

Research Article

What Concerns Consumers about Hypertension? A Comparison between the Online Health Community and the Q&A Forum

Ye Chen¹,, Ting Dong¹, Qunwei Ban¹, Yating Li^{2,*}¹School of Information Management, Central China Normal University, Wuhan, 430079, China²National Engineering Laboratory for Educational Big Data, Central China Normal University, Wuhan, 430079, China**ARTICLE INFO***Article History*

Received 19 Sep 2020

Accepted 31 Jan 2021

Keywords

Hypertension information needs

Social media platform

Topic modeling

Biterm topic model

ABSTRACT

In this paper, the Biterm topic modeling method and comparative analysis were employed to identify consumers' information needs on hypertension and their differences between the Online Health Community and the Q&A Forum. There are common information needs on both platforms but consumers on MedHelp discussed more about pathology and pharmacology, and mental health of hypertension than those on Quora. The results can help consumers, social media platform designers, and medical professionals better understand consumers' information needs on hypertension.

© 2021 The Authors. Published by Atlantis Press B.V.

This is an open access article distributed under the CC BY-NC 4.0 license (<http://creativecommons.org/licenses/by-nc/4.0/>).**1. INTRODUCTION**

The Internet is a vast source of information. People are increasingly using it to search for health information, consult with health professionals, and participate in health support groups. According to the study conducted by the Pew Internet & American Life Project [1], about 80% of Internet users in the United States are reported seeking health-related information online and the results show that it is one of the most prevalent activities to search information about health or medicine online.

In recent years, social media platforms have become important data sources for following infectious diseases spread trail [2], detecting consumers' mental condition [3,4] or identifying consumers' information needs on hypertension. Hypertension, or high blood pressure, is one of the most common chronic diseases and has become a global health threat causing high mortality and heavy health-care cost. It is also known as “the silent killer” because it shows no early obvious symptoms [5]. Although it is a preventable and treatable disease, it will lead to serious complications and result in disability when consumers do not take correct measures to control it. There are several barriers related to hypertension control and the most important barrier in controlling this disease is the lack of information about diverse aspects of hypertension [6]. Thus, one of the most significant challenges in controlling hypertension is

to figure out consumers' information needs and then offer related information to increase their awareness level. Moreover, the role of information needs on hypertension requires particular attention because evidence has shown that consumers' information needs differ from health providers' perception of those needs, and when they are left unresolved this may lead to hypertension control difficulty, such as lower medication adherence [7]. The study has also shown that nurses underestimated patients' information needs leading to poor concordance between them in chronic heart failure [8]. Understanding consumers' information needs is crucial to help health-care providers offer the right and needed information to consumers, to help consumers improve hypertension awareness more efficiently, and to help consumers take appropriate efforts to control or prevent hypertension.

Social media platforms are a group of Internet-based applications that build on the ideological and technological foundations of Web 2.0, and that allow the creation and exchange of user-generated content [9]. They provide general users, patients, and their relatives the ability to access health information from other users, to ask for help and advices from other users, to make contributions to others, to receive assistance from the forum, and to share their experiences in the community. Today, social media platforms are pervasive, rapidly evolving, and increasingly influencing people's daily life and their health behavior. A researcher suggested that “social media is where the future is, and most importantly, that's where our patients are going to be” [10].

*Corresponding author. Email: liyt@ccnu.edu.cn

Millions of consumers engaged in social media platforms to search health information and release information about their health and related questions. Over the past several years, a boom emerged in social media platforms in the health domain. Given the lack of knowledge on the use of health social media platforms, further understanding of hypertension information needs on different social media platforms would help consumers quickly obtain required information. Meanwhile, it helps social media platform designers to construct their own content advantages, and helps health professionals provide more consumer-centered care to hypertension control. Prior research has discovered hypertension information needs on several social media platforms [11–14]. In this paper, we extend previous work by making comparison of information needs on different social media platforms.

In this study, two typical social media platforms, the online community and the question and answer forum, are chosen to investigate:

- RQ1: What kinds of hypertension-related information do consumers share in the online health community?
- RQ2: What kinds of hypertension-related information do consumers describe in their questions in the question and answer forum?
- RQ3: Do consumers' information needs on hypertension in the online health community differ from the question and answer forum?

2. LITERATURE REVIEW

2.1. Hypertension Information Needs

Some researchers have explored consumers' hypertension information needs on several social media platforms such as the online community and the social network site. Abdullah *et al.* [12] collected posts about hypertension from the online community (www.MedHelp.org) and manually classified them into 10 categories: risk factors, pharmacologic and nonpharmacologic management, diet, diagnostics, symptoms, sequence, causes, blood pressure readings, and others. To explore what kind of hypertension information consumers preferred on YouTube. Two hundred and nine videos about hypertension on YouTube were assessed. It was found that consumers tended to watch videos with exactly targeted patients. The result suggested that consumers would like to view videos that had the highest domain on the alternative treatments, which indicated that the information about hypertension alternative treatments were what consumers were interested in [13]. In the meantime, Mamun *et al.* [14] chose hypertension-related Facebook groups as a data source. This first research on hypertension systematic search on Facebook aimed to characterize consumers' main purpose, main discussion topic, and other features. By employing the content analysis method, these groups were assigned into 7 major categories, such as awareness-creating groups and experience-sharing groups. The results showed that these groups were created mainly for improving consumers' hypertension awareness (59.9%) and some groups functioned as support groups for caregivers and patients (11.2%). When content of the top-displayed most recent wall posts were analyzed, researchers discovered that the themes of product promotion had the highest coverage on posts (21.3%) while

some posts discussed hypertension-related information (20.1%) and sharing-related Web addresses (13.4%).

2.2. Health Information Need on Online Communities

Online communities are regarded as a significant source of health information and are beneficial for consumers to search health information, take advice, and shared health-related experiences [15].

