The Emotional World of Health Online Communities

Bei Yu
School of Information Studies
Syracuse University

byu@syr.edu

ABSTRACT

This article presents a preliminary study on the emotional world of health online communities. Using sentiment analysis and natural language processing techniques, this study aims to (1) examine the strength of various kinds of emotions (positivity, optimism, negativity, anxiety, anger, and sadness) in online health forum discussions, and (2) compare the emotional status and expression of forum participants under different roles, such as askers and answerers, men and women, and caregivers and patients. This study is expected to improve the understanding of the emotional communication in online health communities.

Categories and Subject Descriptors

H.3.1 [**Information Storage and Retrieval**]: Content Analysis and Indexing – *Linguistic processing*; J.4.3 [**Computer Applications**]: Social and Behavioral Sciences – *Psychology and Sociology.*

General Terms

Emotion, gender, health, computer-mediated communication

Keywords

Emotion, sentiment analysis, health, patient online community

1. INTRODUCTION

According to the Pew Internet Survey on the social life of health information [1], 61% of American adults look online for health information. 41% of them have accessed the large amount of content contributed by the members of health online communities who account for only 6% of the health information seekers. Patients and their families and friends join the health online communities for both informational support and social/emotional support. While concerns arise regarding the accuracy of online medical information, the shared personal experience and emotional support are valued as the signature feature of health online communities around the world [2,3,4]. [6] reported that online diet and weight loss information seekers who visited health forums progressed through the dietary and weight loss plans faster than those who did not visit health forums.

Health online communities have been the subjects of many studies in computer-mediated communication (CMC). Most of the studies

focused on either behavioral patterns through qualitative analysis [3] or themes of discussions through content analysis [2, 4]. The emotional world of the communities has not been sufficiently studied. What kinds of emotions do people express online? To what extent are they positive, optimistic, negative, angry, anxious, or sad? Is there any difference between those who post questions and those who respond? If the characteristics of masculinity and femininity do extend to the virtual world, how are they expressed in online health discussions? Since 52% of all online health inquiries are from caregivers such as family members and friends rather than the patients themselves, how do caregivers and patients differ in emotional status?

This study aims to measure the emotional status of the community members, and discover the patterns in the emotional expression of the members under different roles using sentiment analysis and natural language processing techniques. The remainder of this article is organized as follows. Section 2 describes the research methods. Section 3 describes data preparation and experiment design. Section 4 reports the preliminary results. Section 5 discusses the limitations of current study.

2. RESEARCH METHOD

2.1 Identifying emotion type and strength

LIWC (Linguistic Inquiry and Word Count) is a text analysis program designed by psychologists to gauge the linguistic expression of emotions in texts of a wide range of genres, especially conversations [5]. LIWC counts the occurrences of 2300 words or word stems in 70 categories, including overall affect (sentiment), positive emotion including positive feel and optimism, and negative emotion including anger, anxiety, and sadness. The proportions of words in these categories indicate the emotion strength along the corresponding dimensions. Given a piece of text, like a post, a thread, or even a whole forum, LIWC outputs the strength of each kind of emotion.

2.2 Role identification

Pronouns are important clues to indicate the role of an author. For example, the phrase "my husband" in a post in a prostate cancer forum indicates the author as the wife, while "my PSA" or "my result" indicates the author as the patient self. Of course some posts do not leave any clues for role identification, however based on our sampled examination such posts are not common in health forums and bear few emotional words as well. In this study the OpenNLP toolkit will be used to extract pronouns and noun phrases containing pronouns.

2.3 Role and emotion association

The role and emotion are associated at sentence level. If a sentence contains both role and emotions, for example "my husband is depressed", the emotion expression is associated with

the patient not the author (the wife). The role is propagated to the subsequent sentences until a new role appears in a sentence. Since online discussions are more similar to informal speech than formal writing, we assume that sentences with multiple roles are rare and apply the rule that an emotion expression is associated with its immediate preceding role.

