A Recommender System for Connecting Patients to the Right Doctors in the HealthNet Social Network

Fedelucio Narducci[†‡], Cataldo Musto[†‡], Marco Polignano[†],
Marco de Gemmis[†], Pasquale Lops[†], Giovanni Semeraro[†]
[†]Dept. of Computer Science, University of Bari Aldo Moro, Italy
[‡]Murex CS srl, Bari, Italy
name.surname@uniba.it

ABSTRACT

In this work we present a semantic recommender system able to suggest doctors and hospitals that best fit a specific patient profile. The recommender system is the core component of the social network named HealthNet (HN). The recommendation algorithm first computes similarities among patients, and then generates a ranked list of doctors and hospitals suitable for a given patient profile, by exploiting health data shared by the community. Accordingly, the HN user can find her most similar patients, look how they cured their diseases, and receive suggestions for solving her problem. Currently, the alpha version of HN is available only for Italian users, but in the next future we want to extend the platform to other languages. We organized three focus groups with patients, practitioners, and health organizations in order to obtain comments and suggestions. All of them proved to be very enthusiastic by using the HN platform[1].

Keywords

e-health, social network, smart health, recommender system

1. INTRODUCTION E MOTIVATION

The main desire when experiencing a health problem is to find a doctor or hospital with the best expertise for solving our health condition. In the U.S., a nonprofit group named Consumers' Checkbook (CB) won a lawsuit allowing it to have access to Medicare's doctors records, but the government appealed the decision[2]. The goal of the group was to have access to a database for analyzing how often a doctor performs a procedure (e.g., knee replacements, prostate surgery) in order to define a first quality indicator associated with proficiency. Indeed, choosing a doctor can be viewed as a typical problem of information asymmetry, because of the information available to the patient is too weak to make an informed choice. In the last years, the new healthcare practice supported by electronic process and communication (i.e., *e-health*) [1], is changing the landscape of clinical practice and health care. Indeed, 72% of U.S. Internet users looked online for health information within the past years[3]. Similarly, in Italy, 84% of young people aged between 18 and 35 use the Web for searching health information[4]. Sharing information and knowledge can empower patients and lead the patient/care team relationship towards a patient-centered medicine. One of the most relevant initiatives in that direction is the U.S. social network PatientsLikeMe (PLM)[5]. This social network enables patients to share, compare and contrast different diagnoses and treatments with people in the same conditions anywhere in the world. PLM counts 300,000 patients sharing 2,300 different conditions. In addition to PLM, many forums, blogs, and more generally web sites deal with health problems, but the information is often confused, difficult to understand, and can lead easily to wrong self-diagnosis [2].

In this paper, we present a recommender system that helps users in finding solutions for their health conditions.. This recommender system is embedded in the HealthNet (HN) social network, whose main goal is similar to PLM: sharing knowledge, finding similar patients, looking at their experiences. The main difference between HN and PLM is that HN embeds a recommender system that is able not only to discover similarity between patients, but also to provide suggestions about practitioners and hospitals that best fit the patient profile, based on the data coming from the community. In this way, HN prevents self-diagnosis but at the same time helps patients to find a solution represented by a doctor or a health facility.

Other Health Related Recommender Systems (HRS) are presented in the literature [3]. To the best of our knowledge, the HRS implemented in HN is the first one able to suggest doctors and hospitals by performing an advanced, semantic matching between patient profiles.

In the following Section 2, we provide a descriptions of the platform and describe the recommender system, Section 3 shows the results of a preliminary experimental evaluation, and finally Section 4 draws conclusions and future work.

2. THE HN RECOMMENDER SYSTEM

HN is implemented as a standard social network where users are patients. The first interaction with the system is

[1]An English demo of HN is available at http://193.204.187.192:8080/HealthNetVideo/

[2]http://www.nytimes.com/2008/09/30/health/30find.html?_r=0

[3]http://www.pewinternet.org/2013/01/15/health-online-2013/

[4]http://it.ejo.ch/tag/eikon-strategic-consulting

[5]http://www.patientslikeme.com

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2742748 .

the registration step. Then, the patient can enter personal health data: conditions, treatments (e.g., drugs, dosages, side effects, surgeries), health indicators (e.g., blood pressure, body weight, laboratory analysis, etc.), consulted doctors, hospitalizations. In this way, HN centralizes individual health data and allows a simple and organized access to them. Furthermore, users can take advantage from sharing their data by obtaining suggestions about doctors and health facilities. In order to receive recommendations, the user should enter at least one condition she is affected by. For each condition, a simple click on the "How can I cure it?" button allows the patient to receive suggestions. It is worth noting that the HN user can decide to be anonymous, by indicating only a nickname during the registration step. Accordingly, health data entered in HN will not be linkable to a real identity, thus preserving user privacy.