Park *et al.* [16] identified consumers' information needs on cancer based on the data from MissyUSA, one of the biggest online communities for Korean Americans. By calculating the frequencies and percentages of a cancer-related term, their results showed most cancer-related posts (71.4%) were linked to medical topics consisting of 9 subtopics. The most frequently discussed subtopic was treatment (24.1%). Nath *et al.* [17] extracted and categorized the URLs from WebMD, one of the most active online health communities, to explore what kinds of websites consumers needed and shared. Their results revealed consumers rarely shared social media websites (0.15%) and most shared.com (59.16%) and WebMD internal (23.2%) websites. In addition, Mi *et al.*'s study [18] investigated the types of social support information consumers needed by analyzing the posts content obtained from QuitStop, a community designed for people suffering from the tobacco quitting process. The results indicated two main kinds of social support, nurturant support, and informational support. Consumers are more concerned about the former (533/881) than the latter (422/881).

2.3. Health Information Needs on Question and Answer Forums

Question and answer forums create a consumer-centered environment where consumers are free to post questions related to the topic and other consumers, who have experience or knowledge on this topic, will offer the answers to these questions. It provides a new way for researchers to better understand consumers' information needs, and has been identified as an important information resource for consumers to receive health-related information especially on chronic disease [19].

Leanne *et al.* [20] collected the questions on the topic of eating disorders posed by teens from Yahoo! Answers to understand their information needs. Through a content analysis of posts, the schema of questions mainly contained seeking information, seeking communication, seeking self-expression, seeking help to complete a task, and seeking emotional support. Lynn and Zhang [21] randomly selected 200 posts about cervical cancer from Yahoo! Answers and characterize their information needs using content analysis. Their experiment reflected that the information consumers were most concerned about encompassed four aspects: sexual behavior, disease intensity, time, and control over disease. To discover which autistic topics consumers were interested in, an LDA model was applied to identify topics in autism-related questions on Quora. This experiment showed clinical information had drawn most of consumers' attention (34% from 2010 to 2017; 68% in 2019) [22].

According to Timmins [23], in the health field information needs can be simply explained as what patients need to know. In a broad sense, it is widely used to represent gaps or deficiencies in patient/family knowledge that may be corrected through information. Moreover, Yang et al. [24] suggested that the topics hidden in the online health information represented the health information the consumers want to obtain. Thus, it is crucial to identify related topics, analyze, and summarize topics from online health information datasets for discovering consumers' information need.

Consumers' health information needs have become a popular research interest. Previous research on consumers' hypertension information needs mainly paid attention to online communities or social network sites but rarely to the question and answer forums. Most of existing research chose to calculate the frequencies of term [16], analyze content [18,20,21], and to clarify specific information needs (such as cancer [16], smoking quitting [18], eating disorder [20], autism [21]). Moreover, there is no research conducting a comparative analysis of consumers' hypertension information needs on different type of social media platforms.

3. RESEARCH METHODS

3.1. Selection of Social Media Platforms

This study selected the online community MedHelp [25] and the Q&A forum Quora [26] as the research data sources. MedHelp is one of the most popular online health communities which attracts more than 12 million users monthly browsing the homepage [27]. On MedHelp, nearly 300 sub-communities were constructed for consumers to discuss and share information of 164 different diseases/issues, such as diabetes, pregnancy, diet and fitness, depression, high blood pressure, and so on [26]. Quora, publicly available in 2010, has become one of the most commonly used Q&A forums and has attracted about 300 million unique, monthly users [27].

The data analysis process is outlined as Figure 1. The details are described in the following parts.

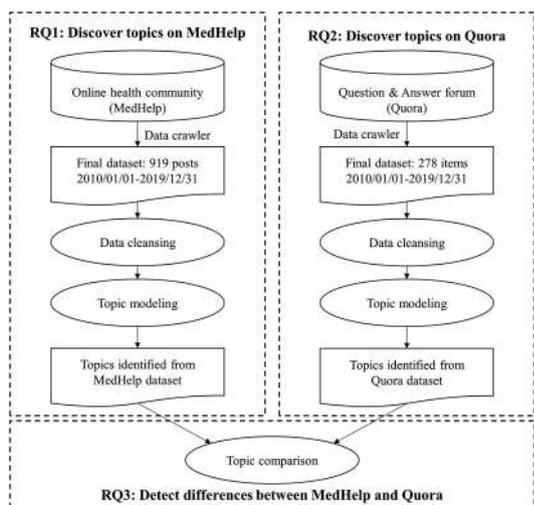


Figure 1 | Data analysis process.

3.2. Data Collection and Cleansing

To collect research data, a web crawler program developed in Python was used to obtain all items (each item consisted of one question and its answers) in the topic of Hypertension on Quora and all posts (each post consisted of one topic and its comments) in the sub-community High Blood Pressure on MedHelp on January 15, 2020. Compare to MedHelp which was founded in 1994, Quora was opened to the public in 2010. To ensure the consistency of research data, the posts and items published from January 1, 2010, to January 31, 2019, were involved in this study. And posts or items which were irrelevant to hypertension were manually removed. As a result, 919 posts and 278 items related to hypertension were collected. All posts or items were converted into a standard record with a post/item number, post/item time, topic/question text, and comment/answer text. After the raw data was collected, the Natural Language Toolkit was implemented for the three-step data cleansing process: letter case formalization, punctuation removal, and lemmatization.

3.3. Topic Modeling

Topic modeling is one of the most frequently used and powerful techniques in text mining for latent data and text documents relationship discovery [28]. In this research, the input data collection of topic modeling are posts or items obtained from the social media platforms. Most of them are short text that are made up of only a few words and lack contexts. Therefore, the corpus for topic modeling was sparse. Conventional topic methods, such as probabilistic latent semantic analysis (PLSA) [29] and Latent Dirichlet allocation (LDA) [30], cannot perform well on short texts. Biterm topic model (BTM) was employed to carry out the topic modeling process. This alleviated the data sparsity problem existing in topic modeling based on short texts [31].

BTM is an unsupervised machine learning generative probabilistic model, which can identify latent topics by modeling the generation of biterms in corpus [31]. A biterm consists of any two single words. The biterms extracted from all the documents make up the training data of BTM. Unlike PLSA and LDA modeling generative process of documents to discover latent topic, the basic idea of BTM is modeling the generation of biterms. Suppose α and β are the Dirichlet priors, the generative process of BTM is described as Figure 2.

