Users' Continuance Participation in the Online Peer-to-peer Healthcare Community: A Text Mining Approach

Full Papers

Yanyan Shang
Dakota State University
yshang@pluto.dsu.edu

Jun Liu
Dakota State University
Jun.liu@dsu.edu

Abstract

The online peer-to-peer healthcare communities are known as the platform where dispersed groups of patients and their families query information, seek and offer support, and connect with others. The success of such communities relies on users' ongoing involvement to generate benefits for both individuals and the communities. This study attempts to understand users' continuance participation in online peer-to-peer healthcare community by classifying users' goals of participation based on the user-generated text contents. We proposed a rule-based classification framework to categorize users' goals of posting contents into four categories: information seeking, experience sharing, information sharing, and social interaction. We formalize and test the relationship between users' continuance participation and all four posting goals, and find that the first three goals have significant impact on users' continuance participation. Our findings can help researchers and practitioners better understand users' behavior in the online peer-to-peer healthcare community.

Keywords

Online peer-to-peer healthcare community, continuance participation, text mining, rule-based classification

Introduction

The past decade has seen rapid development in health-related social media services, including patient blogs, social networking sites, and online peer-to-peer healthcare communities. The Harris poll (Taylor 2010) reported that the number of adults looking for health information on the Internet increased from 71% to 88% during the last decade. Online healthcare services appear to be a means to disseminate healthcare information, enhance communication, and facilitate a wide range of interactions between patients and healthcare delivery systems (Baker, Wagner, Singer, & Bundorf, 2003; Umefjord, Petersson, & Hamberg, 2003). One of the most promising health-related social media services is the widespread availability of online peer-to-peer healthcare communities, where people with common interests or similar health conditions gather virtually to ask questions, share experiences, and provide support, as well as exchange healthcare knowledge (Greene et al. 2011). The widespread use of such online peer-to-peer healthcare communities has dramatically changed illness management, as they are used for both a source of information and psychosocial support (Brandtzaeg et al. 2007; Johnson et al. 2006; Klemm et al. 2003; Zrebiec et al. 2001). Existing research on online peer-to-peer healthcare communities includes outcomes of support and resources for cancer survivors (Chou et al. 2011; Hesse et al. 2006), examination of drug use (Barratt et al. 2010), health effects of e-cigarette users (Alfi et al. 2013; Yamin et al. 2010), and mental health benefits (Kummervold et al. 2002) and other healthcare communities.

The success of an online peer-to-peer healthcare community depends on end-user loyalty in terms of continuance participation (Brandtzæg et al. 2008); in other words, an online community will not survive without lasting user motivation and participation. As such, it is necessary to understand the people who will use the service, the goals or tasks they have, and their context of use (Hackos et al. 1998), since the goals or tasks users have in online communities are often seen in relation to motivational issues (Brandtzæg et al. 2008). Fail to attract enough members to sustain themselves has been a primary reason that many online communities stall (Cummings et al. 2002). Motivation theory has guided researchers to study factors that inspire people to take part in an online community (Waterson 2006). Existing literature on loyalty from the perspective of motivation suggested several powerful factors: people with shared interest; experiences and needs; supportive and sociable relationships; strong social feelings of belonging; and a sense of shared identity (Diker 2004; Waterson 2006). There is also a well-developed research stream that used self-concept theory to explain the phenomenon of contribution to online communities, which includes social identity theory (Stryker 1987; Tajfel et al. 2004), self-presentation theory (Beach et al. 1990), and self-efficacy theory (Bandura 1995).

However, previous studies on users' motivation rely on survey methods to investigate users' intentions and behaviors. Studying on probability sampling from large populations, survey researches might suffer from inadequate coverage of population and data errors due to non-response or low-response. In this study, we seek to understand users' continuance participation in an online peer-to-peer healthcare community from a different perspective using text mining approaches. The text mining approaches are widely adapted in healthcare domain while facing unstructured data and tremendous data volume (Collier et al. 2008; Rebholz-Schuhmann et al. 2012). With hundreds of billions of data archived in online peer-to-peer healthcare communities, text mining is an effective way to help researchers understanding users' behaviors by discovering and extracting knowledge from the user-generated content. We propose a rule-based classification framework to mine users' goals in posting content by categorizing users' posts based on their purposes and goals. Literature on content classification in online peer-to-peer healthcare communities includes using text cluster methods to detect hot topics (Hong et al. 2004) and to recommend education materials for diabetes patients (Lee et al. 2005). Different from previous studies, our research classifies user-generated content based on the purposes and goals of users' initial posts in an online peer-to-peer healthcare community and predicts users' continuance participation by examining and analyzing their posts' contents. From the initial post in discussion threads, the users' purposes and goals can be categorized into four different categories: information seeking, experience sharing, information sharing, and social interaction. We find that the posts on information seeking, experience sharing, and information sharing have significant effect on how long the user will stay in an online peer-to-peer healthcare community.

The remainder of this paper is organized as follows. In the next section, we propose a theoretical background to explain how online communities evolve in ways that are more or less sustainable and propose our hypotheses based on the theoretical basis. In Section 3, we design the goal mining approach to classify user-generated text contents and demonstrate the proposed framework. A preliminary case study is conducted in Section 4. We then discuss the findings regarding users' behavior in online peer-to-peer healthcare communities. The last section discusses future research directions.

Theoretical Background

Users' Continuance Participation

With the purpose of exchanging information and knowledge and providing support, online communities rely on members' participation and contributions in terms of user-generated content to generate benefits for each other and for the community as a whole (Cummings et al. 2002; Ling et al. 2005). Based on the social identity theory (Lazar et al. 2002), the uses and gratifications theory (Phang et al. 2009), and other motivation theories, previous studies have largely sought to explore what and how to attract people to participate in online communities (Cummings et al. 2002; Waterson 2006). However, the prosperity of online communities depends on users' ongoing participation rather than initial acceptance. As a matter of fact, many initially active online communities suffer from a retention problem (Ridings et al. 2004). Without lasting user motivation and ongoing user participation, online communities will dissolve into nothing more than random interactions (Faraj et al. 2011).

