View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by AIS Electronic Library (AISeL)

A KNOWLEDGE ADOPTION MODEL BASED FRAMEWORK FOR FINDING HELPFUL USER-GENERATED CONTENTS IN ONLINE COMMUNITIES

Research-in-Progress

G. Alan Wang
Department of Business Information Technology, Virginia Tech
Blacksburg, Virginia
alanwang@vt.edu

Xiaomo Liu
Department of Computer Science, Virginia Tech
Blacksburg, Virginia
xiaomliu@vt.edu

Weiguo Fan
Department of Accounting and Information Systems, Virginia Tech
Blacksburg, Virginia
wfan@vt.edu

Abstract

Many online communities allow their members to provide information helpfulness judgments that can be used to guide other users to useful contents quickly. However, it is a serious challenge to solicit enough user participation in providing feedbacks in online communities. Existing studies on assessing the helpfulness of user-generated contents are mainly based on heuristics and lack of a unifying theoretical framework. In this article we propose a text classification framework for finding helpful user-generated contents in online knowledge-sharing communities. The objective of our framework is to help a knowledge seeker find helpful information that can be potentially adopted. The framework is built on the Knowledge Adoption Model that considers both content-based argument quality and information source credibility. We identify 6 argument quality dimensions and 3 source credibility dimensions based on information quality and psychological theories. Using data extracted from a popular online community, our empirical evaluations show that all the dimensions improve the performance over a traditional text classification technique that considers word-based lexical features only.

Keywords: Text classification, user-generated content, information helpfulness, knowledge adoption, online community

Introduction

Internet based online communities have become a popular and effective knowledge seeking and sharing platform in recent years. They often involve a large number of voluntary participants who communicate and interact with each other in a computer-based virtual space to seek shared purpose (De Souza and Preece 2004). Professionals in organizations use online communities as an external knowledge source because of the increasing demand and limited availability of expertise and knowledge within their organizations (Constant et al. 1996; Wasko and Faraj 2005; Zhang and Watts 2003). Gray and Meister (2006) found that IT-based knowledge sources are complementary to traditionally non-IT-based sources because different forms of knowledge sourcing are not directly interchangeable. Online communities are also one of the most common platforms for customers to share their experience and learn from each other (Lee et al. 2006). Customer-generated knowledge has become an important drive force for improved product designs and marketing strategies (Kozinets et al. 2008). In general, online communities provide a web-based platform to harness problem solutions using the collective intelligence of a distributed network of individuals (Brabham 2004).

Each of the large online communities over time has accumulated a large knowledge repository, usually consisting of millions of text-based user postings. However, due to the voluntary nature of the participations, knowledge management in online communities faces a number of challenges. First, their knowledge repositories usually suffer from low information quality (David and Pinch 2006). It has been found that information quality at an online community is often inversely related to the size of its membership (Gu and Konana 2007). According to Wikipedia, the largest community-based open-access collaborative knowledge repository, merely 0.09% Wikipedia articles met a set of information quality assessment criteria and qualified as featured articles as of August 2011 (Wikipedia 2011). The second challenge lies in the large amount of unstructured text messages accumulated over time. Community participants are subject to information overload when interpreting those unstructured data. Many online communities (e.g., Amazon Reviews) allow their members to provide information helpfulness judgments that can be used to guide other users to useful contents quickly. However, it is a serious challenge to solicit enough user participation in providing feedback in online communities (Rashid and Ling 2006). Moreover, Lampe and Resnick (2004) found that user provided feedbacks may lead to the problem of premature negative consent. Thus, there is a strong need to develop a method that can automatically assess information helpfulness for the text messages in large-scale online communities.

Quality assessment of user-generated contents has attracted many attentions in recent years. Chai et al. (2009) reviewed 19 content quality assessment studies conducted in various social media communities such as forums, question and answering portals, wikis, weblogs, and review portals. The survey reveals that the development of most of these assessment approaches employ heuristics without a unifying theoretical background. In fact, various theories have been developed with regard to the knowledge seeking behavior in online communities. For example, the heuristic-systematic model (HSM) also observes that people make tradeoffs between systematically processing cognitive information contents and heuristically assessing information validity (Eagly and Chaiken 1993). In the context of electronic communications, the knowledge adoption model (KAM) shows that information quality and source credibility are important factors positively affecting people's intention to adopt received information (Sussman and Siegal 2003).

In this study we are motivated to build a framework based on knowledge adoption model for finding helpful user-generated contents for online knowledge sharing communities. The objective of our framework is to find helpful information that a knowledge seeker has intention to adopt. The paper is organized as follows. We review literature related to text classification techniques on user generated contents and knowledge adoption model. We then propose our assessment framework followed by an empirical evaluation using data extracted from a real online community. Conclusions and future extensions are provided at the end.