The generation procedures of BTM are as follows:

- (i) For the whole corpus: draw a topic distribution $\theta \sim \text{Dirichlet}(\alpha)$

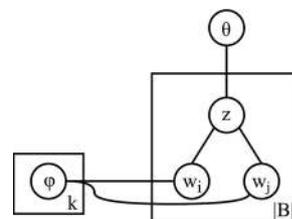


Figure 2 | Graphical model of Biterm topic model (BTM).

- (ii) For each topic $z \in [1, k]$: draw a topic-word distribution $\varphi_k \sim \text{Dirichlet}(\beta)$
- (iii) For each biterm $b_i = (w_i, w_j) \in B$: extract a topic z randomly from the topic distribution $\theta : z \sim \text{Multi}(\theta)$

According to the above procedure, the BTM directly utilizes the co-occurrence words to set up the topic model, which overcomes the data sparsity problem on the social media platforms and considers semantic relationship between words for better understanding short texts.

In the process of topic modeling, the input data were two cleansed datasets (Quora and MedHelp) in a standard record format. The BTM method extracted biterms $B = \{b_1, b_2, b_3 \dots, b_{|B|}\}$ ($b_i = (w_i, w_j)$) and generated the word co-occurrence patterns from the input datasets. The output of BTM consisted of topics distribution $\theta = \{\theta_{T_1}, \theta_{T_2}, \theta_{T_3} \dots, \theta_{T_K}\}$ and topic-word distribution $\varphi = \{\varphi_{T_1}, \varphi_{T_2}, \varphi_{T_3} \dots, \varphi_{T_K}\}$ for two datasets, to discover their latent topics $T = \{T_1, T_2, T_3 \dots T_K\}$.

In the model of BTM, K , α and β have a strong influence on performance [32]. The two hyperparameters are denoted for distributions of topics over document (α) and of words over topic (β). Distribution can concentrate strongly on the center of the simplex and main probabilities are close to $1/K$ with high value of α , which means a very flat probability result. On the other hand, according to Tapi Nzali [33], smaller value of α makes it difficult to interpret as probabilities being further away from $1/K$. The value of β has implications for granularity of model. It means β sets different scales, at which document can be classified into a series of topics. A high value of β , under scientific disciplines, can help a model to decrease the number of topics as it reduces the sparsity of $p(w|z)$. While, a small value of β can lead to more topics that emphasize other rather specific field of research [34].

However, in spite of the variety of topic modeling algorithms proposed, choosing an appropriate number of topics (K) has still been a common challenge in applying these algorithms successfully. Results can be overly broad or “over-clustering,” when the topics are too few or too many for a given corpus [35]. Thus, it is important to determine the optimal number of topics. Moreover, it can minimize the modeling time and make the results contain as much information as possible [36]. According to researchers [37], the optimal number of topics was usually decided by perplexity and topic stability analysis. Perplexity is a common indicator to measure the performance of probabilistic models [38]. Its algebraic format is the geometric mean per-token likelihood [39]. In BTM, a biterm consisting of two words is the token. Perplexity value stands for the uncertainty of topic prediction ability. Therefore, the smaller the perplexity value is, the stronger the model prediction ability is and the better performance the model has [40]. Perplexity is calculated in BTM as

$$Perplexity = p\left(\vec{b}|M\right)^{-\frac{1}{|B|}} = \prod_b^{ |B| } p(b|M)^{-\frac{1}{|B|}} \quad (1)$$

where \vec{b} represents corpus of biterms extracted from documents, and $p(b|M)$ is the probability that the model generates biterms b .

$p(b|M)$ is calculated as

$$p(b|M) = \sum_z^k p(z)p(w_i|z)p(w_j|z) = \sum_z^k \theta_z \varphi_{z,b_{w_i}} \varphi_{z,b_{w_j}} \quad (2)$$

Thus, the perplexity in BTM can be illustrated as

$$Perplexity = \left[\prod_b^{|B|} \left(\sum_z^k \theta_z \varphi_{z,b_{w_i}} \varphi_{z,b_{w_j}} \right) \right]^{-\frac{1}{|B|}} = \exp\left| -\frac{\sum_b^{|B|} \log\left(\sum_z^k \theta_z \varphi_{z,b_{w_i}} \varphi_{z,b_{w_j}}\right)}{|B|} \right| \quad (3)$$

Topic stability analysis measures the average semantic distance between topics generated by BTM. Between-topic semantic similarity (BTS) treats each topic as a vector and judges the semantic similarity between topics based on the cosine of the angle formed by topic vectors. BTS is calculated as

$$BTS = Sim(Topic_A, Topic_B) = \frac{\sum_{i=1}^n ww_{iA} ww_{iB}}{\sqrt{\sum_{i=1}^n (ww_{iA})^2} \sqrt{\sum_{i=1}^n (ww_{iB})^2}} \quad (4)$$

where $Topic_A, Topic_B$ represent any two topic vectors. Average semantic distance between topics is calculated as

$$BTS_arg = \frac{\sum_{A,B=1}^k Sim(Topic_A, Topic_B)}{C_k^2} = \frac{2 \sum_{A,B=1(A<B)}^k \sum_{i=1}^n ww_{iA} ww_{iB}}{k(k-1) \sqrt{\sum_{i=1}^n (ww_{iA})^2} \sqrt{\sum_{i=1}^n (ww_{iB})^2}} \quad (5)$$

Therefore, the smaller the cosine value between-topic vectors the smaller the semantic similarity is and the farther apart topic semantic is, the weaker the topic semantic independence is and the worse the performance of the model is.

When the number of topics is set differently, perplexity and average topic semantic distance of the model results vary. As mentioned above, the smaller values of perplexity and average topic semantic distance, the better performance the model has, because for a given corpus, smaller perplexity value means better model prediction. Meanwhile, smaller average topic semantic distance represents higher topic stability. However, the optimal number of topics cannot be decided entirely by the smallest value of perplexity or the average topic semantic distance. Because the topic model is based on sampling statistics and probability, the model result is not precise. Thus, the selection of optimal number of topics should not only take the quantitative results into consideration, but also include the perspective of qualitative.

