Artificial Intelligence in Medicine xxx (2017) xxx–xxx

Contents lists available at ScienceDirect

Artificial Intelligence in Medicine

journal homepage: www.elsevier.com/locate/aiim

User recommendation in healthcare social media by assessing user similarity in heterogeneous network

Ling Jiang, Christopher C. Yang*

College of Computing and Informatics, Drexel University, 3141 Chestnut Street, Philadelphia, PA, 19104, United States

ARTICLE INFO

Article history:
Received 2 March 2017
Accepted 3 March 2017

Keywords:
Heterogeneous network mining
Similarity analysis
Healthcare informatics
Social media analytics
Recommendation systems

ABSTRACT

Objective: The rapid growth of online health social websites has captured a vast amount of healthcare information and made the information easy to access for health consumers. E-patients often use these social websites for informational and emotional support. However, health consumers could be easily overwhelmed by the overloaded information. Healthcare information searching can be very difficult for consumers, not to mention most of them are not skilled information searcher. In this work, we investigate the approaches for measuring user similarity in online health social websites. By recommending similar users to consumers, we can help them to seek informational and emotional support in a more efficient way.
Methods: We propose to represent the healthcare social media data as a heterogeneous healthcare information network and introduce the local and global structural approaches for measuring user similarity in a heterogeneous network. We compare the proposed structural approaches with the content-based approach.
Results: Experiments were conducted on a dataset collected from a popular online health social website, and the results showed that content-based approach performed better for inactive users, while structural approaches performed better for active users. Moreover, global structural approach outperformed local structural approach for all user groups. In addition, we conducted experiments on local and global structural approaches using different weight schemas for the edges in the network. Leverage performed the best for both local and global approaches. Finally, we integrated different approaches and demonstrated that hybrid method yielded better performance than the individual approach.
Conclusion: The results indicate that content-based methods can effectively capture the similarity of inactive users who usually have focused interests, while structural methods can achieve better performance when rich structural information is available. Local structural approach only considers direct connections between nodes in the network, while global structural approach takes the indirect connections into account. Therefore, the global similarity approach can deal with sparse networks and capture the implicit similarity between two users. Different approaches may capture different aspects of the similarity relationship between two users. When we combine different methods together, we could achieve a better performance than using each individual method.

1. Introduction

Online health social websites have grown substantially over the past decades. Health consumers participated in these websites to share healthcare information with peers. Through social networking with other consumers, they also provide social support for each other. Social support has been found to be important in helping patients to cope with stressful health conditions. It was demonstrated that social support could contribute to health outcomes by enhancing patient adherence to medical treatment [1]. Informational and emotional support are the most prevalent types of social support found in online health social websites [2].

Informational support help consumers to reduce uncertainty by offering facts or knowledge, including advice, information referral, insight from personal experiences, or opinions [2,3]. It was found that a majority of online posts in healthcare social websites were intended to seek informational support, such as asking

* Corresponding author.
E-mail addresses: ling.jiang@drexel.edu (L. Jiang), chris.yang@drexel.edu (C.C. Yang).

http://dx.doi.org/10.1016/j.artmed.2017.03.002
0933-3657/

for advice on drugs or treatments [2]. The information was considered as being particularly helpful to consumers newly joined the online communities [4]. They can learn highly technical and factual information about various medical conditions. The online social support was fulfilling a learning as well as support function, especially for consumers concerned with chronic or mental illness [4,5]. The rapid development of online healthcare social websites has supported consumers' healthcare information searching by significantly reducing barriers to healthcare information. Consumers can see what their peers are experiencing, and discuss healthcare topics with people who have similar concerns. However, this vast amount of healthcare information comes with challenges. Consumers could easily be overwhelmed by the ocean of information and realize that it is difficult to find the most relevant piece. In an online health social website with hundreds of thousands of threads on all kinds of health-related topics, searching can be very difficult for consumers, not to mention most of them are not skilled information searcher. Consumers use very different language from professional terms to express the healthcare issues [6], and this language problem directly leads to poor query formation [7]. In addition, consumers usually cannot fully understand their health conditions due to the lack of medical knowledge, which would further hinder effective information searching [7]. Under such circumstance, asking for informational support from similar consumers could be much more efficient in seeking health information than relying on browsing or using the website search function. Consequently, a good recommendation system that can match users to users for their information needs is desired.

