

# Sentiment Analysis in the Online Health Community

Bing Wu, Yu Peng

School of Economics and Management, Tongji University

**Keywords:** Sentiment Analysis, Social Network Analysis, Online Health Community

**Abstract:** Nowadays, accelerated urbanization and improving living standard have brought some unexpected negative influences, making modern citizens more suffered from chronic diseases. On the other hand, with the innovation of information technology, Web 2.0 is able to provide an accessible and unobstructed channel of communication, which attracts more patients to seek support of health problems through the Internet. Therefore, increasing researches have been focusing on improvement of effective management of online health community. It should be mentioned that community activities reflect the relationship and communication of people, which involves more or less emotional elements. Further, this may influence the management of the entire community. In this research, we utilize social network analysis and sentiment analysis to understand the pattern of communication between patients, in order to provide theoretical and technical support.

## Introduction

Taking diabetes and diabetic complications as an example, a study indicates that the total number of people with diabetes is projected to nearly double in a period of 30 years, from 171 million in 2000 to 366 million in 2030 and the proportion of patients is estimated to rise from 2.8% to 4.4% [1]. Therefore, we are forced to find new countermeasure. In fact, back in 2006, leaders of the UK's National Health Service made a prescient observation: to extend citizens' lives while preventing a budget meltdown, the government must foster a patient-led revolution in healthcare[2].

On the other hand, with the innovation of information technology, the conception of Web 2.0, which emphasizes participation, shared data and collective intelligence, has greatly changed people's daily lives during the last decade. This makes increasing number of people to seek health-related communication through Internet. Aimed at providing a more convenient service, Online Health Community emerges as the times require. It builds a bridge to connect different patients who are physically and mentally suffering from various diseases.

At this point, how to manage the Online Health Community in a more effective and more comprehensive way has become a new research topic. In general, it is the relationship between people that the community activity actually reflects, which contains more or less emotional elements. Further, these elements will indeed influent the pattern of communication and management strategy of the whole community. This paper focuses on utilizing the method of Social Network Analysis and Sentiment Analysis to find general regularity of online communication between patients

## Literature Review

**Online health community.** According to a study by the Pew Research Center [3], 80% of adult Internet users in the U.S. use the Internet for health-related purposes. As one of applications, there are increasing number of people tying Furthermore, the impact of new information technology still keeps increasing, which shows an enormous potential of this new-mode health community[4]. The

combination of traditional healthcare issues and modern information technology is also called "Health 2.0", whose core conception is to provide patients an unprecedented way to connect with each other online with Web 2.0. Though a review by Tom et al. says that Health 2.0 is still a developing and unaccomplished conception, it also indicates this new conception has caught a positive attention of government and researcher[5].

Currently, most Online Health Communities provide basic social media functionality such as forums, blogs and member profiles to promote the connection between patients and provide either informational or emotional support through online social networks. These Online Health Community has accumulated massive information of interaction between users. According to a survey by Boase et al. [6], 81% of those users asked for health information from their core ties, and 46% from at least one of their significant ties.

In fact, there have been a numerous researches focusing on extracting therapic experience from the communication of patients by statistics methods or machine learning algorithm, as the original contents are unstructured data. For instance, Chen [7]used cluster analyses to understand various patient experiences. And Lu [8] tried SVM, Naive Bayes and C4.5 to build an automatic topic identification of health-related messages, in order to assist patients in searching their interesting topics.

On the other hand, however, less attention has been paid to how patient collaborative friendships form in Online Health Communities. To take full advantages of Health2.0, we should realize that the main distinction between traditional health community and Health2.0 is just the convenient and unobstructed channel of communication. We are supposed to utilize it to manage the Online Health Community in a better way.

**Social network analysis.** Online Health Community is defined as a collection of people who concern on common healthcare problems. If we regard the patients in this collection as "actors" and the communication between them as the "relation between actors", then we could get a basic model of Social Network Analysis.

Actually, Patient network analysis is suggested by Chen et al. [9] as one of the most promising areas for health information technology research. Social Network Analysis is widely used to make deeper understanding of features of the online health community. According to a review by Song et al. [10] Social Network Analysis is the most frequently used method in the study of selected patient network, which generally describes the structural characteristics of the patient network and the positions of individuals.

## Research Methods

Text-based Sentiment Analysis, which aims to convert unstructured human emotions into measurable values, is an interdisciplinary research. In general, the research object of the Sentiment Analysis is the texts on the Web, especially those posted by users on their own initiative and related to one certain topic.

Based on the mode above, Wilson et al. introduced a universal method of Sentiment Analysis: OpinionFinder (OF). This system utilizes several existing natural language processing tools to finish analyzing of the polarity of sentences, POS tagging, affix trimming and other operations. Then according to necessary context information, a classifier would be used to get the final result.

Despite OF could identify the subjectivity of opinion, it is not capable to measure emotional expression exactly. To solve this problem, Esuli and Sebastiani [14]indicated that SWN lexicon (Senti Word Net lexicon) could help. The system is supplemented by the Senti Word Net (SWN) lexicon, which provides information on the intensity of sentiment expressions Its comprehensive

and systematic record of the sentiment index information of each individual entry of words, but its weakness is not discrimination based on the context of sub-words were ambiguous.