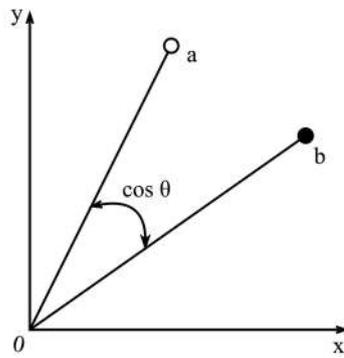


Figure 3 The degree of topic similarity.

3.4. Comparative Analysis

By applying BTM, each topic is represented as a word vector $T_i = (w_1, w_2, \dots, w_n)$. These word vectors are imagined as line segments starting from origin and pointing to different directions. Similar topics share proximal vectors and dissimilar topics have vectors which are far away in the vector space [41]. Cosine similarity is a traditional and efficient method to measure the degree of similarity between two vectors. Many researchers applied this method to do text mining process to find topic similarity [42–45]. The smaller the angle formed between two vectors, the larger the degree of topic similarity is. As shown in Figure 3, a, b are word vectors representing T_a and T_b , respectively, and θ is the angle between the two word vectors. Based on the cosine similarity, the similarity of T_a and T_b is weighted by [41]

$$\cos \theta = \frac{a \times b}{|a| \times |b|} \quad (6)$$

According to cosine principal, cosine values of the angle ranges from -1 to 1 . The value is more proximal to 1 when the angle is more proximal to 0° . Thus, topics will be more similar when the cosine value is more proximal to 1 .

4. DATA ANALYSIS AND RESULTS

4.1. The Parameter Setting of BTM

According to a previous experiment [31], this paper sets $\alpha = 50/k$ and $\beta = 0.01$. Meanwhile, two evaluations, perplexity and BTS_arg, were used to choose the best value of k and the number of iterations was set as 1000.

The perplexity and BTS_arg values are shown in Figure 4. For both datasets, when k was set to 10, their values of BTS_arg reached the bottom simultaneously; which means their topics reached the highest steady state. Although their perplexity value was still in a downward tendency, the value descended slowly. Considering that the corpus is about hypertension posts and items, too large a topic number will bring barriers in comprehension of the semantics of topics, and greatly reduce computing efficiency. Thus, in this research, 10 was selected as the optimal topic number.

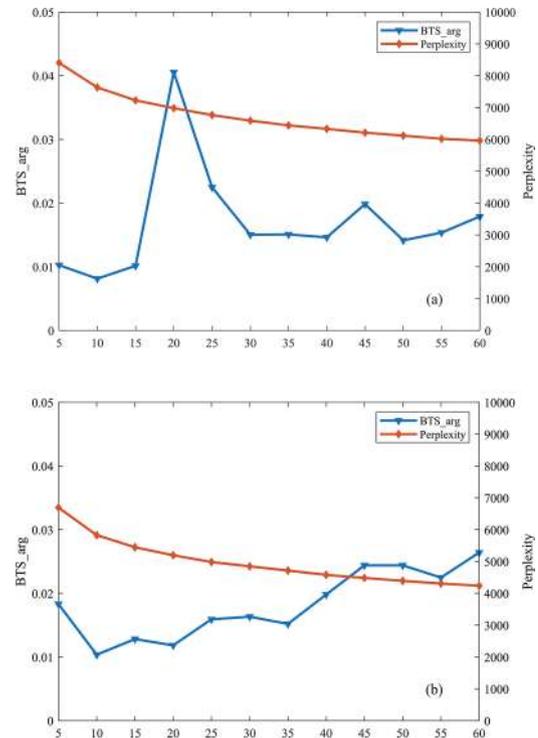


Figure 4 The perplexity and between-topic semantic similarity (BTS_arg) curves of (a) MedHelp and (b) Quora.

4.2. Topics on MedHelp

The topic-word distribution of the BTM on MedHelp dataset is displayed in Table 1. Top 10 words were listed in each topic, since too less words might not distinguish different topics, and too many words might bring understanding confusions. Topics were named according to the top 10 most likely word components, along with topic similarity degrees, and t -test results. Topic 1 mainly associated the way to manage the blood pressure—for example, how to increase or reduce the blood pressure. Terms like medication, doctor, and exercise were popular content. Queries like “Hello... I’m taking two B/P a day, what are some ways to control your B/P?” or “What is the best way to control high blood pressure?” occurred frequently under this topic. Topic 2 contained the terms about diagnosis. Many consumers did not know the basis of hypertension diagnosis and had a doubt about diagnostic results. One consumer asked “My bp was 213/109 at that point and for whatever reason the doctor that evening initially said there weren’t really anything to be done as I had no real ‘symptoms’ even though the attending nurses kept emphasizing it was the ‘silent killer.’” Topic 3 was associated with body sign; consumers tended to figure out if the sign related to hypertension. Queries included “I have a friend that gets bad headaches when stressed and he says he can feel his pressure rising.” “I’m otherwise healthy but when I get it I can feel it, I feel like a weird feeling in my chest, and sometimes pain in the left side, sort of hard to breathe but not really. I went and checked it at one of those free blood pressure things at Walmart when I was feeling bad and couldn’t sleep, and it was like only 159/80, then later I felt better and it was about 150/88.” The next topic was about related disease, the most likely words including heart, kidney, and echo. Topic 5 mainly reflected the confusion about blood pressure readings that

Table 1 | Top 10 terms in each topic in the MedHelp dataset.

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Management	Diagnosis	Body Sign	Related Disease	Time
Pressure	Day	Pressure	Test	Pressure
Blood	Med	Feel	Blood	Blood
Doctor	Doctor	Blood	Heart	Reading
Medication	Blood	Pain	Normal	Doctor
Exercise	Pressure	Day	Kidney	Day
Hypertension	Heart	Time	Doctor	Time
Heart	Start	Heart	Pressure	Monitor
Stress	Time	Symptom	Hypertension	Check
Med	Feel	Start	Echo	Arm
Diet	Normal	Doctor	Hour	Cuff
Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Medication	Dietary	Posture	Dark Thought	Lifestyle
Doctor	Food	Blood	Suffer	Avoid
Blocker	Eat	Pressure	Heart	Technique
Drug	Day	Heart	Anxiety	Relaxation
Beta	Drink	Rate	Hypertension	Salt
Blood	Blood	MmHg	Time	Exercise
Med	Exercise	Sit	Month	Practice
Calcium	Salt	Low	Father	Add
Pressure	Solidum	Bpm	Relate	Lifestyle
Medication	Diet	Stand	Ecg	Level
Day	Pressure	Pulse	History	Food

