Understanding and Predicting Question Subjectivity in Social Question and Answering

Zhe Liu and Bernard J. Jansen

***Abstract*— The explosive popularity of social networking sites has provided an additional venue for online information seeking. By posting questions in their status updates, more and more people are turning to social networks to fulfill their information needs. Given that understanding individuals' information needs could improve the performance of question answering, in this paper, we model the task of intent detection as a binary classification problem, and thus for each question, two classes are defined: subjective and objective. We use a comprehensive set of lexical, syntactical, and contextual features to build the classifier and the experimental results show satisfactory classification performance. By applying the classifier on a larger dataset, we then present in-depth analyses to compare subjective and objective questions, in terms of the way they are being asked and answered. We find that the two types of questions exhibited very different characteristics, and further validate the expected benefits of differentiating questions according to their subjectivity orientations.**

***Index Terms*— Information seeking, social question and answering (social Q&A), social search, subjectivity analysis social network, Twitter.**

I. Introduction

THE emergence of social networking sites (SNSs), such as Facebook and Twitter, has made the communication among individuals more diverse and convenient [1], [2]. Besides using those social platforms for relationship maintenance, many people also perceive SNSs as valuable information sources and engage in what has been referred to as social question and answering (social Q&A) [3]. Compared with the typical search engine services, such as Google and Bing, social Q&A provides people a more direct and easier way to express their information needs, as individuals can publicly broadcast their request for help in natural languages to all friends or followers online, and to receive more personalized and trustworthy responses [4].

As a result of the ever-increasing popularity of social Q&A, a variety of different questions are being asked on SNSs. Some seek for subjective objective knowledge or factual truth, such as *How do I update to IOS 8?* Others request for more subjective information, such as personal opinions or recommendations on certain topics, like *What should I say when asking her out for a meal?* Objective questions focus more on the accuracy of the responses and are expected to be answered by more reliable sources, whereas subjective questions require more diverse replies that rely on personal opinion and perspective. Considering the distinct intents behind, we believe that there does not exist a one-size-fits-all approach to answer both types of questions, and it is necessary to differentiate subjective questions from the objective ones.

With the above-mentioned aim in mind, in this paper, we focus on conducting subjectivity analysis on the questions asked in social Q&A. We build a model to predict whether a question is subjective or objective using a comprehensive set of features from lexical, syntactical, and contextual perspectives. We evaluate the classifier on 3000 randomly sampled questions extracted from Replyz.com, a twitter-based Q&A site. Next, with the classifier on question subjectivity, we also conduct comprehensive analyses on a set of 10 386 information-seeking tweets and 102 131 corresponding answers. We investigate how subjective and objective questions differ in terms of the way they are being asked and answered. We show that subjective questions contain more contextual information and are being asked more during the working hours. Compared with the subjective information-seeking tweets, objective questions experience a shorter time lag between posting and receiving responses and tend to receive less but informative responses. Moreover, we also observed that subjective questions attracted more responses from strangers than the objective ones.

Our contributions are as follows. Using simple features extracted from the question text, our method can automatically detect the subjectivity orientation of a questioner's intent. We believe that by automatically distinguishing subjective questions from the objective ones, one could ultimately build question routing systems that can direct a question to its potential answerers according to its underlying intent. For instance, given a subjective question, we could route it to somebody who knows the questioner well to provide more personalized responses. However, for an objective question, we could discover authorities within a particular domain or could automatically answer a new question using the archived question–answer pairs.

II. Related Work

We reviewed a number of recent studies in the literature on both social Q&A and question classification. To the best of our knowledge, this is the first study investigating the subjectivity of information needs on SNSs.

A. Question Asking in Social Q&A

As an emerging concept, social Q&A has been given very high expectations due to its potential as an alternative to traditional information-seeking tools (e.g., search

Manuscript received June 2, 2015; revised January 17, 2016; accepted April 20, 2016. Date of current version June 17, 2016.

Z. Liu is with the IBM Almaden Research Laboratory, San Jose, CA 95120 USA (e-mail: zul112@ist.psu.edu).

B. J. Jansen is with the Social Computing Group, Qatar Computing Research Institute, Doha 5825, Qatar (e-mail: jjansen@acm.org).