A range of literature has touched on the issue of communities' sustainability, suggesting that online communities provide benefits and experiences that member seek in order to gain end-user loyalty (Brandtzæg et al. 2008; Ridings et al. 2004). Researchers have proposed rich descriptions of design features to increase members' likelihood of joining and remaining in online communities, for instance, (Lazar et al. 2002; Ling et al. 2005; Phang et al. 2009). These studies provide rich insights into online community design and management, but neglect the role of members' individual characteristics and goals and how these will affect their decisions on continuing participation.

Some studies have made solid theoretical contributions to the literature by investigating online communities' phenomena from an individual level of analysis. These studies suggested that the reasons individuals participate in online communities include being attracted by community benefits (Ridings et al. 2004), a sense of reciprocity (Hall et al. 2004; Wasko et al. 2000), and a desire to help the community (Constant et al. 1996; Lakhani et al. 2003). However, these studies mainly focused on personal utilitarian motivations of knowledge sharing (Hall et al. 2004; Wasko et al. 2000) but neglected the hedonic factors that may be important in the context of online communities (Faraj et al. 2011).

This research gap is particularly crucial when we are looking at health-related communities, given the healthcare-domain designated features. For example, in addition to information seeking for the purpose of self-education and self-management, patients are more willing to share their own experiences of living and fighting with certain diseases with the goals of obtaining sympathy and emotional support from others. Based on users' goals and purposes, their text posts can be categorized into four categories: information seeking posts, experience sharing posts, information sharing posts, and social interaction posts. In the framework design section, we will introduce how we collect data and categorize them into different categories.

Utilitarian and Hedonic Motivations

Originating from consumer behavior theory, utilitarian and hedonic motivations (Laurent et al. 1985; Park et al. 1986) are two dimensions of an individual's overall perceived value that can drive specific outcome behavior. The term utilitarian is more task-oriented in nature, whereas hedonic is related to entertainment, fun-seeking, and other emotional desire behavior (Constant et al. 1996). Recently, researchers have addressed the role of both utilitarian and hedonic values in the study of online service usage (Cotte et al. 2006; Hong et al. 2004; Lee et al. 2005; Lee et al. 2012). Utilitarian value (purposive value) is objective and task-focused with the goal of achieving a pre-determined task, which in online peer-to-peer healthcare communities refers to the perceived usefulness of obtaining required information and knowledge. The technology acceptance model (TAM) is frequently adapted by scholars to study computer-technology acceptance in information systems disciplines (Taylor et al. 1995). The TAM takes forwards the idea that an individual's behavioral intention can be predicted from two factors: perceived ease-of-use and perceived usefulness. Both perceived ease-of-use and perceived usefulness jointly affect attitude, while perceived ease-of-use has a direct impact on perceived usefulness. Perceived usefulness thereby has a direct, driving impact on users' initial acceptance Information seeking posts are primarily looking for answers to issues they had/experienced or asking questions on general knowledge for self-management. This kind of post aims at finding information and answers. It is the task-oriented function which is related to utilitarian values. An example of an information seeking post follows:

> *"Hello there, I'm type-1 diabetes, but my brother was lately found out to be type-2 (or perhaps LADA)... anyway, I was wondering is there a need for carb counting in type-2 diabetes? how does it done and what-s the difference? I do understand and do my best carb-counting for insulin-caculating purposes, but what is the benefits for type-2's? thanks alon(t)"*

Therefore, we suggest the goal of information seeking initially motivates users join an online community. To investigate why users continue using an information system, the IS continuance model proposed by Bhattacherjee (2001) goes beyond prior IS adoption models to include the users' post-adoption psychological motivations, for example, satisfaction and confirmation. The IS continuance model was validated empirically by studies in different fields (e.g., online banking users, online communities users), suggesting that users' satisfaction with prior IS use was the strongest predictor of continuance behavior. Without considering satisfaction, the perceived usefulness only has a weaker effect on the users' continuance participation (Jin et al. 2010). In an online peer-to-peer healthcare community, users with information seeking goals initially post questions when they join the communities. However, they usually don't keep engaging with others because they are focusing on the information they requested and the answers they received. Their interest in the community may not last if they cannot get proper answers, or once they have the correct answers. As time passes by, information seeking users' may lose their interests of continuance participation. Additionally, Van der Heijden (2004) suggests that with the effect of the hedonic nature of an information system, perceived of usefulness loses its dominant predictive value in favor of perceived ease of use and enjoyment. As such, we propose hypothesis 1:

H1: The degree to which a user's posts are motivated by information seeking likely has a negative relationship with the user's continuance participation.

Consumers are often faced with choices between utilitarian and hedonic alternatives that are at least partly driven by emotional desires rather than cognitive deliberations (Hsu et al. 2004). Venkatesh et al. (2002) found that along with utilitarian value, hedonic value is also an important value that users want to fulfil through using online communities. Unlike utilitarian value, which is gratified by task achievement, hedonic value (Zhang 2013) focuses on the emotional desires fulfilled through participation, including enjoyment and comfort. Venkatesh et al. (2002) redefined the TAM by adapting the emotional element and added perceived enjoyment as an intrinsic motivation to explain technology acceptance. Lee et al. (2005) also adapted the hedonic value in terms of perceived enjoyment into the TAM, arguing that users' behavior is evoked from feelings of pleasure, joy, and fun. In an online peer-to-peer healthcare community, experience sharing posts focus on sharing personal experiences or calling for others' experiences with the topic, such as on illness management. In this kind of post, users are more interested in discussing personal experiences and feelings than finding solutions. An example of experience sharing follows:

> *"Struggled yesterday with highs on my Paradigm - ended up changing cartridge, set and tubing 3 times and finally resolved overnight. It's days like this that I appreciate all the more my medical team and the support from Minimed. After speaking to Dr late in the day took a manual injection for evening meal and called Minimed 24 hour support. Rep there ran tests on my pump and the infusion set and found it to be operational. Rep suggested one of 3 issues: (1) onset of an illness (not true so far), (2) scar tissue at set site or (3) a suggestion that I use an infusion set with longer tubing for better delivery. I'm quite perplexed how longer tubing will enhance delivery - would logically seem to be just the opposite."*

We believe feeling comfort and sympathy can be a strong emotional driver that impacts behavior. As such, we propose hypothesis 2:

H2: The degree to which a user's posts are motivated by experience sharing likely has a positive relationship with the user's continuance participation.