were different at different time. Consumers did not know which reading was correct when they got different ones. Queries like “I try to sit down and rest a couple of minutes before taking the readings but always big swings. Occasionally 165/90 and a couple of hours later 113/71.” or “Now an Omron wrist cuff (all cuffs make me anxious) shows numbers all over the place, normal before lunch but very high (180’s) before bed, so i do not know which one is accurate. Does anyone know which one is accurate?” were typical questions under this topic. Topic 6, linked to the medication, was chiefly composed of medicine-related terms such as blocker, beta, and calcium. Under this topic, consumers mainly asked about side effect of a certain medicine and consequences of not taking doctor’s directions—for example, stop taking medicine without authorization. Queries contained “The physician recommended for me to take 20 mg of Benicar. I’m on it for a week, but I realized after 30 mins of running, my performance went down, and a finish my 10k really tired, with difficulties to breathe, which didn’t happen before. My resting pulse rate was between 50 and 60, now with Benicar is over 70.” and “I read today that it’s very dangerous to stop it. What should I do? Should I start taking it again?” Topic 7 included dietary-related terms such as eat, drink, and salt. Consumers were curious to know the relation between food and blood pressure. One consumer asked “Has drinking hibiscus tea helped with lowering blood pressure?,” and another asked “I recently had blood work done and doctor says my sodium level is really low and that I have to start eating salt more but my blood pressure is high (the bottom number is always in the 90’s). Why would my sodium be low if I have high blood pressure?” The next topic was about the effect of different body positions on blood pressure readings. Many consumers were confused about why changing body

position had an effect on blood pressure reading. One consumer thought “I try to sit down and rest a couple of minutes before taking the readings but always big swings.” Topic 9 reflects consumers’ concern on dark thoughts of hypertension and the need of emotional support. One consumer asked “Clearly, I’m an anxious person and this is really feeding my anxieties. I’m sure this contributes, but how much?” Finally in Topic 10, consumers have concerns on lifestyle of the patients. There are many terms about workout in Topic 10, like exercise and practice. The queries showed, as to relaxation, consumers have questions about adapting their lifestyle. For example, one consumer described the question as “I’ve always had poor diet consisting of fast food, soda, snacks, fatty foods, etc. The Xanax helped me relax and stimulate my appetite, thus eating more than ever.”

4.3. Topics on Quora

Similarly, the topic-word distribution of the Quora dataset is displayed in Table 2. Topic 1 was about the management, and included some terms about how to reduce blood pressure, such as medication, exercise, diet, and control. Topic 2 terms were associated with blood pressure reading. Consumers expressed their worries in some cases, such as when finishing exercise, the readings appeared to be inaccurate or fluctuated, and they wanted to figure out if it was normal. Queries included “Can a blood pressure read high if taken a day or two after lifting weights.” Topic 3 mainly included terms about the cardiovascular system, such as vessel and artery. Under this topic, most queries discussed habits or symptoms from the perspective of the cardiovascular system, such as “Do I lay on my left

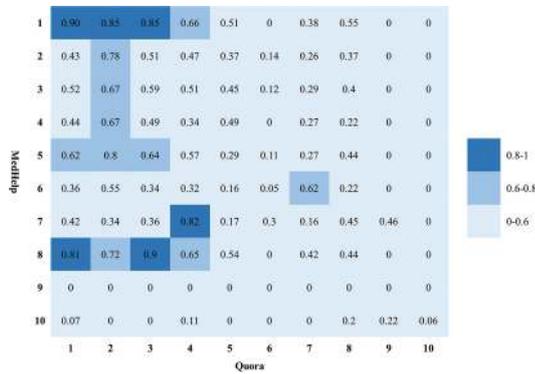


Figure 5 | Topic similarity degrees between MedHelp and Quora.

side or right side after eating to avoid indigestion.” and “Why is my heart rate in the morning around 80? I feel jittery until I eat salt and then my blood pressure is normal and then I feel okay.” The next topic was related to daily intake, because food-related terms chiefly comprised the most likely content for Topic 4, such as drink, water, salt and sodium. Many consumers wanted to figure out if items they intake influenced blood pressure, like “How does water affect high blood pressure?” or “Do bananas help with blood pressure?” Topic 5 obtained major terms about related disease of hypertension, such as stroke and kidney. Consumers were curious about which diseases were link with hypertension. Queries included “What diseases are caused by high blood pressure?” Next topic contained the terms about overweight, such as calorie, lose, and weight. There were some queries about weight influence on blood pressure, like “Can being exceptionally full from overweight actually cause a hbp?” Topic 7 was associated with medication. Drug, blocker, beta, and many medicine-related terms were included. Queries focused on side effect of medicine and principle of drug action. Consumers asked some questions like “Is it true that high blood pressure medicine can cause diabetes?” or “How does Moringa powders reduce blood pressure?” The next topic was about physical activity. Consumers sought answers to whether person with hypertension can take exercise and asked questions like “Can a person with high blood pressure exercise?,” and “Do some exercises raise blood pressure?” Topic 9 was linked to the influence of junk food. Queries included “If you exercise all the time but eat more bad food than good food what will happen?” Topic 10 related to hormone with many associated terms such as glucose, insulin, and protein. Consumers were concerned about the relationship between hormone and hypertension, such as “Is whey protein good for high blood pressure?,” or “Is it true that reducing insulin resistance reverses type II diabetes?”

4.4. Topic Similarities on MedHelp and Quora

The cosine similarity algorithm was applied to explore topic similarity and the heatmap was employed to vividly illustrate the degree of similarity. The topic similarities are shown in Figure 5.

In Figure 5, Topic 1 from the MedHelp dataset was defined as M1 and topic 1 from the Quora dataset was defined as Q1. The color spectrum represented the similarity degree of each of the two topics. The lighter color meant smaller similarity degree. The similarity thresholds were set at 0.8 and 0.6. If the similarity degree of two

topics exceeded 0.8, it meant that the two topics were very similar, and if the degree was less than 0.8 and more than 0.6, it meant that the two topics were similar. If the degree was less than 0.6 it meant that they were less similar.

