A Thematic Portfolio and Recommended Study on the Usefulness of Online Medical Reviews Based on QCA Methods

Wang, Yonghong; Huang, Lei

Published in:
Journal of Physics: Conference Series

Published: 01/01/2021

A Thematic Portfolio and Recommended Study on the Usefulness of Online Medical Reviews Based on QCA Methods
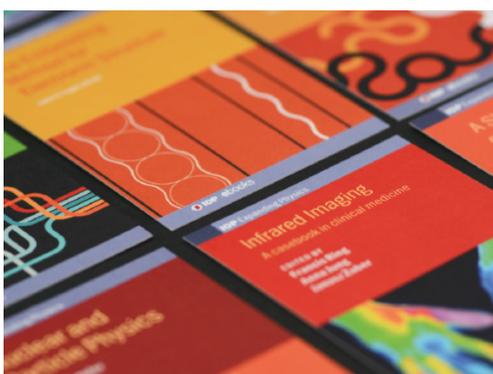
View the article online for updates and enhancements.

A Thematic Portfolio and Recommended Study on the Usefulness of Online Medical Reviews Based on QCA Methods

YonghongWang[2], LeiHuang[1,*]

[1]Hunan City University, No.508, Yingbing Road, Yiyang, China;

[2]Hunan Provincial Engineering Research Center, Intelligent Monitoring of Ecological, No.508, Yingbing Road, Yiyang, China; City University of Hong Kong, PhD of Computer Science.

 *45506034@qq.com

Abstract. The online comment for doctors plays an important role in assisting patients to master the real medical situation and choose medical treatment, which greatly reduces the adverse effects of online medical market information asymmetry. but because ordinary patients lack professional medical knowledge and can not accurately and efficiently write comments that are useful to similar patients, we identify useful topics from online medical reviews and make combination recommendations based on patients with different diseases. guide patients to write comments from the most useful dimensions to alleviate the problem of information overload, thus maximizing the limited human resource utility. This study aimed at online medical platforms such as good doctors and micro-doctors collected online doctor reviews for major diseases and used LDA topic mining and fsQCA fuzzy set qualitative comparative analysis to analyze key topics that affect the usefulness of reviews and optimal topic combinations under different disease types.

Keywords: online medical review, topic combination, usefulness, fuzzy set analysis

1. Introduction

Online medical treatment means that hospitals and doctors use the Internet or mobile Internet to provide medical and health services to ordinary users and patients, including online health care, online diagnosis and treatment, medical information query and other forms of medical and health services.

According to China Online Medical Industry Data Monitor Reporter, in the field of online medical treatment, because of the information asymmetry caused by the limitation of time and region, it is difficult for patients to master the real medical effect of hospitals and doctors(Zeng et al., 2017)[1], So online medical reviews are an effective way for patients to obtain reliable information, and research has confirmed that online medical reviews can reflect the true quality of doctors' services(Gao et al.,

2015)[2] Products reviews in similar product areas are important adjuncts to consumer decision-making behavior, and online medical reviews, especially for doctors, have positive effects on patients' medical choices(Wu et al., 2016)[3].

To explore how to improve the usefulness of online medical reviews, this study takes the online medical platform like good doctor online and micro-doctor as the object and identifies four types of diseases, such as acute, chronic, pediatric, and major diseases, using LDA topic mining methods to identify comment topics, analyzes the effects of different topic combinations on the usefulness of comments based on fuzzy set comparative qualitative analysis (fs QCA) methods, studies the effects of disease types on the usefulness of topics and thematic combinations on the usefulness of comments, and realizes the function of recommending usefulness topics for online medical reviews, Optimizing the content organization of online medical reviews can help guide users to write more useful doctor reviews and improve the utility value of reviews for similar patients.

2. Related research

At present, foreign scholars mainly study how to use computer technology and Internet technology in online medical treatment. The electronic medical records in the field of online medicine are explored by Charles Safran(Safran et al., 1999)[4] so that the medical system can better supervise and monitor the care of patients, and people can focus more on interpersonal communication and provide health care services. The Internet is regarded as a search tool by Houston JD and Fiore DC(Houston et al., 1998)[5], which uses the Internet to collect user information to provide more accurate user positioning for online medical treatment.

In recent years, domestic scholars mainly focus on the relationship between online medical doctors and users, the role of online medical services to stakeholders, and influencing factors. Li Yanganalyzed how doctors participate continuously on online medical websites(Li et al., 1998). There are two main ways to encourage doctors to participate, one is the feedback mechanism of the website, the other is that different behaviors of patients will have an impact on doctors' contribution behavior. And behavioral differences between doctors with different diseases. However, there are few scholars to study the usefulness of online medical comments on their own attributes and text content, and the research of combining feedback with the online medical platform is not deep.