Text Classification on User-Generated Contents

We consider the problem of finding helpful user-generated contents as a classification problem. We consider each discussion thread as a document. Each document can be classified into a helpful or non-helpful document. A helpful document means that it contains information helpful in solving the problem raised in the first posting of the thread. Natural Language Processing (NLP) based text classification has been commonly used for classifying documents into a fixed number of predefined classes. The first step in text classification is to transform documents into a feature vector where each distinct word corresponds to a feature. This representation often leads to very high-dimensional feature spaces due to the large size of its vocabulary. The second step is to apply a feature selection technique to select a subset of features. Finally, machine learning classifiers such as decision tree and Support Vector Machine (SVM) can be trained and used to make classification decisions (Dumais 1998).

Although word-based lexical features used in text classification have been shown effective, recent studies have explored other features that can be useful in predicting the quality of user-generated contents. In the context of online learning, Kim et al. (2006) analyzed student discussion threads in order to support question answering by extracting useful information from the discussion corpus. The features that they found related to the type of contribution include the number of posts in a thread, the average post length, the average number of replies, and topic coherence (i.e., speech act analysis). Weimer et al. (2007) proposed an algorithm for assessing the quality of posts in a software discussion forum. The features that they considered include surface features (e.g., post length and capital word frequency), lexical features (e.g., spelling error frequency), syntactic features (e.g., the percentage of part-of-speech tags), forum specific features (e.g., existence of HTML tags), and similarity features (e.g., the topical relevancy of a post with regard to the forum). Some studies focus on identifying helpful online product reviews. For example, Kim et al. (2006) considered document structure, syntactic features, semantic features, and meta-data features in addition to lexical features when predicting helpful product reviews. They found that the most useful features were the length of the review, its unigrams, and its product rating. In addition, Duan et al. (2010) examined the impact of text-based review features such as basic features (e.g., whether the review is positive or negative, how many days since the posting date, whether there are extreme opinions across different reviewers), stylistic features (e.g., writing styles), and semantic features (e.g., topics contained in each review) on the number of helpfulness votes those reviews receive. They found that the combination of the three feature types had the best prediction performance. Ghose and Ipeirotis (2011) built a prediction model on the perceived helpfulness of online product reviews. They considered three feature sets including review subjectivity, review readability, and reviewer characteristics.

The various feature sets, which previous studies have considered in predicting the quality of user-generated contents, are mostly based on heuristics and lack of a unifying theoretical framework. The categorization of the features can be contradictory from one study to another. For example, the features related to spelling errors belong to lexical features in (Weimer et al. 2007) while they are considered as review readability features in (Ghose and Ipeirotis 2011). There is a strong need to design a theory-based text classification framework for predicting the quality of user-generated contents. In online knowledge sharing communities, the quality of user-generated contents is often subjective to individual users. Therefore, quality assessment in online knowledge sharing communities must not only consider the quality of the textual contents but also the perception of individual users.

In this research we aim to propose a theory-grounded text classification framework based on the Knowledge Adoption Model for the quality assessment of user-generated contents in online knowledge-sharing communities. We review the Knowledge Adoption Model in the next section.

The Knowledge Adoption Model

Current studies mostly focused on online community participants' motivation to share knowledge or information. Very little attention was paid to the actual impact of the information received from online communities. It has led researchers to study the online community as an information adoption process and to understand the extent of information influence on people (Cheung, Lee, and Rabjohn 2008).

Sussman and Siegal (2003) proposed the Knowledge Adoption Model (KAM) in a study that investigated the information adoption process using email communication. KAM extends a communication theory, the Elaboration Likelihood Model (ELM) (Petty and Cacioppo 1986), to the context of electronic communication. ELM posits that a message can influence the message recipient's attitude and behavior by the message's central and peripheral cues. Central cues refer to the arguments contained in the message while peripheral cues refer to issues that are not directly related to the subject matter of the message.

KAM considers message recipients' perceived information usefulness as the direct determinant of knowledge adoption. The determinants of perceived information usefulness include the perceived argument quality and source credibility of the received message, being moderated by the recipients' domain expertise and involvement.

Perceived argument quality determines the degree of information influence towards the message recipient based on the message content only. It refers to the perceived quality of information content such as relevancy, accuracy, timeliness, and comprehensiveness (Bailey and Pearson 1983; Rabjohn et al. 2008). Information quality literature has provided a systematic and thorough discussion on the various information quality dimensions. In a seminal work on data quality, Wang and Strong (1996) discussed 20 quality dimensions that include not only content-based quality features but also features inferred from information sources. Based on their work we identify six quality dimensions that are content-based and can be automatically extracted using text mining techniques: relevancy, timeliness, completeness, appropriate amount of data, ease of understanding, and objectivity.