Information sharing posts have the primary goal of sharing news, information, new studies, or developments in a certain disease. Posts classified to this category usually have an external link or book name showing the sources of information. An example of an information sharing post:

> *"Maybe you've already seen this, but my brother sent me the article today, and I just found it on the Internet. I knew that Dr. Faustman has been doing important work on curing diabetes, but didn't know about this. http://www.thenewstribune.com/2011/06/25/1720244/tb-drug-might-help…"*

Some researchers believe information sharing posts are task-related in light of the information exchange function. However, information sharing activity is different from information seeking activity. While information seeking posts expect answers and solutions to the posted questions, information sharing users seek to help others and obtain gratitude and recognition. Based on Maslow's hierarchy of human needs (Maslow et al. 1970) and Self-Determination Theory (Deci et al. 1985), human seek to control outcomes and be affirmed by others. When users share information in online peer-to-peer healthcare communities, they feel they have information and knowledge that can benefit both themselves and others. Satisfied by the sharing activity itself, the users have more lasting motivation to participate in the online peer-to-peer healthcare community. As such, we propose hypothesis 3:

H3: The degree to which a user's posts are motivated by information sharing likely has a positive relationship with the user's continuance participation.

Social interaction posts discuss non-health-related topics, including greetings, chat, and other contents with no purposive value but to build a friendly environment in the online peer-to-peer healthcare community. Users' purpose for posting social interaction posts usually has nothing to do with obtaining information and knowledge to manage their illness or expressing anxiety to get emotional support. The social interaction posts goal is hedonic value since it is gratified by the engagement itself. An example of a social interaction post:

> *"I just started reading a Jane Austen w(h)ich is Pride and Prejudice. I just wanted to see how many people like her. So for i like the book. I have also seen the movie Emma_based on the Jane Austen book, and the movie Clueless is an updated version of Emma. I bet the person who wrote Clueless was a Jane Austen fan."*

The Self-Determination Theory (Deci et al. 1985) suggests three basic psychological needs that must be satisfied to foster well-being and health: competence, relatedness, and autonomy. Relatedness refers to an individual's tendency to want to interact with and experience caring for other people. SDT suggests that intrinsic motivation will tend to flourish in contexts characterized by a sense of relatedness (Deci et al. 2000). In an online peer-to-peer healthcare community, social interaction posts focus on fun and enjoyment, and can lead to long-term relationships between users, thereby, providing/proving to be a positive driver on users' continuance participation in the community. As such, we propose hypothesis 4:

H4: The degree to which a user's posts are motivated by social interaction likely has a positive relationship with the user's' continuance participation.

Methodology and Framework Design

Text Mining

Text mining methods aid researchers in coping with information overload (unstructured data) problems by using techniques from data mining, machine learning, natural language processing (NLP), information retrieval (IR), and knowledge management (Feldman et al. 2007). Text mining operates at a finer level of granularity than information retrieval (IR) and text summarization (TS) and examines the relationships between specific kinds of information contained both within and between documents (Cohen et al. 2005). Text mining applications are found in different fields, including monitoring and analysis of online text sources for security breaches (Gegick et al. 2010), analyzing customer relationships in business management (Coussement et al. 2008), predicting stock prices based on text documents (Mittermayer 2004), as well as detecting emotions in the field of affective computing (AC) (Calvo et al. 2010). One promising use of text mining is adapted by biomedical research, which aids researchers and practitioners in discovering knowledge and putting it to practical use in the forms of diagnosis, prevention, and treatment (Cohen et al. 2005). For example, classification methods such as decision trees and machine learning are used to classify cancer using gene-expression data (Lakhani et al. 2003; Statnikov et al. 2005), tumors of the tongue (Schwarzer 2004), and the deferral of rectal examination in trauma patients (Guldner et al. 2004). Collier et al. (2008) designed an ontology-based text mining system for detecting and tracking the distribution of infectious disease outbreaks from linguistic signals on the web. Rebholz-Schuhmann et al. (2012) proposed a text mining solution for automated literature analysis with the purpose of extracting knowledge from biomedical information. In this study, we propose a text mining processing framework to analyze user-generated content with the purpose of discovering the users' continuance participation based on their posts in an online peer-to-peer healthcare community. Traditional studies on users' intention and behavior are normally based on user surveys Our study, however, aims to understand users' behavior by analyzing user-generated contents on an online healthcare website site rather than surveying users.

Users' Goals Mining Framework

Figure 1 illustrates the framework for classifying user-generated text content in online peer-to-peer healthcare communities based on their goals for posting. There are three modules in the framework: data collection, data pre-processing, and users' goals mining. Each module has two sub-modules.