Digital Object Identifier 10.1109/TCSS.2016.2564400

engines, online catalogs, and databases). Jansen *et al.* [1], in their work examining Twitter as a mechanism for word-of-mouth advertising, reported that 11.1% of the brand-related tweets were information providing, while 18.1% were information seeking. Li *et al.* [5] revealed that there were about 11% of general tweets containing questions and 6% of tweets having information needs. Going one step further, Efron and Winget [6] analyzed 100 question tweets on Twitter and proposed a taxonomy of questions asked on microblogging platforms. Morris *et al.* [3] manually labeled a set of questions posted on social networking platforms and identified eight question types in social Q&A, including recommendation, opinion, factual knowledge rhetorical, invitation, favor, social connection, and offer. In the set of tweets they analyzed, recommendation (29%) and opinion (22%) questions accounted for the majority of cases. Different from [3], Paul *et al.* [7] observed more rhetorical (42%) questions on Twitter, followed by the categories of factual knowledge (16%), and polls (15%). Adapting the categorization scheme proposed in [3], Ellison *et al.* [8] labeled a set of 20 000 status updates on Facebook and presented multiple types of mobilization requests beyond information-seeking attempts.

B. Automatic Question Classification

Most of the above-mentioned studies performed the question classification task manually based on handcrafted rules. There are only a few papers that touch on the problem of automatic question classification based on machine learning techniques. Li *et al.* [5] proposed a cascade approach, which first detected interrogative tweets and then questions revealing real information needs (referred to as qweets in their paper). They relied on both rule-based (as proposed in [6]) and learning-based approaches for interrogative tweets detection and some Twitter-specific features, such as retweet, mentioned to extract qweets. Through their experiment, Efron and Winget [6] observed that rule-based approach actually outperformed the learning-based method in identifying interrogative tweets. Zhao and Mei [9] classified question tweets into two categories: tweets conveying information needs and tweets not conveying information needs. They manually labeled 5000 tweets and built an automatic text classifier based on lexical, part-of-speech (POS) tagging, and meta features. With the classifier, they further investigated the temporal characteristics of those information-seeking tweets. We view our work as a further step of the above-mentioned studies in the direction of understanding and comprehending the question's intent in social Q&A.

Besides the two works in social Q&A, most of the existing studies on automatic question classification were conducted in the context of community question and answering (community Q&A), which are sites specifically designed for asking questions. Analyzing questions from three popular community Q&A sites, Harper *et al.* [10] automatically classified questions into conversational and informational and reached an accuracy of 89.7% in their experiments. As a result of their analysis, they claimed that conversational questions typically have much lower potential archival value than the informational ones. Kim *et al.* [11] classified questions from Yahoo! Answers into four categories: information, suggestion, opinion, and other. They pointed out that the criteria of selecting best answer differed across categories. Pal *et al.* [12] introduced the concept of question temporality based on when the answers provided on the questions would expire. They labeled questions into five categories, with permanent, long, medium, short, and other temporal durations. Their results showed that question temporality can be automatically detected using question vocabulary and other simple features.

C. Subjectivity Analysis

As for the task of subjectivity analysis, Wilson *et al.* [13] developed a system called OpinionFinder, which performs subjectivity analysis by automatically identifying subjective sentences and to mark the source of the subjectivity and words expressing positive or negative sentiments. By identifying subjective sentences that contain strong subjective clues based on the General Inquirer dictionary, Jiang and Argamon [14] classified a political blog as either liberal or conservative, based on its political leaning. Biyani *et al.* [15], [16] analyzed subjectivity orientation of online forum threads using the combinations of words and their parts-of-speech tags as features as extracted from the title of the thread and initial post, as well as the entire thread. Li *et al.* [17] labeled 987 resolved questions from Yahoo!Answers and explored a supervised learning algorithm utilizing features from both the perspectives of questions and answers to predict the subjectivity of a question. Zhou *et al.* [18] proposed an approach to automatically collect training data based on social signals, such as vote and answer number, in community Q&A sites. The results of their experiment demonstrated that leveraging social interactions in community Q&A portals could significantly improve prediction performance. Chen *et al.* [19] classified questions from Yahoo! Answers into subjective, objective, and social. They built a predictive model based on both text and metadata features and cotraining them. Their experimental results showed that cotraining worked better than simply pooling these two types of features together. Aikawa *et al.* [20] employed a supervised approach in detecting Japanese subjective questions in Yahoo! Chiebukuro. Unlike the other studies, they evaluated the classification results using weighed accuracy that reflected the confidence of annotation.