According to Figure 5, M1(Management), M5(Time), and M8 (Posture) had a strong and wide similarity degree with topics generated from Quora. M1(Management) had a strong link with four topics, namely Q1(Management, 0.9), Q2(BP reading, 0.85), Q3(Cardiovascular, 0.85), and Q4(Dietary, 0.66). M8(Posture) tended to share common content with Q1(Management, 0.81), Q2(BP reading, 0.81), Q3(Cardiovascular, 0.9), and Q4(Dietary, 0.65). Three topics, Q1(Management, 0.62), Q2(BP reading, 0.8), and Q3(Cardiovascular, 0.64) were heavily related to M5(Time).

In addition, M7(Dietary) hold solely strong link with Q4(Dietary). Besides, for M2(Diagnosis), M3(body sign), and M4(Related disease), they only had association with Q2(BP reading) with similarity degrees of 0.78, 0.67, and 0.67, respectively. Meanwhile, as to M6(Medication), Q7(Medication) was the just topic that it linked with.

Among all the topics, M9(Dark thought), M10(Lifestyle), Q9(Junk food), and Q10(Hormone) were so unique that their similarity degrees with other topics almost measured 0. For the rest of topics, their topic similarity degrees with other topics were small, mostly fluctuating between 0 and 0.2.

5. DISCUSSION

For the first and second research question, this study found that on both platforms, MedHelp and Quora, hypertension management was a heated discussion topic, which was consistent with the research conducted by Mohammed [12]. Consumers on these two platforms did not know what to eat or how much to eat, such as appropriate salt intake. This finding revealed that many consumers were lacking in dietary knowledge to prevent or control hypertension [46]. Meanwhile, most of the consumers on MedHelp and Quora, expressed their confusion of medications, including drug side effects and drug withdrawal indications. Many consumers asked for help when they noticed that medication accelerated kidney failure or made them dizzy, and some stopped taking medication once their blood pressure reading got normal [47]. A related study [48] revealed that this kind of information needs may contribute to the inefficient communication between health professionals and patient as well as to a lack of continuity of follow up. Many consumers complained that they were just given a prescription and not given enough time to ask their unmet need for medication. Furthermore, because their conditions after prescription were rarely followed up, problems of medication adjustment was not perceived or settled quickly.

According to our results, the concern about how to read their blood pressure level presented in Oliveria’s study [49] has receded; how to measure blood pressure accurately became a dominant concern. Posts on MedHelp and Quora both cared about the most accurate measurement time and position when measuring blood pressure. Consumers’ focus changed from how to interpret readings to how to measure accurate indicated that most consumers could have the basic knowledge of blood pressure readings. Meanwhile, consistent with another study [48], consumers overwhelmingly thought

Table 2 | Top 10 terms in each topic in the Quora dataset

Topic 1	Topic 2	Topic 3	Topic 4	Topic 5
Management	BP Reading	Cardiovascular	Dietary	Related Disease
Blood	Pressure	Blood	Blood	Heart
Pressure	Blood	Pressure	Pressure	Disease
Hypertension	Doctor	Heart	Day	Hypertension
Medication	Test	Artery	Eat	Blood
Exercise	Time	Increase	Drink	Symptom
Diet	Med	Vessel	Water	Pressure
Lifestyle	Heart	Flow	Salt	Stroke
Control	Day	Pump	Food	Kidney
Change	Reading	Reduce	Sodium	People
Reduce	Normal	Low	Diet	Attack
Topic 6	Topic 7	Topic 8	Topic 9	Topic 10
Overweight	Medication	Physical Activity	Junk Food	Hormone
Eat	Drug	Exercise	Food	Glucose
Weight	Blocker	Pressure	Eat	Insulin
Lose	Blood	Blood	Natural	Muscle
People	Type	Week	Junk	Cell
Day	Deta	Day	Diet	Protein
Overweight	Body	Activity	Healthy	Tissue
Calorie	Inhibitor	Time	Process	Level
Time	Ace	Minute	Nutrient	Fat
Healthy	Heart	Walk	Junkie	Liver
Lot	Pressure	Cardio	Yeah	Body

blood pressure readings were influenced by external factors. They regarded the blood pressure readings as a vital principle to monitoring their condition, thus, the readings' accuracy was valued.

Moreover, hypertension-related diseases were discussed on two platforms. Among all the related diseases directly or indirectly caused by hypertension, consumers on MedHelp were concerned more about kidney diseases, while consumers on Quora seem to care more for cardiovascular system health.

For the third research question, there were some differences of consumers' information needs on hypertension between the online health community and the question and answer forum. First, consumers on MedHelp paid more attention to pathology and pharmacology of hypertension. Among the ten topics discovered from the MedHelp dataset, there were several topics about medication, causes, and diagnosis. However, consumers on Quora tended to ask questions about how to control blood pressure from a good lifestyle, such as daily recipe, overweight, and workout. Second, the proportion of posts about seeking the solutions for getting relief of anxiety on MedHelp was higher than Quora. It seems that consumers on MedHelp have more information needs on mental health.

The reason of differences might be that it was easier to construct a strong relationship in a community [50] which led consumers to describe their conditions more carefully, including more details about pathology and pharmacology of hypertension. And lastly, mental health problem caused by hypertension have drawn more and more attention from the public [50,51]. Consumers with mental health problem also needed assistant from the professionals.

6. CONCLUSIONS

This paper investigated consumer hypertension information needs on two different kinds of social media platforms, the online community and the question and answer forum. Datasets were obtained from two typical social media platforms, MedHelp and Quora. Ten topics were discovered from each dataset through topic modeling method. After employing comparative analysis on the two datasets, it was found that there were common information needs on hypertension on both platforms, such as hypertension management, recommended dietary practice, medication, blood pressure reading, and related disease. In the meantime, there were differences of consumer hypertension information needs between the two platforms. Consumers on MedHelp discussed more about pathology and pharmacology of hypertension than those on Quora. In addition, they had more information needs on the mental health of hypertension.