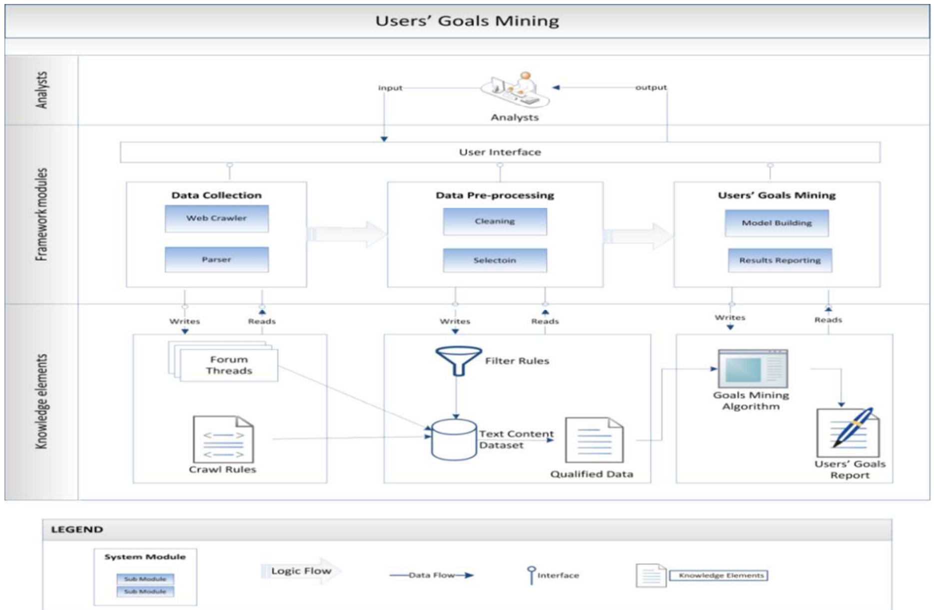

Figure 1. Users' Goals Mining Framework

In the data collection module, we build a web crawler using Python programming language to collect the user-generated text content thread by thread from web pages, as well as user profiles that are provided by users when they register Several parsers have been used in the web crawling process in order to retrieve the data we needed, such as HTML parser, DateTime parser, and so on. In order to avoid overloading the website, we set the program to sleep for 30 seconds for every 100 pages retrieved.

In the data pre-processing module, we first clean data by removing unnecessary punctuation and symbols that may impede data format and data processing. Then we removed meaningless words in the text content to prepare the text corpus for mining. Additionally, we developed a set of filtering rules to rule out the data that does not match the requirement of our study (e.g. users who didn't register in the study period; users who only had reply posts but no initial posts).

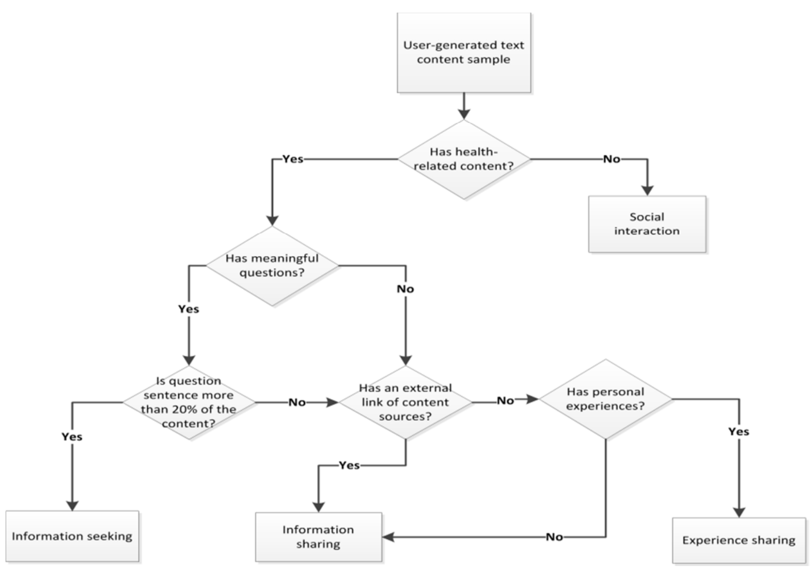

Figure 2. The Rule-based Classifier for User-generated Content

In the users' goals mining module, we design an algorithm to classify the users' posting goals into four different categories: information seeking, experience sharing, information sharing, and social interaction. A rule-based classifier consists of a set of rules, used in a given order during the prediction process, to classify unlabeled objects (Giacometti et al. 2008). Rule-based classification methods have been used in biomedical field for assisting diagnosis (Li et al. 2008; Zaıane et al. 2002) and detecting diseases (Solt et al. 2009).

```
Algorithm: The Users' Goals Mining Algorithm
Input: t ∈ T is a set of user-generated content documents.
Input: r ∈ R is a set of classification rules.
Output: g ∈ G is a matrix of categories indicating users' goals for user-generated contents.
//Initialize
 g = ∅
i, j, k = 0
 //Classify user-generated contents
For t in T:
  Label t_j by the owner of the document u_i
  For r in R:
    If ( t_j matches r_k):
      g_{i,k} = (u_i, r_k)
//Categorize users' goals
For user, category in group by user:
    Count for each category c_k ∈ C
    g_i = (u_i, C)
Return G
```

Figure 3. The Users' Goals Mining Algorithm

In the proposed users' goals mining algorithm, we use rule-based classifier to classify users' goals. A close scrutiny of the posts indicates that many posts that include question sentences or question marks show no intention of seeking information but focus on expressing feelings or looking for support or recognition. These posts often include questions such as: "How could this happen to me?", "Do you feel bad with…?", etc. There are also question sentences that appear in the narratives when users try to share information or experiences. We hence set up a threshold of 20% (i.e., only posts with over 20% of question sentences would be classified as information seeking), based on investigating the manually classified sample data. Figure 2 shows the rule-based classifier and Figure 3 shows the pseudo code for users' goals mining Algorithm.

Evaluation: A Preliminary Case Study

Of all the health information searched online, diabetes is one of the most common searched disease-related topics. In 2013, it was estimated that over 382 million people throughout the world had diabetes (Melmed et al. 2011). As a chronic illness, diabetes requires continuing medical care and patient self-management education to prevent acute complications and to reduce the risk of long-term complications (AmericanDiabetesAssociation 2009). Online diabetes peer-to-peer communities can help patients with self-education and psychosocial support. Previous studies on online diabetes peer-to-peer communities focused on topic discovery and categorization (Franklin et al. 2008; Ravert et al. 2003), and health outcome benefits (Bond et al. 2006; Franklin et al. 2008; Savolainen 2011) . However, to our knowledge, no study on online diabetes peer-to-peer communities focuses on users' continuance participation by mining user-generated content based on their purposes and goals.