3. EXPERIMENT SETUP
3.1 Data Preparation
All posts from three forums are downloaded from the WebMD health community: the breast cancer support group, the prostate cancer support group, both moderated by WebMD staff, and the cancer treatment group moderated by a physician. The breast cancer and prostate cancer forums are purposefully chosen in order to compare gender impact on emotion expression. Although [4] reported that prostate cancer forums in UK were actually dominated by women, our sampled examination indicates significant number of men participants in the prostate cancer forum on WebMD. The physician-moderated forum is used as a contrast group. We hypothesize that this group has more balanced men and women participants and also low strength of emotion because of the presence of the physician.

3.2 Experiment design
We design two experiments to answer the questions raised in the introduction section. The purpose of the first experiment is to compare the overall emotion strength in each forum, and examine the frequent emotional expressions for each kind of emotion. The second experiment aims to group the sentences by the authors' roles, namely original posters and the responders, men and women, and caregivers and patients, and then compare the emotion strength of different groups.

4. PRELIMINARY RESULT
Table 1 reports the initial results of the first and second experiment. As for the average emotion strength of the original posts, the physician-moderated cancer treatment forum demonstrates the weakest strength in every kind of emotion except for negativity. A possible reason is that the posts (mostly questions for the physicians) focused on the symptoms of diseases like pain and discomfort, which count as negative words. The breast cancer forum demonstrates the strongest overall emotion strength characterized by strong positive emotion and optimism. This result suggests that higher number of female participants correlate with stronger emotional expression in forum discussions.

The average emotion strength of the responding posts follows the same pattern: breast cancer forum > prostate cancer forum > cancer treatment forum. The physician's scientific-styled responses account for nearly half of the total responding posts in the cancer treatment forum, which contributes to the lowest emotion strength. The three forums demonstrate a unanimous pattern that the responding posts are more positive and less negative than the original posts, as evidence of social and emotional support. The breast cancer support group demonstrates the strongest social support.

An initial examination of the frequent emotion expressions reveals some interesting phenomena. For example, "depressed" occur quite some times in the prostate cancer forum but rarely in the breast cancer forum. Further examination shows that almost all of the posts which used "depressed" also used "my husband", which means that depression is often observed by caregivers but rarely self-reported in patient's own posts.

Table 1. Emotion strength in the original/responding posts

Emotion	Breast cancer	Prostate cancer	Cancer treatment
Overall	5.4/7.8	3.4/3.5	3.2/3.1
positivity	4.1/6.5	2.3/2.6	1.8/2.1
Optimism	1.2/1.9	0.6/0.8	0.4/0.5
Negativity	1.3/1.2	1.1/0.9	1.4/0.9
Anger	0.2/0.2	0.2/0.1	0.1/0.1
Anxiety	0.3/0.3	0.3/0.2	0.4/0.2
Sadness	0.4/0.3	0.3/0.2	0.3/0.2

5. LIMITATIONS
Precise gender identification is a challenge in order to measure the gender impact on emotion expression. The accuracy of the NLP-based role and emotion association algorithm will be evaluated. Counting affective word is the simplest sentiment analysis technique with limited accuracy. For example, dealing with words of multiple senses could be tricky. More sophisticated techniques will be further explored.

6. REFERENCES

[1] Fox, S. 2009. The social life of health information. Pew Internet and American Life Project.

[2] Meier, A., Lyons, E.J., Frydman, G., Forlenza, M., and Rimer, B.K. 2007. How cancer survivors provide support on cancer-related internet mailing lists," *Journal of Medical Internet Research* 9(2):e12

[3] Preece, J. 1999. Empathic communities: balancing emotional and factual communication," *Interacting with computers* 12(1): 63–77.

[4] Seale, C. 2006. Gender accommodation in online cancer support groups. *Health* 10(3): 345-360.

[5] Pennebaker, J. W. & Francis, M. E. 1999. Linguistic Inquiry and Word Count (LIWC). Mahwah, NJ: LEA Software and Alternative Media/Erlbaum

[6] White, R.W., André, P. and Tan, D. 2009. Investigating web search strategies and forum use to support diet and weight loss. *CHI' 2009*, Boston, USA, April 2009, pp. 3829–3834.