This study can help consumers, social media platform designers, and medical professionals better understand consumers' information needs on hypertension. Topics found can provide guidance for hypertension-related information organization and social media platform design. For instance, topic names can be used as tags to classify user-generated content on MedHelp and Quora, and then provide guidance for users to browse hypertension-related information. Recognizing of the differences of two kinds of social media platforms can lead consumers to improve the efficiency when searching information about hypertension. If the users want to search for information about pathology, pharmacology, and mental health of hypertension, MedHelp will work better; if the users

need some advises for blood pressure control, Quora will give them satisfying answers.

However, there are two major limitations that can be addressed in the future work. First, each topic was labelled by only one or two keywords, the semantics of the topic may be simplified. Topic modeling method might be mixed with content analysis methods or statistical analysis methods to improve the interpretability of the results. Second, only the online community and the question and answer forum were involved in this study. Other kinds of social media platforms may use different information flow patterns. What is consumers' information needs on hypertension on other kinds of social media platforms, such as a social networking site, still needs more research work.

DATA AVAILABILITY STATEMENT

This manuscript selected the online community MedHelp (<https://www.medhelp.org/>) and the Q&A forum Quora (<https://www.quora.com/>) as the research data sources. In order to collect research data, two web crawler programs (<https://github.com/banqunwei/Quora-Scraper-master> and <https://github.com/banqunwei/MedHelp>) were developed using Python. Then this manuscript used the BTM model to analyze the data (<https://github.com/banqunwei/BtmModel>). The data that support the findings of this study are available from the corresponding author upon request.

CONFLICTS OF INTEREST

The authors declare there is no conflicts of interest regarding the publication of this paper.

AUTHORS' CONTRIBUTIONS

Yating Li contributed to the conception of the study; Ting Dong and Qunwei Ban performed the experiment; Ye Chen and Ting Dong contributed significantly to data analysis and manuscript preparation; Yating Li helped perform the analysis with constructive discussions.

ACKNOWLEDGMENTS

This research has been made possible through the financial support of the National Natural Science Foundation of China under Grants No. 71904057 and No.71871102, and the Postdoctoral Research Foundation of China under Grants No. 2018M642884.

REFERENCES

- [1] C. Barrett, Pew internet and American life project, in: M.D. Gellman, J.R. Turner (Eds.), *Encyclopedia of Behavioral Medicine*, Springer, New York, NY, USA, 2013, pp. 1464–1465.
- [2] E.H.-J. Kim, et al., Topic-based content and sentiment analysis of Ebola virus on Twitter and in the news, *J. Inf. Sci.* 42 (2015), 763–781.
- [3] M.A. Moreno, et al., Feeling bad on Facebook: depression disclosures by college students on a social networking site, *Depress Anxiety*. 28 (2011), 447–455.
- [4] S.C. Guntuku, et al., Detecting depression and mental illness on social media: an integrative review, *Curr. Opin. Behav. Sci.* 18 (2017), 43–49.
- [5] World Health Organization, A Global Brief on Hypertension: Silent Killer, Global Public Health Crisis: World Health Day 2013, World Health Organization: Geneva 2013, pp. 1–40. https://www.who.int/cardiovascular_diseases/publications/global_brief_hypertension/en/
- [6] A.J. Viera, et al., High blood pressure knowledge among primary care patients with known hypertension: a North Carolina Family Medicine Research Network (NC-FM-RN) study, *J. Am. Board Fam. Med.* 21 (2008), 300–308.
- [7] C. Stavropoulou, Perceived information needs and non-adherence: evidence from Greek patients with hypertension, *Health Expect.* 15 (2012), 187–196.
- [8] I. Ekman, et al., Standard medication information is not enough: poor concordance of patient and nurse perceptions, *J. Adv. Nurs.* 60 (2007), 181–186.
- [9] A.M. Kaplan, M. Haenlein, Users of the world, unite! The challenges and opportunities of social media, *Bus. Horiz.* 53 (2010), 59–68.
- [10] B. Prasad, Social media, health care, and social networking, *Gastrointest. Endosc.* 77 (2013), 492–495.
- [11] T. Kear, M. Harrington, A. Bhattacharya, Partnering with patients using social media to develop a hypertension management instrument, *J. Am. Soc. Hypertens.* 9 (2015), 725–734.
- [12] M. Abdullah, K. Abdelrahman, A. Prabhu, Hypertension outside of the clinic: what are patients are discussing online in the Med-Help community?, *J. Am. Coll. Cardiol.* 73 (2019), 1114–1172.
- [13] N. Kumar, et al., Are video sharing Web sites a useful source of information on hypertension?, *J. Am. Soc. Hypertens.* 8 (2014), 481–490.
- [14] M. Al Mamun, H.M. Ibrahim, T.C. Turin, Social media in communicating health information: an analysis of Facebook groups related to hypertension, *Prev. Chronic Dis.* 12 (2015), 1–10.
- [15] W. Macias, L.S. Lewis, T.L. Smith, Health-related message board-s/chat rooms on the web: discussion content and implications for pharmaceutical sponsorships, *J. Health Commun.* 10 (2005), 209–223.
- [16] H. Park, M.S. Park, Cancer information-seeking behaviors and information needs among Korean Americans in the online community, *J. Commu. Health.* 39 (2014), 213–220.
- [17] C. Nath, et al., Website sharing in online health communities: a descriptive analysis, *J. Med. Internet Res.* 18 (2016), 1–14.
- [18] M. Zhang, C.C. Yang, Using content and network analysis to understand the social support exchange patterns and user behaviors of an online smoking cessation intervention program, *J. Assoc. Inf. Sci. Technol.* 66 (2015), 564–575.
- [19] J. Zhang, Y. Zhao, A. Dimitroff, A study on health care consumers' diabetes term usage across identified categories, *Aslib J. Inf. Manag.* 66 (2014), 443–463.
- [20] L. Bowler, et al., Eating disorder questions in Yahoo! Answers: information, conversation, or reflection?, *Proc. Am. Soc. Inf. Sci. Technol.* 49 (2012), 1–11.
- [21] L. Westbrook, Y. Zhang, Questioning strangers about critical medical decisions: “what happens if you have sex between the HPV shots?”, *Inf. Res.* 20 (2015), 1–12. <https://www.>