Data collection

We collected data from a website called tudiabetes.org. It is an online peer-to-peer healthcare community funded by the Diabetes Hands Foundation, which is aimed at providing a platform for people who have diabetes to get in touch with others, help each other out, educate themselves and share the steps they take every day to stay healthy while living with this very serious condition. We collected both structured data and unstructured data from user profiles and the discussion threads from the webpage using a Python web crawler. Up to August 2014, there were 22,304 discussion threads and a total of 244,511 discussion posts in the forum. On average, there were 10 reply posts for each initial post. There were 32,299 users in the user profile dataset, and 9,939 of them had posted in at least one discussion post. On average, the users' participation time spanned 181 days. Table 1 summarizes the users' participation time span.

N	Mean	Std Dev	Minimum	Maximum
9939	181.21	329.07	0	1933.00

Table 1. Summary of Users' Participation Time Span

Operationalization

Description of variables

Users' continuance participation is measured by the time span that a user has been active in this online diabetes peer-to-peer community. We calculate it by counting the number of days from a user's first post to the last post in the discussion forum.

User-generated content in terms of thread posts are classified into four different categories: information seeking, experience sharing, information sharing, and social interaction. Table 2 describes the analysis variables.

Dependent Variable	Users' continuance participation	➢ Users' stay time length (in days, Standardized to [0,1])
Independent Variables	Users' goals of posting user-generated content	➢ Users' goals category 1: information seeking ➢ Users' goals category 2: experience sharing ➢ Users' goals category 3: information sharing ➢ Users' goals category 4: social interaction (Standardized to [0,1])

Table 2. Description of Variables

For the empirical study we select a 2-year study period that includes users who issued their first post between January 1, 2011 and December 31, 2012. Our study aims to understand users' goals of posting based on the content analysis. We believe the initial post in each thread will more clearly indicate the users' purposes or goals than the reply posts, since most reply posts don't initiate a topic but follow a topic. The reply posts can imply the users' interests and the way they like to participate in a discussion, but they are hardly purpose-oriented. Therefore we focus our text analysis on the initial post.

The descriptive statistics on the data collected from Tudiabetes.org shows users' average continuance participation time span is 6 months. Closer investigation of the dataset showed that most users are more active in the first 6 months. As such, we process each user's content for the first 6 months of their membership using the aforementioned algorithm. To validate the classification results, two colleagues manually read through the text content of the sample data to make sure the classification results are consistent. Figure 4 summarizes the classification results of users' goals, and table 3 shows descriptive statistics of sampling dataset.

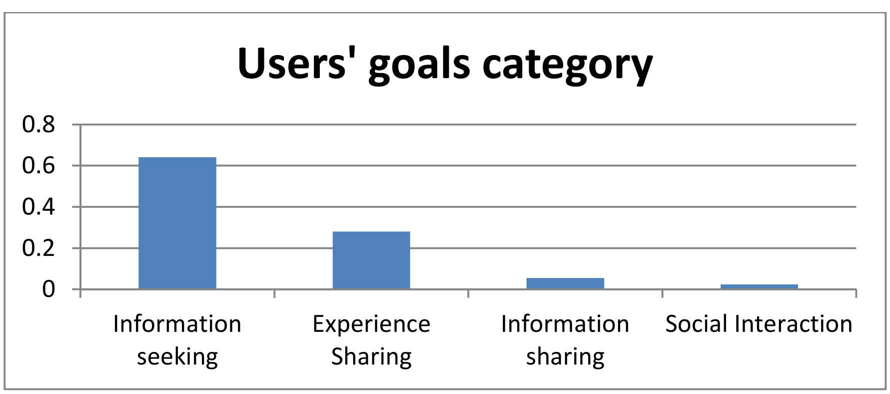

Figure 4. Users' Posting Goals

Descriptive Statistics						
Variable	N	Mean	Std Dev	Sum	Minimum	Maximum
Information Sharing	565	0.05439	0.20193	30.73000	0	1.00000
Information Seeking	565	0.64150	0.41594	362.45000	0	1.00000
Experience Sharing	565	0.27991	0.38481	158.15000	0	1.00000
Social Interaction	565	0.02448	0.12805	13.83000	0	1.00000
ContnParticipationInDays	565	234.85310	315.86988	132692	1.00000	1361

Table 3. Descriptive Statistics of Sampling Dataset

Results

A between subjects multivariate correlation was used to compare the effects of users' continuance participation and users' goals of posting content. Table 3 shows the correlation analysis results from the SAS Enterprise Guide.

Pearson Correlation Coefficients, N = 565 Prob > |r| under H0: Rho=0	
	ContnParticipationInDays
Information Seeking	-0.14972
Standardized Information Seeking: mean = 0 standard deviation = 1	0.0004
Experience Sharing	0.12627
Standardized Experience Sharing: mean = 0 standard deviation = 1	0.0026
Information Sharing	0.10157
Standardized Information Sharing: mean = 0 standard deviation = 1	0.0157
Social Interaction	-0.05500
Standardized Social Interaction: mean = 0 standard deviation = 1	0.1917

Table 4. Correlation Analysis Results

The correlation analysis shows the users' continuance participation in the online diabetes peer-to-peer community is negatively related with their goal of information seeking ($p<0.001$) and positively related with their goal of experience sharing ($p<0.005$) and information sharing ($p<0.05$). However, the results show a non-significant negative relationship between users' continuance participation and social interaction, which rejects Hypothesis 4. Table 4 lists the summary of the 4 hypothesis tests.