The studies that are most related to our work are described above. However, some important differences between our work and the past ones are as follows. First, compared with questions asked on community Q&A platforms, twitter questions are relatively short and informally phrased. This definitely adds difficulties to the task of question intention detection. Therefore, in this paper, we propose an approach that is particularly developed to identify question subjectivity in social Q&A context, instead of community Q&A sites. In addition, given the casual nature of SNS, questions asked in social Q&A may contain different intentions compared with those asked on community Q&A platforms. Our work addresses these differences and further explores the question asking and answering patterns that happen in the social

Q&A process. We believe that our result can contribute to provide better Q&A services on social platforms in the future. Finally, compared with [17] and [18], our method relies only on the question itself and does not depend on any information from the answers or the users, and thus can be applied to all questions, no matter with or without solutions.

III. Research Objectives

To address the gaps as mentioned in Section II, we propose two overarching research objectives in this study.

Objective 1: Automatically determine the subjectivity orientation of an information-seeking question on Twitter.

Our first research objective aims to examine whether by monitoring the way a question is phrased, one can tell if it is subjective or objective. To accomplish this research objective, we explore the lexical, syntactical, and contextual differences between subjective and objective questions posted on Twitter and build a predictive model that can reliably distinguish the two types of questions using machine learning algorithms.

Objective 2: Further analyze the differences between subjective and objective questions in terms of the way they are being asked and answered.

To measure the differences, we introduce metrics including question length, phrasing, posting time, response speed, informativeness, and the characteristics of the respondent. Due to the distinct nature of the two types of questions, we anticipate significant differences in all proposed metrics.

IV. Problem Formulation and Feature Engineering

A. Problem Formulation

As we discussed earlier, questions posted on SNSs can be either subjective or objective. To achieve an unambiguous understanding of question subjectivity, here we provide the definition of subjective and objective information-seeking questions in social context, respectively. We define *subjective information-seeking questions* as SNS posts asking for responses reflecting the answerer's *personal opinions, advices, preferences, or experiences*. A subjective information-seeking tweet is usually with a survey purpose, which encourages the audience to provide their personal answers. In contrast, *objective questions* are characterized as SNS posts requesting answers based on some *factual knowledge* or *common experiences*. The purpose of the objective questions is to receive one or more correct answers, instead of responses based on the answerer's personal experience. Questions asking how to do something usually belong to the objective category.

To better illustrate our annotation criteria used in this paper, in Table I we listed a number of sample questions with objective or subjective intents.

TABLE I
Definitions of Objective and Subjective Questions

Question Type	Sample Questions
Subjective	• Can anyone recommend a decent electric toothbrush? • I wanna change my username. Any ideas...? • How does the rest of the first season compare to the pilot? Same? Better? Worse?
Objective	• When is the debate on UK time? • Mac question. If I want to print a doc to a color printer but in B&W how do I do it? • Anyone know how to say "fun smasher" in Spanish? #help

B. Feature Engineering

We modeled question subjectivity using three groups of features: lexical features, syntactic features, and contextual features. Again, we adopt only features extracted from the question content and ignore all information from either the answer or answer provider's perspective. In this way, our classification model can be applied to all questions posted in social Q&A, no matter with or without solutions.

1) Lexical Features:

a) N-gram: We assume that given the different information needs behind, there should be a different usage of lexical terms between subjective and objective information-seeking tweets. Hence, in this study, we adopted word-level n-gram features. We counted the frequencies of all unigram, bigram, and trigram tokens that appeared in the training data, as they have been proved to be useful in [9] and [20]. Before feature extraction, we lowercase and stemmed all the tokens using the Porter stemmer [21]. We discarded rare terms with observed frequencies of less than 5 to reduce the sparsity of the data. This left us with 960 n-gram features.

b) POS tagging: We believed that POS tagging may also help in distinguishing the two types of questions, as it can add more context to the words used in the interrogative tweets. To tag the POS of each tweet, we used the Stanford tagger [27]. Again, we counted the frequencies of all unigram, bigram, and trigram POS that appeared in the training data. POS sequences with frequencies less than 5 were also eliminated. This left us with 1070 features of POS. taggings.

c) MPQA subjectivity lexicon: In addition to the n-gram and POS tagging features, we also counted the number of subjective clues [22] that appeared in each question using MPQA Subjectivity Lexicon[1] [23]. The lexicon contains in total 8222 subjective clues. Among them, 5569 are strongly subjective clues, while the rest 2653 are weak ones. According to [24], strongly subjective clues are seldom used without subjective meanings, whereas weakly subjective clues are often of ambiguous subjectivity orientations. Therefore, in this study, we counted only the frequencies of the lexical clues that are considered to be strongly subjective in each question. We then normalized the frequency of subjectivity clues by the total number of words in the corresponding question.