The *Recommender System* is the core component of HN. It exploits patient profiles for suggesting other similar patients, doctors, hospitals (the list of suggested, patients, doctors and hospitals can be further filtered by position and disease). The similarity between two patients p, p' is computed in terms of conditions and treatments. The semantic matching between the conditions exploits the HN disease hierarchy[6]. More formally, the similarity score between two patients is computed as follows:

$$s(p,p') = \alpha \frac{\sum_{i=1}^{k}\sum_{j=1}^{n} s_c(p_{c_i}, p'_{c_j})}{k*n} + \\ +(1-\alpha)\frac{\sum_{i=1}^{z}\sum_{j=1}^{r} s_t(p_{t_i}, p'_{t_j})}{z*r}, \quad (1)$$

where k (respectively n) is the number of conditions p (respectively p') is affected by, p_c is a condition of the patient p, z (respectively r) is the number of treatments for p (respectively p'), p_t is a treatment for the patient p. $s_c(p_{c_i}, p'_{c_j})$ is the condition similarity between c_i, and c_j, while $s_t(p_{t_i}, p'_{t_j})$ is the treatment similarity between t_i, and t_j. They are computed as follows:

$$s_c(p_{c_i}, p'_{c_j}) = \begin{cases} log\frac{P}{P_{c_i}}, & \text{if } c_i = c_j \\ \frac{1}{sp(c_i,c_j)}, & \text{otherwise} \end{cases} \quad (2)$$

$$s_t(p_{t_i}, p'_{t_j}) = \begin{cases} 1, & \text{if } t_i = t_j \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

The similarity score s_c for patients experiencing the same condition c_i is computed as the logarithm of the ratio between the total number of patients in the database (P) and the number of patients affected by that condition (P_{c_i}). The aim is to give higher similarity to patients that share rare diseases. For different conditions, the score is simply the reciprocal of the length (number of edges) of the shortest path *sp* that connects the two conditions in the disease hierarchy. The treatment similarity is a simple binary score that is equal to 1 when the treatments are the same (or they are based on the same active ingredient). Treatment similarity and condition similarity scores can differently contribute to the patient similarity score, by varying the α value. The patient similarity is used for computing a ranked list of suggested doctors and hospitals. Given the *target* patient p_i (for whom suggestions must be provided), doctors and hospitals are ranked according to the *scoreDoc* and *scoreH*. The *scoreDoc* for the doctor d_z and patient p_i is computed by taking into account the rating r_j assigned to d_z by the other patients in the database, weighted by the similarity score between each one of them and p_i:

$$scoreDoc(d_z, p_i) = \sum_{j=1}^{P} s(p_i, p_j) \cdot r_j(d_z). \quad (4)$$

The *scoreH* takes into account patient similarity, the rating r_j assigned by patient p_j to a given hospital h_m, and also a quality indicator provided by the Italian Health Ministry for every Italian hospital[7]. The community indicator and the ministry indicator can be weighted differently by changing the β value:

$$scoreH(h_m, p_i) = \beta \left(\sum_{j=1}^{P} s(p_i, p_j) * r_j(h_m) \right) + (1-\beta) \cdot qi(h_m) \quad (5)$$

3. EXPERIMENTAL EVALUATION

A preliminary evaluation has been carried out to compare our semantic approach based on the desease hierarchy to a simple string matching baseline (SM), at different α values. Supported by three practitioners, 12 patient stereotypes (gold standard) have been defined. For all pairs of patients, the three domain experts heuristically provided a similarity score in the range $[0, 1]$, which is compared by that given by our recommender system, and the Mean Absolute Error is computed. Results in Table 1 demonstrated that

Table 1: MAE with different alpha values

α	0.5	1	0	0.7	0.3	SM
MAE	.044	.052	.127	.043	.053	.190

the best configuration is obtained by $\alpha = 0.7$. All configurations have a statistically significant difference compared to the SM (Paired T-test p-value <0.01), confirming the usefulness of the semantic matching (e.g., with $\alpha = 0.7$, the similarity score between a patient with *prostate cancer* and another with *testicular cancer* was 0.82, while SM was 0.31).

4. CONCLUSION AND FUTURE WORK

We described a recommender system for suggesting doctors and hospitals to patients, based on their semantic similarity, computed by exploiting a desease hierarchy. We are currently working to design an extensive evaluation involving real patients, while in the future we want to extend the system to other languages, to include other similarity measures and to allow user to export and share (e.g, with her practitioners) all her health data.

Acknowledgments. This work is partially funded by the "HEALTHNET: The Health Network - Il social network della salute" project, P.O. Puglia 2007-2013 Asse I - L. di Int. 1.4 - Az. 1.4.2 - Living Labs Smart Puglia 2020.

5. REFERENCES

[1] V. Della Mea. What is e-health (2): The death of telemedicine? *J Med Internet Res*, 3(2):e22, June 2001.

[2] B. W. Hesse, D. E. Nelson, G. L. Kreps, R. T. Croyle, N. K. Arora, B. K. Rimer, and K. Viswanath. Trust and sources of health information. *Archives of Internal Medicine*, 165(22):2618–2624, 2005.

[3] M. Wiesner and D. Pfeifer. Health recommender systems: Concepts, requirements, technical basics and challenges. *Int. J. of Environmental Research and Public Health*, 11(3):2580–2607, 2014.

[6]The HN disease hierarchy counts 12, 286 diseases on 7 levels

[7]http://95.110.213.190/PNEed13/