Perceived source credibility indicates a message recipient's perception on the credibility and authority of the information source reflecting nothing about the information content (Chaiken 1980). While different proposals on the underlying dimensions of source credibility have been discussed in various studies, two dimensions, competence and trustworthiness, have consistently emerged (Sussman and Siegal 2003). Competence-based source credibility refers to past experience and expertise while trustworthiness measures believability. Expertise is often considered as the dominant dimension of source credibility (Homer and Kahle 1990). Computationally, expertise finding techniques can be used to evaluate one's expertise with regard to a specific topic based his/her authored documents. Social computing studies have indicated that network centrality measures such as in-degree and betweenness correlate with trustworthiness (Prell 2003). Therefore, we can measure the trustworthiness of a posting user using social network centrality measures.

Each of the two factors in KAM may independently lead to a positive or negative usefulness perception on the received message. But they often contribute collectively to knowledge adoption in a complex way. According to ELM, argument quality is a critical determinant for information influence when the message recipient is able to comprehend the message content well. The source credibility of the message will become a critical factor only when the recipient is either unable or unwilling to process the message content.

A KAM-Based Framework for Finding Helpful User-Generated Contents

Based on KAM, we propose a text classification framework for finding helpful user-generated contents. This framework considers collectively argument quality and source credibility that both are positively related to perceived information usefulness and information adoption intention. The framework contains 6 argument quality dimensions, which measure the quality of textual contents, and 3 source credibility dimensions (Table 1). For each dimension we develop quantitative metrics based on the measures proposed in previous studies (Abbasi et al. 2008; Chiu et al. 2006; Otterbacher 2009; Prell 2003; Stone et al. 1962; Wasko and Faraj 2005). We define the metrics for information completeness that is defined as the breadth, depth, and scope of information by Wang and Strong (1996). The metrics are developed with the consideration that they can be applied to most online communities where user-generated contents are text documents and organized into discussion threads. Trustworthiness metrics need the construction of a social network based on past post/reply interactions in an online community. The network has community participants as its nodes with each directed edge linking from a replier to the original poster of the same thread. The network becomes a directed graph G_s with an adjacency matrix M, where $M(u, v)=1$ if member u replies to v in a thread and $M(u, v)=0$ otherwise. The design of the quantitative metrics is exploratory in the nature of this study. The usefulness of each metric can be evaluated using such approaches as information gain. For post-related metrics we calculate the metrics for the original question post and the minimum, maximum, average, and summation of the metrics for all replying posts in the same thread.

Table 1. Quantitative Metrics for the Dimensions of Argument Quality and Source Credibility

Category	Dimensions	Metrics
Argument Quality	F1: Appropriate Amount of Data (Abbasi et al. 2008; Otterbacher 2009)	1-5. Number of characters in a thread
		6-10. Number of words in a thread
		11-15. Number of unique words in a thread
		16-20. Number of sentences in a thread
		21-25. Number of web links in a thread
		26-30. Number of nomenclature (e.g., programming code, math formula) in a thread
		31-35. Number of quotations in a thread
	F2: Ease of Understanding (Otterbacher 2009)	36-40. Characters of sentences ratio in a thread
		41-45. Words to sentences ratio in a thread
		46-50. Number of wh-type words in a thread
		51-55. Number of question marks in a thread
		56-90. Ratio of nouns, adjectives, comparatives, verbs, adverbs, punctuations and symbols in a thread
	F3: Relevance (Abbasi et al. 2008; Otterbacher 2009)	91-94. Number of words overlapping between a question and its replies
		95-98. Cosine similarity between a question and its replies
		99-102. KL divergence between a question and its replies
		103-105. Centroid of a reply to all other replies in a thread
		104-106. Perplexity of a reply to all replies in a thread
		107-109. Entropy of a reply to all replies in a thread
	F4: Objectivity (Stone et al. 1962)	110-111. Ratio of positive and negative words in a question
		112. Number of "thank" words in a question
		113-114. Ratio of positive and negative words of the original poster (OP) to repliers in a thread
		115. Number of "thank" words of the OP to repliers in a thread
		116. If "thank" words appear in the last post of the OP in a thread
		117-124. Ratio of positive and negative words of repliers to the OP in a thread
		125-132. Ratio of positive and negative words of repliers to other repliers in a thread
	F5: Timeliness (Otterbacher 2009)	151-155. Question post time as hour-of-day, day-of-week, day-of-month, month-of-year
		156-158. Time lapse between a question and each reply
		159-161. Time lapse between a question and each direct reply
		162-164. Time span between consecutive replies in a thread
		165-167. Time span between consecutive replies replied to the question
	F6: Completeness	168. Number of replies in the thread
		169. Number of repliers in the thread
		170. Number of posts by the OP in the thread
		171. Number of posts directly replied to the OP
Source Credibility	F7: Past Experience (Wasko and Faraj 2005)	172-176. Tenure of the OP and repliers
		177-186. Number of questions and replies between the OP and repliers in the past
		187-189. The OP's user-feedback history
		190-201. Repliers' user-feedback history
	F8: Expertise (Chiu et al. 2006)	242-245. Cosine similarity between a replier's expertise profile and the question
		246-249. KL divergence between a replier's expertise profile and the question
		250-253. A replier's LDA-based expertise score on the question