Hypotheses	Support
H1: Information Seeking → Continuance Participation	YES
H2: Experience Sharing → Continuance Participation	YES
H3: Information Sharing → Continuance Participation	YES
H4: Social Interaction → Continuance Participation	NO

Table 5. Summary of Hypothesis Tests

Discussion and Implications

Our results suggest that the users whose primary goal of participating in the online diabetes community is information seeking most likely lose interest in continuance participation in the long term. This is slightly different from previous studies on online communities which believe that task-oriented activity should have a positive relationship with user acceptance. We believe this is reasonable for the following reasons. First, previous studies verified that perceived usefulness (Lee et al. 2005) is positively related to users' behavior intention. In conjunction with this, we argue that the perceived usefulness has a strong impact on users' initial intention of acceptance, yet has a very weak impact on users' continuance participation in the long term. As many other studies proposed (Bhattacherjee 2001; Jin et al. 2010), if the users' expectation has been satisfied, they will lose motivation to continue. Secondly, if users get expected answers from replies to information seeking posts, they may likely come back to ask other questions. However, how many questions will they have? What happens when users have asked all their questions? On the other hand, if the users' goal focuses on asking questions and finding solutions, it would be easier and faster to search for relevant questions. One study (Nonnecke et al. 2004) showed that information seeking users usually read but seldom post. Nonnecke et al. (2004) posited that the information seeking goal can be achieved in the form of easily searched and browsable archives and other online informative resources such as FAQs. Therefore, information seeking can be an initial motivation to join an online peer-to-peer healthcare community, but is not a lasting driver.

Those looking for emotional support are most likely to continue in the online peer-to-peer healthcare community. Users who are motivated by utilitarian value participate in the online community for the sake of informative benefits. Users who are hedonically motivated typically enjoy participating in online communities for the sake of participation itself (Cotte et al. 2006). The users who like to share personal experiences usually feel a sense of connection and belonging with other community members. Commitment is a psychological bond that characterizes an individual's relationship with an organization (Wykes 1998). Bateman et al. (2011) adapted commitment theory in an online community setting and identified three types of commitment. Affective commitment is the bond between a member and a particular community that is based on the member's strong emotional attachment to that community. With this bond, members are more willing to share personal experiences and feel happy to stay in the community to support each other.

Hypothesis 3 finds that users' goal of information sharing is also positively related with users' continuance participation. Information sharing can be driven by a sense of reciprocity (Hall et al. 2004) or altruism (Lakhani et al. 2003; Wasko et al. 2005). A rich set of existing research in online communities suggest that individuals are motivated intrinsically to share information and knowledge with others because engaging in intellectual pursuits is challenging and fun, and also because they enjoy helping others (Wasko et al. 2000; Wasko et al. 2005).

Our results reject hypothesis 4, that there is a positive relationship between users' continuance participation and social interaction. This is probably because our sample dataset is not balanced. As shown in Figure 4, only 2% of the content was classified as social interaction. With an excessively low proportion, it is very hard to find a significant relationship. Additionally, Tudiabetes.org is a peer-to-peer community rather than a social networking site. Members are attracted to the community by their expectation of congruency (Ridings et al. 2004) between their personal interests and the community's topics, goals, and activities, which in this case are self-management of diabetes and emotional support. Maintaining a social network and online relationships wasn't an important goal in this online peer-to-peer community.

Conclusion

There are several benefits of online peer-to-peer healthcare communities that attract patients. Online peer-to-peer healthcare communities are structured around a central theme (a certain disease) and can function as a meeting place for people who are at least to a certain extent interested in or familiar with the topic and may have more understanding of what someone is going through than those in one's offline social network. We design a users' goals mining framework to analyze user-generated text contents and formalize 4 hypotheses to investigate the relationship between users' continuance participation and their posting goals. Our findings help us to understanding users' intention and behavior in an online peer-to-peer healthcare community.

Different from traditional users' intention and behavior research on online communities, our study uses text mining techniques to discover users' goals instead of conducting a survey. Survey methodology is frequently used in studies that investigate individuals' opinions and behaviors. However, it normally suffers from the problem of low response rate. With the tremendous data we can get from Internet archives, texting mining methods help us discover knowledge from a large dataset or the whole population.

In this study, we only consider the impact of user-generated text contents on users' continuance participation in an online peer-to-peer healthcare community. However, there are other factors that may have influence. For example, a preliminary study shows that users' diagnosis date and their short-term activities both impact users' continuance participation. Future study will take these factors into account and try to find in-depth reasons that keep users' active in online peer-to-peer healthcare communities.

REFERENCES

Alfi, M., and Talbot, P. 2013. "Health-related effects reported by electronic cigarette users in online forums," *Journal of medical Internet research* (15:4).

AmericanDiabetesAssociation 2009. "Standards of medical care in diabetes—2009," *Diabetes care* (32:Suppl 1), p S13.

Bandura, A. 1995. *Self-efficacy in changing societies*, Cambridge university press.

Barratt, M. J., and Lenton, S. 2010. "Beyond recruitment? Participatory online research with people who use drugs," *International Journal of Internet Research Ethics* (3:1), pp 69-86.

Bateman, P. J., Gray, P. H., and Butler, B. S. 2011. "Research Note-The Impact of Community Commitment on Participation in Online Communities," *Information Systems Research* (22:4), pp 841-854.

Beach, L. R., and Mitchell, T. R. 1990. "Image theory-a behavioral-theory of decision-making in organizations," *Research in organizational behavior* (12), pp 1-41.

Bhattacherjee, A. 2001. "Understanding information systems continuance: an expectation-confirmation model," *MIS quarterly*), pp 351-370.

Bond, G. E., Burr, R., Wolf, F. M., Price, M., McCurry, S. M., and Teri, L. 2006. "Preliminary findings of the effects of comorbidities on a web-based intervention on self-reported blood sugar readings among adults age 60 and older with diabetes," *Telemedicine Journal & E-Health* (12:6), pp 707-710.

Brandtzæg, P. B., and Heim, J. Year. "User loyalty and online communities: why members of online communities are not faithful," Proceedings of the 2nd international conference on INtelligent TEchnologies for interactive enterTAINment, ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering)2008, p. 11.