2) Syntactic Features: The syntactic features describe the format of a subjective or objective information-seeking tweet. The syntactic features that we adopted in this study include the length of the tweet, number of clauses/sentences in the tweet, whether or not there is a question mark in the middle of the

[1] http://www.cs.pitt.edu/mpqa/subj_lexicon.html.

tweet, whether or not there are consecutive capital letters in the tweet.

3) Contextual Features: The Twitter-based contextual features, such as the presence of hashtags, mentions, and emoticons, captured the unique characteristics of content posted on Twitter, and have been widely adopted in past studies on sentiment analysis [25], [26]. We assume that these features can provide extra signals for determining whether a question is subjective or objective. The Twitter-specific features that we adopted in this study are whether or not a question tweet contains a hashtag, a mention to somebody, or an emonticon.

In total, we have extracted 2040 features from the abovementioned perspectives.

V. Classification Experiments

A. Data

Given the high percentage of conversational questions on Twitter [5], [10], in order to collect as many information-seeking questions as possible, in this study, we collected question tweets from a site called Replyz.[2] Replyz is a very popular Twitter-based Q&A site, which searches through Twitter in real time looking for posts that contain questions based on their own algorithm. By collecting questions through Replyz, we filtered out a large number of conversational tweets. Another advantage of using Replyz for data collection is that due to its community Q&A nature, it allows answerers to respond to anybody's questions, not limited to the follower relationship. In that sense, comparing with directly collecting question tweets from Twitter, this way, we guarantee that most of our collected questions have been answered by at least one stranger, so that we can address our second research objective.

For our data collection on Replyz, we employed a snowball sampling approach. To be more specific, we started with the top ten contributors who have signed in Replyz with their Twitter account as listed on Replyz's leaderboard. For each of these users, we crawled all the question tweets that they have answered in the past from their Replyz profile. Then we identified the individuals who posted those collected questions and went to their profile to crawl all the interrogative tweets that they have ever responded. We repeated this process until each seed user yielded at least 500 other unique accounts. After removing those non-Twitter questioners in our collection, in total, we crawled 25 697 question tweets and 271 821 answers from 10 101 unique questioners and 148 639 unique answerers. To build and evaluate our classifier on question subjectivity, we randomly sampled 3000 English questions from our data collection and recruited two human annotators to work on the labeling task based on our proposed definitions on objective and subjective information-seeking tweets. Finally, 2588 out of 3000 questions (86.27%) received agreement on their subjectivity orientation from the two coders. Among the 2588 interrogative tweets, 24 (0.93%) were labeled as with mix intent, 1303 (50.35%) were annotated as noninformation seeking, 536 (20.71%) as subjective information seeking, and the rest 725 (28.01%) as objective information seeking. Our Cohen's kappa is quite high at 0.75.

[2]http://www.replyz.com (shut down on July 31, 2014).

We also examined the 412 questions with annotation differences and found that the major cause of such disagreement is that without knowing the context of a question, annotators interpreted the questioner's intent differently. For instance, for question *Ok as we start to evaluate #Obama legacy who is worse him or #Carter?,* one annotator tagged it as subjective as it surveys the audience about their opinions regarding Obama, while the other treated this tweet as sarcasm and tagged it as noninformation seeking. To build our classifier, we used only the 536 subjective and 725 objective information-seeking questions.

B. Experiment Settings

We next adopted a number of supervised learning algorithms to build the binary classifier on question subjectivity. We experimented with Naïve Bayes support vector machines (SVMs) (sequential minimal optimization) and decision trees (J48) as implemented in WEKA [29]. Classifier parameters were chosen in an experimental way to obtain the best classification performance. To be more specific, a search space is defined by specifying a number of discrete values for each parameter (e.g., J48 was tested with a confidence factor ranging from 0.1 to 1.0 by an increment of 0.1, SVM was tested with c ranging from 10^{-5} to 10^{5} by an increment of 0.01, and γ ranging from 10^{-15} to 1 with the radial basis function kernel). Cross-validation procedure was then performed for each point in the search space. Parameters with the best performance were demonstrated in the Section V-C.

For evaluation purposes, we calculated the classic machine learning evaluation metrics, such as accuracy, precision, recall, F1-measuremet, and area under receiver operating characteristic curve (AUC) values, as they have also been adopted in [31] and [32].

We also adopted the majority induction algorithm, which predicts the majority class in the data set as a baseline model to interpret our classification results and evaluate our classification method. With this approach, our data set got a baseline accuracy of 0.575 as 725 tweets were tagged as objective information-seeking among the overall 1261 informational seeking questions.

