Journal of Biomedical Informatics 46 (2013) 998–1005

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

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

Text classification for assisting moderators in online health communities

Jina Huh [a,*], Meliha Yetisgen-Yildiz [b], Wanda Pratt [b,c]

[a] Department of Telecommunication, Information Studies, and Media, Michigan State University, 404 Wilson Rd, Rm 409, East Lansing, MI 48864, USA
[b] Department of Biomedical Informatics and Medical Education, School of Medicine, University of Washington, 1959 NE Pacific Street, Seattle, WA 98195, USA
[c] The Information School, University of Washington, Box 352840, Mary Gates Hall, Ste 370, Seattle, WA 98195-2840, USA

ARTICLE INFO

Article history:
Received 17 October 2012
Accepted 31 August 2013
Available online 8 September 2013

Keywords:
Online health communities
Consumer health
Human–computer interaction
Text mining
Health information seeking

ABSTRACT

Objectives: Patients increasingly visit online health communities to get help on managing health. The large scale of these online communities makes it impossible for the moderators to engage in all conversations; yet, some conversations need their expertise. Our work explores low-cost text classification methods to this new domain of determining whether a thread in an online health forum needs moderators' help.

Methods: We employed a binary classifier on WebMD's online diabetes community data. To train the classifier, we considered three feature types: (1) word unigram, (2) sentiment analysis features, and (3) thread length. We applied feature selection methods based on χ^2 statistics and under sampling to account for unbalanced data. We then performed a qualitative error analysis to investigate the appropriateness of the gold standard.

Results: Using sentiment analysis features, feature selection methods, and balanced training data increased the AUC value up to 0.75 and the F1-score up to 0.54 compared to the baseline of using word unigrams with no feature selection methods on unbalanced data (0.65 AUC and 0.40 F1-score). The error analysis uncovered additional reasons for why moderators respond to patients' posts.

Discussion: We showed how feature selection methods and balanced training data can improve the overall classification performance. We present implications of weighing precision versus recall for assisting moderators of online health communities. Our error analysis uncovered social, legal, and ethical issues around addressing community members' needs. We also note challenges in producing a gold standard, and discuss potential solutions for addressing these challenges.

Conclusion: Social media environments provide popular venues in which patients gain health-related information. Our work contributes to understanding scalable solutions for providing moderators' expertise in these large-scale, social media environments.

1. Introduction

In clinic-sponsored, face-to-face support groups, moderators bring in clinical expertise when appropriate, such as when patients need clarification of clinical concepts or medical consultation. In these settings, patients not only benefit from peer-patients' knowledge but also from the clinical knowledge of the moderators.

Although online health communities facilitate large-scale information exchange among peer-patients, few communities employ moderators [1]. For those few communities who do provide moderators, time and resource limitations make it challenging for the moderators to answer the hundreds of posts uploaded each day.

Not all posts in online community forums require moderators' response. Some posts discuss everyday strategies, such as recipes for healthier diet, which could be supported purely by patients' responses. However, other posts asking for medical consultation or clarification of medical concepts need moderators' help. Moderators redirect patients to their health care providers, point to related health professionals' blog articles, or provide medical information. Given the large amount of information exchanged online, the challenge lies in helping moderators efficiently select the posts that need their expertise.

In this paper, we present our exploration on low-cost text classification methods—namely the approach not requiring manual annotation for the training data and using readily available features—for prioritizing posts that need moderators' help. Our method can help moderators of online health communities to more efficiently prioritize which patients' posts they need to respond to, which will also help fulfill patients' information needs. We built a binary classifier and explored various feature types, feature selection methods, and under sampling of the training data for performance optimization. We then performed an error analysis on a

* Corresponding author.
E-mail addresses: jinahuh@msu.edu, jina.huh@gmail.com (J. Huh), melihay@uw.edu (M. Yetisgen-Yildiz), wpratt@uw.edu (W. Pratt).

1532-0464/$ - see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.jbi.2013.08.011

Table 1
Characteristic of the WebMD diabetes community.

Dates	Total number of posts	Total number of threads	Patients who posted at least once	Total number of moderators	Threads with no replies	Moderator responded threads	Patient only threads
6/07–5/12	65,922	8549	4383	18	0.82%	29.2% (2499)	67.1% (5740)

subset of the data to find ways to achieve a better gold standard and to further understand the remaining social and technical challenges.