Brandtzaeg, P. B., and Heim, J. 2007. "Initial context, user and social requirements for the Citizen Media applications: Participation and motivations in off-and online communities," *Citizen Media Project*.

Calvo, R. A., and D'Mello, S. 2010. "Affect detection: An interdisciplinary review of models, methods, and their applications," *Affective Computing, IEEE Transactions on* (1:1), pp 18-37.

Chou, W.-y. S., Liu, B., Post, S., and Hesse, B. 2011. "Health-related Internet use among cancer survivors: data from the Health Information National Trends Survey, 2003–2008," *Journal of Cancer Survivorship* (5:3), pp 263-270.

Cohen, A. M., and Hersh, W. R. 2005. "A survey of current work in biomedical text mining," *Briefings in bioinformatics* (6:1), pp 57-71.

Collier, N., Doan, S., Kawazoe, A., Goodwin, R. M., Conway, M., Tateno, Y., Ngo, Q.-H., Dien, D., Kawtrakul, A., and Takeuchi, K. 2008. "BioCaster: detecting public health rumors with a Web-based text mining system," *Bioinformatics* (24:24), pp 2940-2941.

Constant, D., Sproull, L., and Kiesler, S. 1996. "The kindness of strangers: The usefulness of electronic weak ties for technical advice," *Organization science* (7:2), pp 119-135.

Cotte, J., Chowdhury, T. G., Ratneshwar, S., and Ricci, L. M. 2006. "Pleasure or utility? Time planning style and web usage behaviors," *Journal of Interactive Marketing* (20:1), pp 45-57.

Coussement, K., and Van den Poel, D. 2008. "Integrating the voice of customers through call center emails into a decision support system for churn prediction," *Information & Management* (45:3), pp 164-174.

Cummings, J. N., Butler, B., and Kraut, R. 2002. "The quality of online social relationships," *Communications of the ACM* (45:7), pp 103-108.

Deci, E. L., and Ryan, R. M. 1985. *Intrinsic motivation and self-determination in human behavior*, (Springer Science & Business Media.

Deci, E. L., and Ryan, R. M. 2000. "The" what" and" why" of goal pursuits: Human needs and the self-determination of behavior," *Psychological inquiry* (11:4), pp 227-268.

Diker, V. 2004. "A Dynamic Feedback Framework for Studying Growth Policies in Open Online Collaboration Communities," *AMCIS 2004 Proceedings* (2004), p 328.

Faraj, S., and Johnson, S. L. 2011. "Network exchange patterns in online communities," *Organization Science* (22:6), pp 1464-1480.

Feldman, R., and Sanger, J. 2007. *The text mining handbook: advanced approaches in analyzing unstructured data*, (Cambridge University Press.

Franklin, V. L., Greene, A., Waller, A., Greene, S. A., and Pagliari, C. 2008. "Patients' engagement with "Sweet Talk"–a text messaging support system for young people with diabetes," *Journal of Medical Internet Research* (10:2).

Gegick, M., Rotella, P., and Xie, T. Year. "Identifying security bug reports via text mining: An industrial case study," Mining Software Repositories (MSR), 2010 7th IEEE Working Conference on, IEEE2010, pp. 11-20.

Giacometti, A., Khanjari Miyaneh, E., Marcel, P., and Soulet, A. Year. "A generic framework for rule-based classification," Proceedings of LeGo2008, pp. 37-54.

Greene, J. A., Choudhry, N. K., Kilabuk, E., and Shrank, W. H. 2011. "Online social networking by patients with diabetes: a qualitative evaluation of communication with Facebook," *Journal of general internal medicine* (26:3), pp 287-292.

Guldner, G., Babbitt, J., Boulton, M., O'Callaghan, T., Feleke, R., and Hargrove, J. 2004. "Deferral of the rectal examination in blunt trauma patients: a clinical decision rule," *Academic emergency medicine* (11:6), pp 635-641.

Hackos, J. T., and Redish, J. 1998. *User and task analysis for interface design*, (Wiley Computer Publishing: New York.

Hall, H., and Graham, D. 2004. "Creation and recreation: motivating collaboration to generate knowledge capital in online communities," *International Journal of Information Management* (24:3), pp 235-246.

Hesse, B. W., Moser, R. P., Rutten, L. J. F., and Kreps, G. L. 2006. "The health information national trends survey: research from the baseline," *Journal of Health Communication* (11:S1), pp vii-xvi.

Hong, S., and Kim, J. 2004. "Architectural criteria for website evaluation–conceptual framework and empirical validation," *Behaviour & Information Technology* (23:5), pp 337-357.

Hsu, M.-H., and Chiu, C.-M. 2004. "Predicting electronic service continuance with a decomposed theory of planned behaviour," *Behaviour & Information Technology* (23:5), pp 359-373.

Jin, X.-L., Lee, M. K., and Cheung, C. M. 2010. "Predicting continuance in online communities: model development and empirical test," *Behaviour & Information Technology* (29:4), pp 383-394.

Johnson, G. J., and Ambrose, P. J. 2006. "Neo-tribes: the power and potential of online communities in health care," *Communications of the ACM* (49:1), pp 107-113.

Klemm, P., Bunnell, D., Cullen, M., Soneji, R., Gibbons, P., and Holecek, A. 2003. "Online cancer support groups: a review of the research literature," *Computers Informatics Nursing* (21:3), pp 136-142.

Kummervold, P. E., Gammon, D., Bergvik, S., Johnsen, J.-A. K., Hasvold, T., and Rosenvinge, J. H. 2002. "Social support in a wired world: use of online mental health forums in Norway," *Nordic journal of psychiatry* (56:1), pp 59-65.

Lakhani, K. R., and Von Hippel, E. 2003. "How open source software works:"free" user-to-user assistance," *Research policy* (32:6), pp 923-943.