	F9: Trustworthiness (Prell 2003; Newman 2008)	202-209. SNA centrality measures of the OP (in-degree, out-degree, betweennees, closeness, cluster coefficient, PageRank, and HITS scores)
		210-241. SNA centrality measures of repliers (in-degree, out-degree, betweennees, closeness, cluster coefficient, PageRank, and HITS scores)

We consider the problem of assessing the helpfulness of discussion threads as a classification problem. Given a set of discussion thread instances, each thread is converted into a tuple ($\boldsymbol{x}_i$, y_i) where $\boldsymbol{x}_i \in \Re^n$ is a vector of extracted features and $y_i \in L$ ($|L|>1$) is the class label. Figure 1 shows the system design of our proposed helpfulness classification framework.

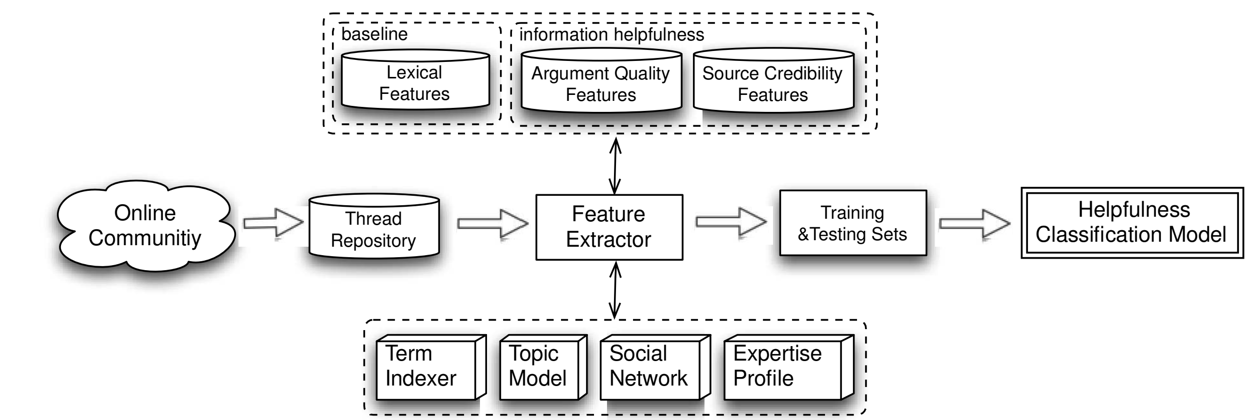

Figure 1. The System Design of the KAM based Framework for Finding Helpful User-Generated Contents

Experiments

Data Collection

We crawled the discussion threads published in the "Using iPhone" sub-forum as of April 6, 2010 from the *Apple Support Communities*. Table 2 summarizes some characteristics of the collected data set. This community allows community members to provide information helpfulness judgment with two tags: *"correct answer"* and "*helpful answer*". We parsed crawled data set and extracted 12,453 threads with *"correct answer"* tag and 1,140 threads with "*helpful answer*" tag. Excluding 6,702 threads without any reply, there were 29,048 threads, roughly 59%, without any user feedback. We randomly sampled 2,700 threads from those with no feedback and asked two domain experts, who are graduate students in Computer Science, to manually examine them. They agreed on 1,017 threads that did not contain any helpful answer (i.e., not-helpful).