Emotional support is another important social support consumers could benefit from an online health social website. Many patients describe their situations as "understandable only if you have gone through a similar situation." This understanding is part of empathy, which naturally stems from going through a similar situation [8]. Therefore, emotional support could be especially effective when it comes from similar consumers. And prior studies have shown that emotional support could help patients to maintain positive attitude towards health issues and to achieve better health outcomes [9,10]. So, social networking with similar consumers not only provides a shortcut to relevant healthcare information, but also helps consumers to find emotional resonance.

However, online health social websites are different from traditional social media sites. In traditional social media sites such as Facebook and Twitter, people keep in contact with a list of friends, and the networks are often built based on one's existing social ties [9]. People can follow their friends and keep track of news in their friends' daily lives. Social networking activities in such social media sites are based on the relationship between the participants but not necessarily based on the desire of in-depth discussion on healthcare specific issues. Nevertheless, in an online health social websites, social networking activities are mainly based on common concerns rather than long-term relationships. Consumers would participate in the same discussion thread or send messages to others because of similar interests, but they seldom maintain a long-term relationship with each other. The social ties are much weaker in online healthcare social websites [11]. In this case, consumers may not be able to quickly find the most similar users with themselves when they need informational or emotional support. Therefore, recommendation of similar users to consumers in online health social websites is desirable.

In this work, we propose to consider an online healthcare social website as a heterogeneous healthcare information network, and develop the content-based approach and structural approach for recommending similar users in such a network. In this paper, we used data collected from a popular online health social website MedHelp.[1] Millions of Consumers post discussion threads on a variety of healthcare topics in MedHelp every day [12], and we can collect the online threads for our study. In most existing studies on network science, information networks are usually assumed to be homogeneous, where nodes are objects of the same entity type and links are relationships from the same relation type. However, most real-world networks are heterogeneous, where nodes and relations are of different types [13]. Healthcare information network is one typical heterogeneous network. Different types of relations convey different semantic meanings, and treating all the nodes or links as of the same type may miss important semantic information [13].

The rest of the paper is organized as follows. In Section 2, we review related work on network construction, including named entity recognition and relation extraction, and then we discuss existing studies on similarity analysis. In Section 3, we introduce methods for constructing a heterogeneous healthcare information network from an online healthcare social website. In Section 4, we propose three different methods based on content-based approach and structural approach for recommending similar users in online health social websites. Section 5 presents the details on experiment, including data collection, ground truth, results and discussion. Finally, we conclude the work in Section 6.

2. Literature review

2.1. Constructing heterogeneous healthcare information network

Given a set of healthcare social media data, the question is how to construct an information network to best represent the data. In order to do that, we need to identify entities and relationship between entities from the data. In this section, we review related work on named entity recognition and relation extraction from healthcare social media data.

Many studies have been reported on named entity recognition from healthcare social media data in the past few years. Some researchers used lexicon to identify entities such as Adverse Drug Reactions (ADRs) and drugs. In [14], Leaman et al. created a lexicon by combining terms and concepts from four resources and used it to identify adverse reactions in user comments collected from DailyStrength. Patki et al. [15] expanded the lexicon used by Leaman et al. [14] by adding addition terms from two more sources: SIDER II [16] and Consumer Health Vocabulary (CHV) [17]. Benton et al. [18] scraped websites and databases that contain lists of dietary supplements, pharmaceuticals, and adverse events for terms to create a controlled vocabulary, and then augmented the vocabulary using the CHV.

Besides lexicon, language patterns were also utilized to automatically extract mentions of ADRs from user reviews about drugs in social network websites [19]. Pimpalkhute et al. proposed an approach to generate most probable misspelled drug name variants for querying social media postings. The underlying intuition is that, faced with the task of writing an unfamiliar, complex word (the drug name), users will tend to revert to phonetic spelling [20].

Some other researchers used machine learning techniques to extract ADRs from social media data. Using Twitter as data source, Ginn et al. [21] created a manually annotated corpus of 10,882 tweets that were used to train automated tools to extract ADRs from tweets. Nikfarjam et al. [22] extracted user posts from DailyStrength and Twitter, and introduced ADRMine, a machine learning sequence tagger for concept extraction from social media. ADRMine utilizes a variety of features such as context features, ADR lexicon, part-of-speech tags, negations, and a novel feature

[1] http://www.medhelp.org/.

Please cite this article in press as: Jiang L, Yang CC. User recommendation in healthcare social media by assessing user similarity in heterogeneous network. Artif Intell Med (2017), http://dx.doi.org/10.1016/j.artmed.2017.03.002