3. Thematic analysis and combination of user reviews on the usefulness

This study suggests that writing online medical reviews with high usefulness requires three foundations:1) the content of the review contains several different topics, each including the attributes of the patient's comments;2) whether the fit between the topics depends on the co-occurrence frequency of the related topics ;3) the combination of different topics have an important impact on the usefulness of the comments. Therefore, the focus of this study is on two aspects: first, how to determine the correlation between the combination of topics and the usefulness of comments; second, how different types of diseases will affect the combination of useful topics. The overall framework is shown in Fig.1.
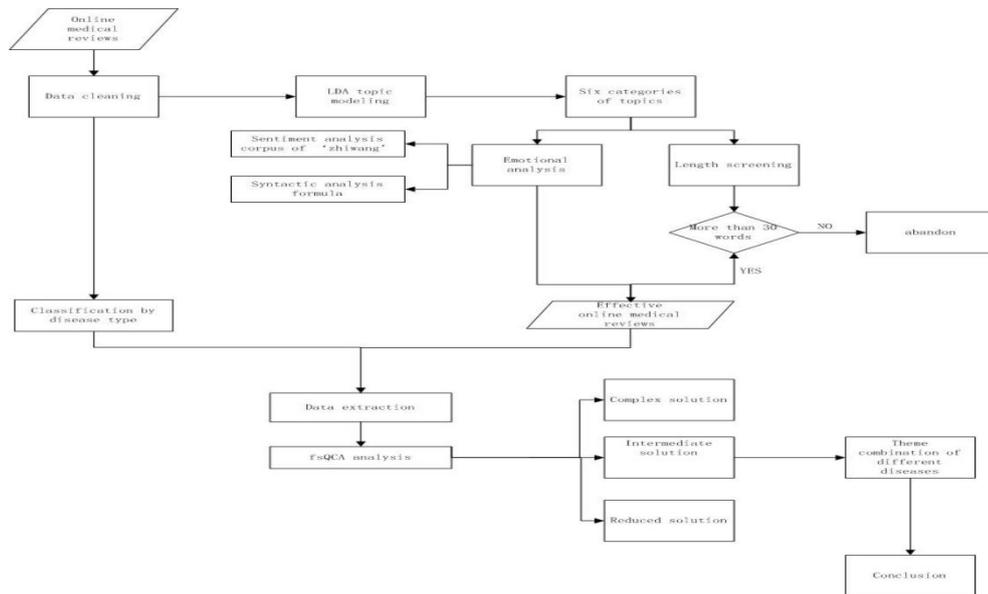
Fig.1 Overall framework

Based on the above theoretical viewpoints, this study constructed a machine learning model for automatically extracting online medical reviews using the Chinese LDA topic model. taking nine diseases under four types of disease, acute, chronic, pediatric, and major diseases as an example, the fuzzy set qualitative comparative analysis (fsQCA) method was used to excavate the combination of key topics that affect the usefulness of reviews and the best topics under different disease types.

4. Experimental analysis

4.1. Experimental data sets

This paper uses the python Selenium(browser automation test framework) module to simulate the browser method to crawl the patient's comment data on the doctor in the two online medical service websites (good doctor and micro-doctor) . The range of crawling includes major diseases: gastric cancer, lung cancer by attending doctors nationwide.

Type	Category
Major diseases	Gastric cancer, lung cancer

Tab.1 Classification of diseases

4.2. Analysis of reasoning results of topic modeling

LDA M`odeling Process. Data analysis is based on the data set after the previous data cleaning. Firstly, the data is partially extracted to determine the availability of the data, so that the data is as complete, effective, and unreadable as possible. After the availability test, the stage of data analysis is

officially started. During LDA training, two parameters are particularly important, one is the number of inferred topics, the other is the number of subject words in each topic. In order to get a better convergence effect, the parameters need to be corrected in the process of analysis. After several revisions, the topic value is set to 10, and the number of words displayed in each topic is 6. The parameter selection is based on the expected dimension. The number of topics extracted is about 6, and the 6 words can comprehensively describe the core points of a topic. Finally, the words that reach the result of topic inference are of practical significance and can be clearly distinguished.

Topic modeling results. The results of this data analysis get 10 topics, each subject takes the first 6 words. The above results show that the theme results are coarse-grained and need to be further refined by artificial judgment. Finally, we have extracted six major themes from the results: medical skill, attitude, procedure, environment, symptoms, and etiology.