Table 2. Data Characteristics

# of Threads	# of Posts	# of Participants	Avg. # of replies per thread	Avg. # of repliers per thread	Vocabulary size
49,343	271,823	55,108	4.13	3.15	56,157

Experimental Design

In this study we considered two different binary classification tasks. Many online knowledge-sharing communities allow their users to provide helpfulness feedbacks. Our first classification task is to predict if a discussion thread contains helpful information. Some online communities, especially those that focus on problem solving, let their users indicate if their problems have been solved. Our second classification task, which is considerably more difficult, is to detect if a thread contains not only helpful information but also solutions (i.e., correct answers) to the original problem. In order to produce a balanced data for our experiments, we randomly selected 500 threads with correct answers, 500 with helpful answers, and 1000 with not-helpful answers for the first classification task. We then randomly selected 1000 threads with correct answers, 500 with helpful answers, and 500 with not-helpful answers for the second task. For each thread we calculated the metrics of the 9 feature dimensions in Table 1. Those metrics are the independent variables of the classification tasks. The classification algorithm will categorize those threads into either helpful/non-helpful groups (task 1) or solution/non-solution groups (task 2). We conducted a five-fold cross validation for each of the classification tasks.

We used four commonly used classification algorithms in our experiments, namely Naïve Bayes (NB), C4.5 Decision Tree, ADA Boosting, and Support Vector Machine (SMO). We used the algorithm implementations in Weka (Hall et al. 2009). The baseline for our performance evaluation is the text classification technique that considers word-based lexical features only. Traditional text classification technique converts each thread document into a binary vector with each element indicating the occurrence of a word in the document (Dumais 1998). A classifier will classify each document into a helpful or not-helpful category based on the occurrences of certain words. The vector usually has a large dimension due to the large vocabulary in the corpus. A comparative study found that χ^2-test (CS), information gain (IG), and document frequency (DF) are the most effective feature selection methods for text categorization tasks (Yang and Pedersen 1997). We followed their feature selection approach to select a subset of terms in order to reduce computation complexity without significantly sacrificing classification performance. To test the effectiveness of each feature dimension in our proposed framework, we used the combination of the term-based lexical features and each feature dimension to predict the quality of each discussion thread. The experiment is designed to answer two research questions: (1) which feature dimension achieves the most performance improvement in finding helpful user-generated contents? (2) which classification algorithm achieves the best performance in finding helpful user-generated contents?

We used precision, recall and *f*-measure, which are commonly used in information retrieval (Salton 1988), to evaluate the performance of our classifiers.

$$Precision = \frac{\text{\# of correctly predicted positive threads}}{\text{\# of predicted positive threads}} \quad Recall = \frac{\text{\# of correctly predicted positive threads}}{\text{\# of acutal positive threads}}$$

$$F - measure = \frac{2 * precision * recall}{(precision + recall)}$$

Experimental Result

Figure 2 shows the F-measure performance of Naïve Bayes and SMO classifiers (other two classifiers with similar curves were omitted for clarification) with the top-*n* terms selected by CS, DF and IG in the baseline technique. We observed that all three term-selection methods achieved the best performance in both classification tasks with top 150-350 terms. CS achieved slightly better performance than DF and IG with top 150-350 terms. We selected the top 200 terms, where NB and SMO achieved collectively the best performance, as our baseline feature set (F0).

Table 3 summarizes the precision, recall and F-measure of the 4 classification algorithms in the two classification tasks. We conducted paired *t*-tests on *F*-measure in order to examine if a feature dimension had a significant performance improvement over the baseline feature set. The *p*-values show that every dimension significantly improved f-measure with at least one classification algorithm over the baseline. For argument quality dimensions, relevance (F3) significantly improved F-measure over the baseline for all 4 classifiers. Subjectivity (F4), timeliness (F5), and completeness (F6) achieved significantly improved F-measure for 3 classifiers. For source credibility dimensions, all three dimensions achieved significantly improved F-measure over the baseline. When combining one feature dimension with the baseline term-based features, naïve Bayes appeared to perform consistently well in Task 1 while Ada Boosting achieve the best performance overall in Task 2. However, when all feature dimensions were considered, SVM achieved the best performance. Although the F-measures in Task 2 were slightly lower than those in Task 1, they were still satisfactory.

