27, 2012.
- [9] H. Kondylakis, L. Koumakis, E. Kazantzaki, M. Chatzimina, M. Psaraki, K. Marias, and M. Tsiknakis. Patient empowerment through personal medical recommendations. In *MEDINFO 2015: eHealth-enabled Health - Proceedings of the 15th World Congress on Health and Biomedical Informatics, São Paulo, Brazil, 19-23 August 2015*, page 1117, 2015.
- [10] H. Kondylakis, L. Koumakis, M. Psaraki, G. Troullinou, M. Chatzimina, E. Kazantzaki, K. Marias, and M. Tsiknakis. Semantically-enabled personal medical information recommender. In *ISWC*, 2015.
- [11] H. Kondylakis, L. Koumakis, M. Tsiknakis, K. Marias, E. Genitsaridi, G. Pravettoni, A. Gorini, and K. Mazzocco. *Smart*

Recommendation Services in Support of Patient Empowerment and Personalized Medicine, pages 39–61. Springer International Publishing, Heidelberg, 2013.

- [12] J. A. Konstan, B. N. Miller, D. Maltz, J. L. Herlocker, L. R. Gordon, and J. Riedl. GroupLens: Applying collaborative filtering to usenet news. *Commun. ACM*, 40(3):77–87, 1997.
- [13] G. Koutrika, B. Bercovitz, and H. Garcia-Molina. Flexrecs: expressing and combining flexible recommendations. In *SIGMOD*, 2009.
- [14] X. Li and T. Murata. Using multidimensional clustering based collaborative filtering approach improving recommendation diversity. In *Web Intelligence/IAT Workshops*, 2012.
- [15] M. McMullan. Patients using the internet to obtain health information: How this affects the patienthealth professional relationship. *Patient Education and Counseling*, 63(12):24 – 28, 2006.
- [16] R. J. Mooney and L. Roy. Content-based book recommending using learning for text categorization. In *DL*, 2000.
- [17] E. Ntoutsi, K. Stefanidis, K. Nørnvåg, and H. Kriegel. Fast group recommendations by applying user clustering. In *ER*, 2012.
- [18] E. Ntoutsi, K. Stefanidis, K. Rausch, and H. Kriegel. Strength lies in differences: Diversifying friends for recommendations through subspace clustering. In *CIKM*, 2014.
- [19] S. Qi, N. Mamoulis, E. Pitoura, and P. Tsaparas. Recommending packages to groups. In *ICDM*, 2016.
- [20] H. Roitman, Y. Messika, Y. Tsimerman, and Y. Maman. Increasing patient safety using explanation-driven personalized content recommendation. In *Proceedings of the 1st ACM International Health Informatics Symposium, IHI '10*, pages 430–434, New York, NY, USA, 2010. ACM.
- [21] S. B. Roy, S. Amer-Yahia, A. Chawla, G. Das, and C. Yu. Space efficiency in group recommendation. *VLDB J.*, 19(6):877–900, 2010.
- [22] J. J. Sandvig, B. Mobasher, and R. D. Burke. A survey of collaborative recommendation and the robustness of model-based algorithms. *IEEE Data Eng. Bull.*, 31(2):3–13, 2008.
- [23] P. Scullard, C. Peacock, and P. Davies. Googling children’s health: reliability of medical advice on the internet. *Archives of Disease in Childhood*, 95(8):580–582, 2010.
- [24] K. Stefanidis, M. Drosou, and E. Pitoura. *You May Also Like* results in relational databases. In *PersDB*, 2009.
- [25] K. Stefanidis, E. Ntoutsi, M. Petropoulos, K. Nørnvåg, and H.-P. Kriegel. A framework for modeling, computing and presenting time-aware recommendations. *T. Large-Scale Data- and Knowledge-Centered Systems*, 10:146–172, 2013.
- [26] J. Stoyanovich, S. Abiteboul, and G. Miklau. Data responsibly: Fairness, neutrality and transparency in data analysis. In *EDBT*, 2016.
- [27] M. Wiesner and D. Pfeifer. Adapting recommender systems to the requirements of personal health record systems. In *Proceedings of the 1st ACM International Health Informatics Symposium, IHI '10*, pages 410–414, New York, NY, USA, 2010. ACM.
- [28] L. Xiang, Q. Yuan, S. Zhao, L. Chen, X. Zhang, Q. Yang, J. Sun, and J. Sun. Temporal recommendation on graphs via long- and short-term preference fusion. In *KDD*, 2010.