Topic inference. According to the topic modeling process, we get the above six topics. We need to further determine the topic combination contained in each comment. In order to get the corresponding combination, we use topic inference to compare the similarity between each comment and the topic result to get the probability that the comment belongs to each topic and synthesize these probability values as the possible result of the topic combination of a comment.

Comments on usefulness. We have to further extract the data we need from the data set. Since the judgment boundary of the usefulness of comments is relatively vague, we comprehensively consider the use of sentiment analysis and the analysis of the length of comments in the process of data analysis to quantitatively evaluate the usefulness of comments.

4.3. Data extraction

This analysis uses the "Zhiwang" emotional analysis words set, the corresponding level of vocabulary has a corresponding score,>1 means to strengthen emotion,<1 means to weaken emotion. In order to ensure that the sampling data cannot be removed from the original characteristics, so the stratified sampling scheme is adopted according to the disease type, and the final result is 300 cases of each type of disease.

4.4. fsQCA Comparison and Analysis of Fuzzy Sets

Selection of Typical Cases. In this paper, the four major diseases, major diseases, acute diseases, chronic diseases, pediatric diseases, each stratified extraction of 300 comments.

Choice and measure of variables. Here, comments within 0-10 points are recorded as more useful comments, assigned 1. Comments beyond 10 points are recorded as useless comments, and 0. In the selection of conditional variables, six main topics are obtained according to topic modeling: medical skill, attitude, procedure, environment, symptoms, and etiology.

Comments	Attitude	Medical Skill	Procedures	Environment	Etiology	Symptoms	usefulness
Comment 1	7.27	0.1	7.98	0.1	4.8	1.72	1

Tab. 2 Findings (Part)

After determining the value of the conditional variable, it is also necessary to calibrate the variable. calibration converts the original value into a fuzzy set, making the variables match and conform to external standards. The fuzzy set is calibrated using theoretical and substantive standards external to the data. three anchors need to be preset: complete membership, intermediate points, and complete non-membership. the transformed set membership is between 0-1. In this paper, the threshold of complete membership is 0.95, the middle point is 0.5, and the total non-membership is 0.05.

Comments	Attitude	Medical Skill	Procedures	Environment	Etiology	Symptoms	usefulness
Comment 1	0.65	0.05	0.77	0.05	0.94	0.21	1

Tab. 3 Comments after calibration (part)

FsQCA results analysis. After determining the variables, using the analysis software FsQCA3.0, three kinds of solutions will be obtained: complex solution, intermediate solution, and refined solution.

Configuration	Raw coverage	Unique coverage	Consistency
Attitude *~ Medical Skill * Environment * Procedures	0.25	0.11	0.88
Attitude * Medical Skill * Environment	0.2	0.15	0.89
Solution coverage	0.37		
Solution consistency	0.89		

Tab. 4 Intermediate solutions for major diseases

Combination one (attitude *~ medical skill * environment * procedure) shows that if patients with major diseases are in a hospital with good treatment environment and procedures and encounter doctors with good communication attitude, even if the doctor's own level is not very high, It can still be provided to patients with reliable medical care. Combination two (attitude * medicine * environment) says that for doctors with both medical ethics and medicine, no matter what kind of patient they face, they can make them keep trust. Of course, the necessary medical environment is also indispensable for the more intractable major diseases.

5. Conclusion

By mining the online doctor reviews of the two mainstream online medical platforms, the results of six topics are summarized, and the optimal topic combination under each type of disease is obtained by a fuzzy set qualitative comparison method. It is recommended to write medical reviews for users of different disease types to guide them to display useful information for similar patients.

References

[1]   Yiqiao Z, Hong W and Naiji L 2017 An Empirical Study on the Influencing Factors of Patients'On-line Comment Behavior - A Case Study of Registration Network*Smart Health* **22** pp12-17

[2]   GaoG and et al2015Vocal minority and silent majority ： how do online ratings reflect population perceptions of quality *J. MIS Quarterly* **39** pp5 65-590

[3]   Wu H and Lu N2016 How your colleagues' reputation impact your patients' odds of posting experiences ： Evidence from an online health community*Electronic Commerce Research &Applications* **16** pp 7-17

[4]   Safran C 1999 Electronic Medical Records ： A Decade of Experienc *J. Jama the Journal of the American Medical Association* **38** pp 308-312

[5]   Houston J D and Fiore D C 1998 Online medical surveys ： using the Internet as a research tool *Md Comput* **15** pp116-120

[6]   Yang L 2015 Study on the influencing factors of doctors' contribution behavior in online medical community