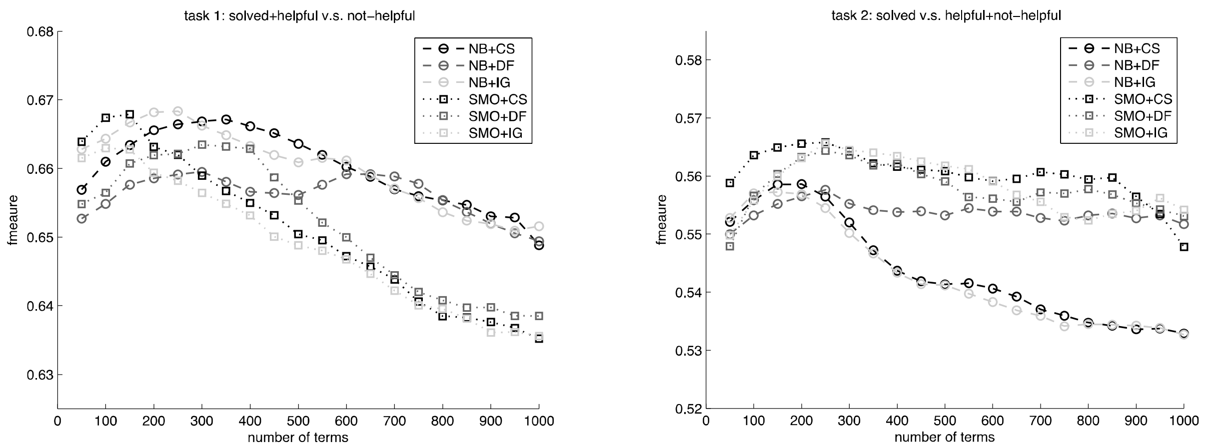

Figure 2. The Performance of the Text Classification Approach with Top-*n* Terms

Table 3. The performance of four classification algorithms and ten feature sets (Significant results are highlighted in bold)

	Naïve Bayes			C4.5			Ada Boosting			SVM		
Task 1: Helpful+Correct vs. Not-helpful												
	Prec.	Rec.	F-mea.	Prec.	Rec.	F-mea.	Prec.	Rec.	F-mea.	Prec.	Rec.	F-mea.
F0 (baseline)	0.761	0.595	0.668	0.652	0.584	0.616	0.709	0.603	0.652	0.694	0.622	0.656
F0+F1	0.760	0.631	0.690	0.645	0.623	0.634	0.645	0.818	**0.721****	0.725	0.646	0.683
F0+F2	0.710	0.745	**0.727****	0.641	0.62	0.630	0.685	0.620	0.651	0.733	0.663	**0.696***
F0+F3	0.720	0.755	**0.737****	0.660	0.645	**0.652***	0.689	0.685	**0.687***	0.732	0.674	**0.702***
F0+F4	0.720	0.745	**0.732****	0.661	0.595	0.626	0.711	0.685	**0.698***	0.734	0.680	**0.706****
F0+F5	0.593	0.930	**0.724****	0.682	0.651	**0.666***	0.687	0.747	**0.716****	0.727	0.648	0.685
F0+F6	0.743	0.732	**0.738****	0.680	0.675	**0.678****	0.677	0.870	**0.738****	0.728	0.645	0.684
F0+F7	0.748	0.725	**0.736****	0.706	0.637	**0.670****	0.709	0.738	**0.723****	0.743	0.704	**0.723****
F0+F8	0.735	0.717	**0.726****	0.668	0.637	**0.651***	0.696	0.686	**0.691***	0.736	0.658	**0.695***
F0+F9	0.753	0.658	**0.702****	0.663	0.647	**0.655***	0.704	0.687	**0.695***	0.732	0.681	**0.705***
All features	0.746	0.719	**0.732****	0.745	0.740	**0.743*****	0.717	0.807	**0.759****	0.799	0.79	**0.794*****
Task 2: Correct v.s. Helpful+Not-helpful												
F0 (baseline)	0.665	0.474	0.553	0.562	0.519	0.540	0.628	0.524	0.571	0.611	0.522	0.563
F0+F1	0.704	0.535	0.608	0.611	0.581	**0.596***	0.675	0.67	**0.673****	0.689	0.598	**0.640***
F0+F2	0.656	0.688	**0.672****	0.597	0.568	0.582	0.663	0.628	**0.645***	0.683	0.624	**0.652****
F0+F3	0.671	0.615	**0.642***	0.647	0.624	**0.635****	0.634	0.824	**0.717****	0.676	0.629	**0.652****
F0+F4	0.669	0.682	**0.675****	0.638	0.601	**0.619****	0.646	0.705	**0.674****	0.681	0.605	**0.641***
F0+F5	0.614	0.761	**0.680****	0.579	0.553	0.566	0.667	0.634	**0.650****	0.649	0.563	0.603
F0+F6	0.692	0.589	**0.636***	0.663	0.634	**0.648****	0.619	0.853	**0.717****	0.682	0.582	**0.628***

F0+F7	0.703	0.647	**0.674****	0.706	0.721	**0.713*****	0.719	0.701	**0.710*****	0.714	0.664	**0.688****
F0+F8	0.663	0.635	**0.649***	0.637	0.591	**0.613****	0.663	0.735	**0.697****	0.670	0.601	**0.634***
F0+F9	0.654	0.721	**0.686****	0.604	0.598	**0.601****	0.672	0.634	**0.652***	0.677	0.615	**0.645***
All features	0.702	0.665	**0.683****	0.757	0.735	**0.746*****	0.758	0.734	**0.746*****	0.763	0.763	**0.764*****