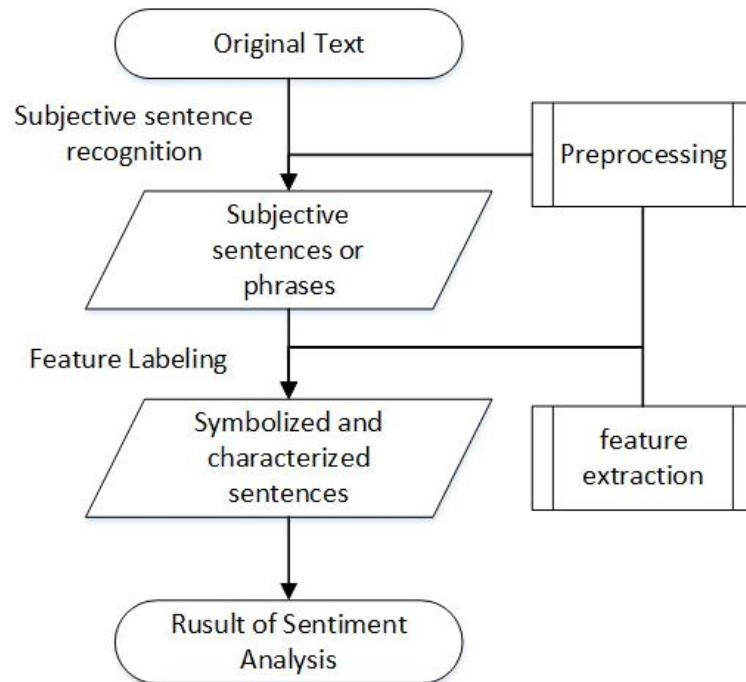


Figure 1 Basic and typical process of Sentiment Analysis.

According to Chang's research, mutual aids of patients in the Online Health Community can be classified as informational support, network support, esteem support and emotional support. The first two are mainly objective contents which embody little emotional elements; and the latter two obviously contain more positive and optimistic keywords. In fact, the primary purpose of patients' participation in the Online Health Community is to seek support, so we can infer that the average sentiment score of the whole community tends to be positive. Thus, we make following hypothesis:

H1: Emotional expression in the social network tends to be positive.

The research of McPherson et al. demonstrated that in the social network, the similarity of personal attributes would lead a closer relationship between each other. In our research, this tendency is reflected by frequent communication of patients who share same experience and hold similar attitudes towards disease. Thus, we make following hypothesis:

H2: In different subgroups, the polarity and intensity of users' emotional expression tend to be similar.

As mentioned before, the main purposes of communication between patients are to get informational or emotional support from other patients who suffer from the identical disease. Therefore, though there are a number of subtopics in the community, users won't digress too much, as the main topic is constrained in the range of one certain disease. This leads to less possibility of isolation of various subtopics. Thus, we make following hypothesis:

H3: There is a significant centralization of the nodes, which means less subgroups.

## Conclusions

In the online health community, patients have frequent communication and integrate as a group in which they are able to provide or receive support. In general, the way patients keep connection with other is positive and beneficial, which demonstrates Online Health Community does create a patient-led service and help patients obtain emotional support. Through the analysis of the results of this experiment, we could get following conclusions:

( 1) The majority of people and information in the whole community are positive. To some Extent, it shows that the online health community is indeed benefit for providing emotional support to patient. Manager of the community could understand the atmosphere of the whole community by measuring the average sentiment score.

( 2) A more positive user engages with more topics and have a deeper degree in the social network. This shows that in Online Health Community, positive users tend to be willing to provide emotional support. These users should be encouraged as key users, because they are the critical point of whether the community could provide emotional support.

( 3) Users who have similar emotional attitude tend to communicate with each other more frequently. This offers an inspiration of friend recommendation, referring to their emotional situation. We could encourage communication of patients with either same or different emotional attitudes, to help patients find someone who could provide either understanding or inspiration.

## References

- [1] Wild S, Roglic G, Green A, et al. Global prevalence of diabetes estimates for the year 2000 and projections for 2030[J]. Diabetes care, 2004, 27(5): 1047-1053.
- [2] Cayton H. The Flat-Pack Patient? Creating Health Together. Patient Education and Counseling. 2006;62(3):288–290.
- [3] Fox, S.: The Social Life of Health Information, 2011. Pew Research Center's Internet & American Life Project (2011)
- [4] Randeree Ebrahim. Exploring technology impacts of Healthcare 2.0 initiatives. Telemed J E Health. 2009 Apr;15(3):255–60. doi: 10.1089/tmj.2008.0093.
- [5] Van De Belt T H, Engelen L J, Berben S A A, et al. Definition of Health 2.0 and Medicine 2.0: a systematic review[J]. Journal of medical Internet research, 2010, 12(2).
- [6] Boase J, Horrigan J B, Wellman B, et al. The strength of Internet ties: The Internet and e-mail aid users in maintaining their social networks and provide pathways to help when people face big decisions[J]. Washington, DC: The Pew Internet and American Life Project. Retrieved on June, 2006, 22: 2007.
- [7] Chen A T. Exploring online support spaces: Using cluster analysis to examine breast cancer, diabetes and fibromyalgia support groups[J]. Patient education and counseling, 2012, 87(2): 250-257.
- [8] Lu Y. Automatic topic identification of health-related messages in online health community using text classification[J]. SpringerPlus, 2013, 2(1): 1-7.
- [9] Chen, H., Chiang, R.H.L., Storey, V.C.: Business intelligence and analytics: From big data to big impact. MIS Quarterly 2012, 36, 1165–1188
- [10] Song X, Jiang S, Yan X, et al. Collaborative Friendship Networks in Online Healthcare Communities: An Exponential Random Graph Model Analysis[M]. Springer International Publishing, 2014: 75-87