- researchgate.net/publication/281654127_Questioning_strangers_about_critical_medical_decisions_What_happens_if_you_have_sex_between_the_HPV_shots
- [22] Y. Zhao, *et al.*, Listening to the user's voice: a temporal analysis of autism-related questions on Quora, *Proc. Assoc. Inf. Sci. Technol.* 56 (2019), 513–516.
- [23] T. Fiona, Exploring the concept of 'information need', *Int. J. Nurs. Pract.* 12 (2006), 375–381.
- [24] L. Yang, Z. Wang, G. Hou, Discovering topics of online health community with Q-LDA model, *Data Anal. Knowl. Discov.* 3 (2019), 52–59.
- [25] Y. Lu, *et al.*, Health-related hot topic detection in online communities using text clustering, *Plos One.* 8 (2013), 1–9.
- [26] C.C. Yang, X. Tang, Estimating user influence in the MedHelp social network, *IEEE Intell. Syst.* 27 (2012), 44–50.
- [27] M. Archibald, 21 Quora statistics marketers need to know for 2020. (2021) Available from: <https://foundationinc.co/lab/quora-statistics/>
- [28] H. Jelodar, *et al.*, Latent Dirichlet Allocation (LDA) and topic modeling: models, applications, a survey, *Multimedia Tools Appl.* 78 (2019), 15169–15211.
- [29] T. Hofmann, Probabilistic latent semantic indexing, in *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, Berkeley, CA, USA, 1999, pp. 50–57.
- [30] D.M. Blei, *et al.*, Latent Dirichlet allocation, *J. Mach. Learn. Res.* 3 (2003), 993–1022.
- [31] X. Cheng, *et al.*, BTM: topic modeling over short texts, *IEEE Trans. Knowl. Data Eng.* 26 (2014), 2928–2941.
- [32] A. Asuncion, *et al.*, On smoothing and inference for topic models, in *Proceedings of the 25th Conference on Uncertainty in Artificial Intelligence (UAI Quebec, Montreal, 2009)*, 2009, pp. 27–34.
- [33] M.D. Tapi Nzali, *et al.*, What patients can tell us: topic analysis for social media on breast cancer, *J. Med. Internet Res.* 5 (2017), 1–17.
- [34] T.L. Griffiths, M. Steyvers, Finding scientific topics, *Proc. Natl. Acad. Sci. U. S. A.* 101 (2004), 5228–5235.
- [35] D. Greene, D. O'Callaghan, P. Cunningham, How many topics? Stability analysis for topic models, in: T. Calders, F. Esposito, E. Hüllermeier, R. Meo (Eds.), *Machine Learning and Knowledge Discovery in Databases*, Springer, Berlin, Heidelberg, Germany, 2014, pp. 498–513.
- [36] L. Huang, J. Ma, C. Chen, Topic detection from microblogs using T-LDA and perplexity, in *2017 24th Asia-Pacific Software Engineering Conference Workshops (APSECW)*, Nanjing, China, 2017, pp. 1–9.
- [37] J. Cao, *et al.*, A density-based method for adaptive LDA model selection, *Neurocomputing.* 72 (2009), 1775–1781.
- [38] E. Elhamifar, R. Vidal, Sparse subspace clustering: algorithm, theory, and applications, *IEEE Trans. Pattern Anal. Mach. Intell.* 35 (2013), 2765–2781.
- [39] H. Attias, A variational Bayesian framework for graphical models, in *Proceedings of the 12th International Conference on Neural Information Processing Systems*, Colorado, Denver, 1999, pp. 209–215. .
- [40] X. Tian, M. Yao, S. Gu, Forum topic mining based on BBS-LDA, *J. Zhejiang Univ. Technol.* 48 (2020), 55–62.
- [41] M. Alodadi, V.P. Janeja, Similarity in patient support forums using TF-IDF and cosine similarity metrics, in *2015 International Conference on Healthcare Informatics*, Dallas, TX, USA, 2015, pp. 521–522.
- [42] D. Gunawan, C.A. Sembiring, M.A. Budiman, The implementation of cosine similarity to calculate text relevance between two documents, in *2nd International Conference on Computing and Applied Informatics*, North-Sumatra, Medan, 2017, pp. 1–7.
- [43] F.S. Al-Anzi, D. AbuZeina, Toward an enhanced Arabic text classification using cosine similarity and latent semantic indexing, *J. King Saud Univ. Comput. Inf. Sci.* 29 (2017), 189–195.
- [44] A.R. Lahitani, A.E. Permanasari, N.A. Setiawan, Cosine similarity to determine similarity measure: study case in online essay assessment, in *2016 4th International Conference on Cyber and IT Service Management*, Bandung, Indonesia, 2016, pp. 1–6.
- [45] K.M. Oh, *et al.*, Cancer information seeking and awareness of cancer information sources among Korean Americans, *J. Cancer Educ.* 26 (2011), 355–364.
- [46] T.A. Gebrihet, *et al.*, Awareness, treatment, and control of hypertension is low among adults in Aksum town, northern Ethiopia: a sequential quantitative-qualitative study, *PloS One.* 12 (2017), 1–16.
- [47] E. Long, M. Ponder, S. Bernard, Knowledge, attitudes, and beliefs related to hypertension and hyperlipidemia self-management among African-American men living in the southeastern United States, *Patient Educ. Couns.* 100 (2017), 1000–1006.
- [48] H. Legido-Quigley, *et al.*, Patients' experiences on accessing health care services for management of hypertension in rural Bangladesh, Pakistan and Sri Lanka: a qualitative study, *PLoS One.* 14 (2019), 1–23.
- [49] S.A. Oliveria, *et al.*, Hypertension knowledge, awareness, and attitudes in a hypertensive population, *J. Gen. Internal Med.* 20 (2005), 219–225.
- [50] L. Yan, J. Peng, Y. Tan, Network dynamics: how can we find patients like us, *Inf. Syst. Res.* 26 (2015), 496–512.
- [51] I.A. Kretchy, F.T. Owusu-Daaku, S.A. Danquah, Mental health in hypertension: assessing symptoms of anxiety, depression and stress on anti-hypertensive medication adherence, *Int. J. Mental Health Syst.* 8 (2014), 1–25.