*:p-value<0.05, **: p-value<0.01, ***:p-value<0.001

Conclusions and Discussions

In this article we proposed a text classification framework for finding helpful user-generated contents in online knowledge-sharing communities. The objective of our framework is to help a knowledge seeker find useful information that can be potentially adopted. The framework was built on the Knowledge Adoption Model that considers both content-based argument quality and information source credibility. We identified 6 argument quality dimensions and 3 source credibility dimensions based on information quality and psychological theories. Using data extracted from a popular online community, our empirical evaluation showed that all the dimensions improved the performance over a traditional text classification technique that used word-based lexical features alone.

This study has significant implications on the field practitioners and users of online communities. The providers of online communities can use the proposed framework to automatically judge the helpfulness of user-generated contents in addition to the limited helpfulness feedback provided by the community participants. The user can be benefited from the framework by easily accessing helpful knowledge embedded in the text documents in online communities. The proposed framework is an essential complement to the human-generated helpfulness judgments, which are often very limited in online communities (Lampe and Resnick 2004).

Several future extensions can be done to this study. First, following a design science approach, we will assess the appropriateness and usefulness of the metrics with respect to the performance of the machine learning algorithms in finding helpful user-generated contents. Second, we will examine the importance of each feature dimension in order to verify and/or update the theories that we build our framework upon. Although KAM has been tested in the context of electronic communications, it will be interesting to learn if KAM could adapt to the unique characteristics of online communities. We will use those findings to construct the optimal set of feature dimensions. Third, our empirical evaluation shows that a feature dimension may achieve different performance with different classification algorithms. It is also interesting to examine the sensitivity of feature dimensions with each classification algorithm. It will provide guidance on the selection of classification algorithms depending on the availability of feature dimensions. Fourth, we did not consider the moderation effect of the users' domain expertise and involvement suggested by the KAM. Most of online community users do not interact with the community on a regular basis. Due to the lack of evidence, it is difficult to assess one's expertise in a particular domain. However, we can study this moderation effect for active users in the community. Lastly, we intend to improve the external validity of our research findings by conducting additional empirical evaluations with other online communities.

Acknowledgements

This research is partly supported by the Natural Science Foundation of China (Grant #70872089 and #71072129) and the National Science Foundation (Grant #DUE-0840719).

References

Abbasi, A., Chen, H., and Nunamaker, J. F. 2008. "Stylometric Identification in Electronic Markets: Scalability and Robustness," *Journal of Management Information Systems*, (25:1), pp. 49-78.

Bailey, J. E., and Pearson, S. W. 1983. "Development of a Tool for Measuring and Analyzing Computer User Satisfaction," *Management Science* (29:5), pp. 530-545.

Brabham, D. C. 2008. "Crowdsourcing as a Model for Problem Solving: An Introduction and Cases," *The International Journal of Research into New Media*, (14:1), pp. 75-90.

Chai, K., Potdar, V., and Dillon, T. 2009. "Content Quality Assessment Related Frameworks for Social Media," in *Lecture Notes in Computer Science: Computational Science and Its Applications*, O. Gervasi, D. Taniar, B. Murgante, and A. Lagan, Y. Mun, and M. Gavrilova (eds.), Berlin, Germany: Springer-Verlag, pp. 791-805.

Chaiken, S. 1980. "Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion," *Journal of Personality and Social Psychology*, (39:5), pp. 752-766.

Cheung, C. M. K., Lee, M. K. O., and Rabjohn, N. 2008. "The Impact of Electronic Word-of-Mouth: The Adoption of Online Opinions in Online Customer Communities," *Internet Research*, (18:3), pp. 229-247.

Chiu, C.-M., Hsu, M.-H., and Wang, E. 2006. "Understanding Knowledge Sharing in Virtual Communities: An Integration of Social Capital and Social Cognitive Theories," *Decision Support Systems*, (42:3), pp. 1872-1888.

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), p. 119.

David, S., and Pinch, T. 2006. "Six Degrees of Reputation: The Use and Abuse of Online Review and Recommendation Systems," *First Monday*, (11:3), pp. 1–29.

De Souza, C. S., and Preece, J. 2004. "A Framework for Analyzing and Understanding Online Communities," *Interacting with Computers*, (16:3), pp. 579–610.

Duan, W., Cao, Q., and Gan, Q. 2010. "Investigating Determinants of Voting for the 'Helpfulness' of Online Consumer Reviews: A Text Mining Approach," In *AMCIS 2010 Proceedings*, p. 497.