2. Background

Social media environments are common venues for health information seeking. Especially for patients with chronic conditions, where the majority of disease management occurs at home, these environments provide a primary resource for people to receive help from their peer patients. The PEW Internet Research found that 23% of patients who have a chronic illness and who use the Internet have gone online to find others with similar health conditions.[1] The survey also showed that, for information on accurate medical diagnosis, prescription drugs, and alternative treatment options, more than 80% preferred to ask professional sources like doctors and nurses rather than fellow patients, friends, and family. On the other hand, the participants preferred to get emotional support and a quick remedy for daily issues from fellow patients. Hartzler and Pratt [2] discussed such distinct expertise roles between (clinical) professionals and peer patients in social media environments. Accordingly, getting balanced help on both emotional and informational support would require both peer patients and professionals as available sources of information. Yet, few researchers have explored ways to weave together peer patients' and professional moderators' knowledge.

In contrast, researchers have long examined novel ways to augment question and answering systems in the medical domain. Lee et al. [3] developed MedQA which provides paragraphs of text for definitional questions in the medical domain, and Cao et al. [4] developed AskHERMES, an online question answering systems for complex clinical questions. Jacquemart and Zweigenbaum [5] explored feasibility of fully automated medical question and answering system and discussed limitations in retrieving specialized knowledge. Liu et al. [6] presented text classification methods for automatically distinguishing consumer questions from health professionals' questions in the health domain to help systems provide appropriate information appropriate for askers' expertise. Cruchet et al. [7] developed supervised approach to recognizing question types in question and answer systems in the health domain. These studies in medical question and answering systems have focused on clinicians' use of the systems, rather than consumers.

From the consumer health perspective, researchers investigated text classification and natural language processing techniques for detecting personal health safety information or health-related expression of consumers in social media. For instance, Himmel et al. [8] developed automatic categorization of medical information requests by users in an online health forum. The authors used manual annotation for generating training data and used word n-grams as features, while refining features using text mining techniques such as principal component analysis [9]. Konovalov et al. [10] developed a classifier using manually selected relevant word unigrams as features to extract combat exposure descriptions from Weblogs. Yang et al. [11] reported best performing information retrieval metrics for detecting adverse drug events from user-contributed content in social media. Qiu et al. [12] showed combining Name and Slang features can improve emotional expression detection in online health communities. While we observe that increasing number of researchers investigate social media for consumer health as a place to incorporate text classification research, little study has examined low-cost approach without having to generate manually annotated training data, and no specific solutions have been proposed for helping to deliver clinical expertise in online health communities.

The increasing number of biomedical information research using text classification in social media and people visiting social media environments seeking health information push us to find scalable solutions for providing clinical expertise to consumers. In this paper, we explore low-cost solutions to assisting moderators of online health communities in delivering clinical expertise to the patient members in the online community.

3. Methods

WebMD.com hosts one of the few online communities that offer moderators in patient forums. Their diabetes community shows the most active participation of both patients and moderators among other WebMD communities. Table 1 shows activity level characteristics of the diabetes community. The WebMD diabetes community has 18 moderators and 4383 patients[2] who posted at least once in the online forum by May of 2012. Thread conversations in WebMD online communities are publicly available. We applied for approval from the University of Washington's Institutional Review Board (IRB) and received a letter stating that our project is unregulated by IRB because the data is publicly available. We wrote a script that crawled WebMD online diabetes community to download all thread conversations.

3.1. Dataset

The dataset consists of 8239 initiated posts made by 2902 unique WebMD diabetes community's patient members from July, 2007 to May, 2012. We tagged each post as *positive* if the question has been responded by a moderator and as *negative* if only patients responded to the post.

As a result, 2499 posts belong to the positive, moderated class—posts answered by moderators, and the remaining 5740 posts belong to the negative, non-moderated class—posts only answered by peer patients. Table A in the Appendices shows example questions answered by moderators and questions that only patients responded to.

3.2. System architecture

In Fig. 1, we illustrate the main components of our system architecture. We used three feature sets. Previous research pointed out the reliability and effectiveness of word unigrams over the

[1] http://pewinternet.org/~/media//Files/Reports/2011/Pew_P2PHealthcare_2011.pdf.

[2] We do not know whether the community members we refer to as "patients" are patients, caregivers, or individuals without diabetes participating in the forum. However, for simplicity, we will use the term "patients" to refer to the community members who are not the WebMD moderators.