Laurent, G., and Kapferer, J.-N. 1985. "Measuring consumer involvement profiles," *Journal of marketing research* (22:1), pp 41-53.

Lazar, J., and Preece, J. 2002. *Social considerations in online communities: Usability, sociability, and success factors*, (Lawrence Erlbaum Associates Inc. Publishers: Mahwah: NJ.

Lee, M. K., Cheung, C. M., and Chen, Z. 2005. "Acceptance of Internet-based learning medium: the role of extrinsic and intrinsic motivation," *Information & management* (42:8), pp 1095-1104.

Lee, Y., Chen, A. N., and Ilie, V. 2012. "Can Online Wait Be Managed? The Effect of Filler Interfaces and Presentation Modes on Perceived Waiting Time Online," *MIS Quarterly* (36:2), pp 365-394.

Li, Q., Li, F., and Doi, K. 2008. "Computerized detection of lung nodules in thin-section CT images by use of selective enhancement filters and an automated rule-based classifier," *Academic radiology* (15:2), pp 165-175.

Ling, K., Beenen, G., Ludford, P., Wang, X., Chang, K., Li, X., Cosley, D., Frankowski, D., Terveen, L., and Rashid, A. M. 2005. "Using social psychology to motivate contributions to online communities," *Journal of Computer-Mediated Communication* (10:4), pp 00-00.

Maslow, A. H., Frager, R., Fadiman, J., McReynolds, C., and Cox, R. 1970. *Motivation and personality*, (Harper & Row New York.

Melmed, S., and Williams, R. H. 2011. *Williams textbook of endocrinology*, (12th ed.) Elsevier/Saunders: Philadelphia.

Mittermayer, M.-A. Year. "Forecasting intraday stock price trends with text mining techniques," System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on, IEEE2004, p. 10 pp.

Nonnecke, B., Preece, J., and Andrews, D. Year. "What lurkers and posters think of each other [online community]," System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on, IEEE2004, p. 9 pp.

Park, C. W., and Young, S. M. 1986. "Consumer response to television commercials: The impact of involvement and background music on brand attitude formation," *Journal of marketing research* (23:1), pp 11-24.
Phang, C. W., Kankanhalli, A., and Sabherwal, R. 2009. "Usability and sociability in online communities: A comparative study of knowledge seeking and contribution," *Journal of the Association for Information Systems* (10:10), p 2.
Ravert, R. D., Hancock, M. D., and Ingersoll, G. M. 2003. "Online forum messages posted by adolescents with type 1 diabetes," *The Diabetes Educator* (30:5), pp 827-834.
Rebholz-Schuhmann, D., Oellrich, A., and Hoehndorf, R. 2012. "Text-mining solutions for biomedical research: enabling integrative biology," *Nature Reviews Genetics* (13:12), pp 829-839.
Ridings, C. M., and Gefen, D. 2004. "Virtual community attraction: Why people hang out online," *Journal of Computer-Mediated Communication* (10:1), pp 00-00.
Savolainen, R. 2011. "Requesting and providing information in blogs and internet discussion forums," *Journal of Documentation* (67:5), pp 863-886.
Schwarzer, R. 2004. *Psychologie des Gesundheitsverhaltens: Einführung in die Gesundheitspsychologie*, (Hogrefe Verlag.
Solt, I., Tikk, D., Gál, V., and Kardkovács, Z. T. 2009. "Semantic classification of diseases in discharge summaries using a context-aware rule-based classifier," *Journal of the American Medical Informatics Association* (16:4), pp 580-584.
Statnikov, A., Aliferis, C. F., Tsamardinos, I., Hardin, D., and Levy, S. 2005. "A comprehensive evaluation of multicategory classification methods for microarray gene expression cancer diagnosis," *Bioinformatics* (21:5), pp 631-643.
Stryker, S. 1987. *Identity theory: developments and extensions*, (John Wiley & Sons: Oxford, England.
Tajfel, H., and Turner, J. C. 2004. *The Social Identity Theory of Intergroup Behavior*, (Pshchology Press: New York, NY, US.
Taylor, H. 2010. ""Cyberchondriacs" on the Rise?."
Taylor, S., and Todd, P. A. 1995. "Understanding information technology usage: A test of competing models," *Information systems research* (6:2), pp 144-176.
Van der Heijden, H. 2004. "User acceptance of hedonic information systems," *MIS quarterly* (28:4), pp 695-704.
Venkatesh, V., Speier, C., and Morris, M. G. 2002. "User acceptance enablers in individual decision making about technology: Toward an integrated model," *Decision Sciences* (33:2), pp 297-316.
Wasko, M. M., and Faraj, S. 2000. "“It is what one does”: why people participate and help others in electronic communities of practice," *The Journal of Strategic Information Systems* (9:2), pp 155-173.
Wasko, M. M., and Faraj, S. 2005. "Why should I share? Examining social capital and knowledge contribution in electronic networks of practice," *MIS quarterly* (29:1), pp 35-57.
Waterson, P. 2006. "Motivation in online communities," *Encyclopedia of virtual communities and technologies*), pp 334-337.
Wykes, L. 1998. "Commitment in the Workplace: Theory, Research, and Application, by John P. Meyer, Natalie J. Allen," *Human Resource Development Quarterly* (9), pp 309-311.
Yamin, C. K., Bitton, A., and Bates, D. W. 2010. "E-cigarettes: a rapidly growing Internet phenomenon," *Annals of internal medicine* (153:9), pp 607-609.
Zaıane, O. R., Antonie, M.-L., and Coman, A. 2002. "Mammography classification by an association rule-based classifier," *MDM/KDD*), pp 62-69.
Zhang, P. 2013. "The affective response model: a theoretical framework of affective concepts and their relationships in the ICT context," *MIS Quarterly* (37:1), pp 247-274.
Zrebiec, J., and Jacobson, A. 2001. "What attracts patients with diabetes to an internet support group? A 21-month longitudinal website study," *Diabetic Medicine* (18:2), pp 154-158.