Dumais, S. 1998. "Using SVMs for Text Categorization," *IEEE Intelligent Systems*, (13:4), pp. 21–23.

Eagly, A. H., and Chaiken, S. 1993. *The Psychology of Attitudes*, Fort Worth, TX: Harcourt Brace Jovanovich College Publishers.

Ghose, A., and Ipeirotis, P. G. 2011. "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics," *IEEE Transactions on Knowledge and Data Engineering*, (23:10), pp. 1498-1512.

Gray, P. H., and Meister, D. B. 2006. "Knowledge Sourcing Methods," *Information & Management*, (43:2), pp. 142–156.

Gu, B., and Konana, P. 2007. "Competition among Virtual Communities and User Valuation: The Case of Investing-Related Communities," *Information Systems Research*, (18:1), pp. 68-85.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. 2009. "The WEKA Data Mining Software: An Update," *ACM SIGKDD Explorations Newsletter*, (11:1), pp. 10–18.

Homer, P. M., and Kahle, L. R. 1990. "Source Expertise, Time of Source Identification, and Involvement in Persuasion: An Elaborative Processing Perspective," *Journal of Advertising*, (19:1), pp. 30-39.

Kim, J., Shaw, E., Feng, D., Beal, C., and Hovy, E. 2006. "Modeling and Assessing Student Activities in Online Discussions," In *Proceedings of the AAAI Workshop on Educational Data Mining*, Boston, MA.

Kim, S.-M., Pantel, P., Chklovski, T., and Pennacchiotti, M. 2006. "Automatically Assessing Review Helpfulness," in *Proceedings of Conference on Empirical Methods in Natural Language Processing*, Sydney, Australia, pp. 423-430.

Kozinets, R. V., Hemetsberger, A., and Schau, H. J. 2008. "The Wisdom of Consumer Crowds," *Journal of Macromarketing*, (28:4), p. 339.

Lampe, C., and Resnick, P. 2004. "Slash (dot) and Burn: Distributed Moderation in a Large Online Conversation Space," In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, Vienna, Austria, pp. 543–550.

Lee, M. K. O., Cheung, C. M. K., Lim, K. H., and Sia, C. L. 2006. "Understanding Customer Knowledge Sharing in Web-based Discussion Boards: An Exploratory Study," *Internet Research*, (16:3), pp. 289-303.

Newman, M. E. J. 2008. "Mathematics of Networks," in *The New Palgrave Encyclopedia of Economics*, L.E. Blume and S. N. Durlauf (eds.), Basingstoke, United Kingdom:Palgrave McMillian.

Petty, R. E., and Cacioppo, J. T. 1986. *Communication and Persuasion: Central and Peripheral Routes to Attitude Change*, New York: Springer-Verlag.

Prell, C. L. 2003. "Community Networking and Social Capital: Early Investigations," *Journal of Computer-Mediated Communication*, (8:3).

Rabjohn, N., Cheung, C. M. K., and Lee, M. K. O. 2008. "Examining the Perceived Credibility of Online Opinions: Information Adoption in the Online Environment," In *Proceedings of the 41st Annual Hawaii International Conference on System Sciences*, Waikoloa, Big Island, Hawaii, p. H3.

Rashid, A., and Ling, K. 2006. "Motivating Participation by Displaying the Value of Contribution," In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, Montréal, Canada, pp. 955--958.

Salton, G. 1988. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*, Addison-Wesley.

Sussman, S. W., and Siegal, W. S. 2003. "Informational Influence in Organizations: An Integrated approach to Knowledge Adoption," *Information Systems Research*, (14:1), pp. 47-65.

Wang, R. Y., and Strong, D. M. 1996. "Beyond Accuracy: What Data Quality Means to Data Consumers," *Journal of Management Information Systems*, (12:4), pp. 5–33.

Wasko, 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.

Weimer, M., Gurevych, I., and Mühlhäuser, M. 2007. "Automatically Assessing the Post Quality in Online Discussions on Software," In *Proceedings of the 45th Annual Meeting of the ACL on Interactive Poster and Demonstration Sessions*, Stroudsburg: Association for Computational Linguistics, pp. 125–128.

Wikipedia 2011. "Wikipedia: Featured article Statistics." http://en.wikipedia.org/wiki/Wikipedia:FAS.

Yang, Y., and Pedersen, J. O. 1997. "A Comparative Study on Feature Selection in Text Categorization," In *Proceedings of International Conference on Machine Learning*, Corvallis, OR, pp. 412--420.

Zhang, W., and Watts, S. 2003. "Knowledge Adoption in Online Communities of Practice," *Systemes d□Information Et Management*, (9:1), pp. 81-102.