C. Classification Results

First, due to the large number of features extracted, before conducting the classification, we performed feature selection using the information gain criterion [28] as implemented in WEKA [29]. The method of information gain helped us to identify the most informative and relevant features in the classification process and to reduce noise from irrelevant or inaccurate features. We evaluated the classification accuracies along with the number of features selected and plotted the results in Fig. 1.

We saw from Fig. 1 that either too few or too many features would result in a decrease in the prediction accuracy. Table II shows the optimum classification performance using all three classifiers in conjunction with corresponding use of the selected features.

We observed from Table II that among all three methods, SVM outperformed the other two in the subjectivity

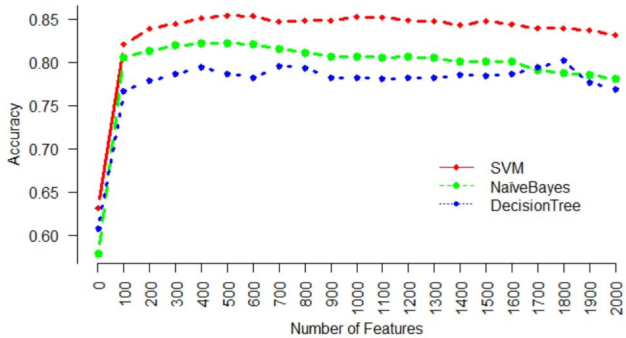


Fig. 1. Classification accuracy as a function of the number of features selected using information gain.

TABLE II

CLASSIFICATION PERFORMANCE OF THE PROPOSED MODELS FOR SUBJECTIVE AND OBJECTIVE CLASSES

Method	Class	Acc	Prec	Rec	F1	AUC
Naïve Bayes	Objective	0.838	0.848	0.838	0.843	0.883
	Subjective	0.797	0.785	0.797	0.791	0.883
	Overall	0.821	0.821	0.821	0.821	0.883
SVM	Objective	0.896	0.856	0.896	0.876	0.846
	Subjective	0.797	0.85	0.797	0.823	0.846
	Overall	**0.854**	**0.854**	**0.854**	**0.853**	**0.846**
Decision Trees	Objective	0.902	0.785	0.902	0.84	0.800
	Subjective	0.667	0.834	0.667	0.741	0.800
	Overall	0.802	0.806	0.802	0.798	0.800

TABLE III

CLASSIFICATION PERFORMANCE USING SVM METHOD WITH DIFFERENT TYPES OF FEATURES

Method	Class	Acc	Prec	Rec	F1	AUC
Lexical	1-gram	0.793	0.812	0.793	0.782	0.764
	2-gram	0.807	0.807	0.807	0.805	0.795
	3-gram	0.753	0.761	0.753	0.744	0.728
	1-gram_POS	0.731	0.730	0.731	0.726	0.714
	2-gram_POS	0.754	0.753	0.754	0.753	0.746
	3-igram_POS	0.725	0.724	0.725	0.724	0.717
	MPQA	0.620	0.613	0.620	0.595	0.584
	Overll	**0.827**	**0.794**	**0.827**	**0.81**	**0.833**
Syntactical		0.576	0.554	0.576	0.524	0.528
Contextual		0.578	0.686	0.578	0.428	0.504

classification process, achieving a relatively satisfactory classification performance, with the prediction accuracy equal to 0.854, with a variance of 0.081. We also noted that the overall classification accuracy of 0.854 was much higher than the majority class baseline of 0.575, which validated the possibility of automatically detecting question subjectivity using only features extracted from the question text.

We next explored the effect of different types of features on predicting question subjectivities. In order to do that, we incorporate only one type of feature at a time to perform the experiment using the SVM method. Table III illustrates the performance of features from different perspectives. We observed that overall lexical features indicated the most discrimination power in differentiating objective questions from subjective ones. Among all lexical features, the bigram word features achieved the best classification performance. Compared with the lexical features, both syntactical and contextual features demonstrated a very limited effect on the performance of subjectivity classification. This result was further supported by the fact that none of the top ten features selected using information gain were from the syntactical or contextual aspects, as shown in Table IV.

TABLE IV

DISTRIBUTION OF THE TOP 10 FEATURES SELECTED USING INFORMATION GAIN ACROSS OBJECTIVE AND SUBJECTIVE QUESTIONS

Features	Objective (%)	Subjective (%)
good	0.00	15.78
recommend	0.00	12.37
MPQA	57.00	73.37
anyone recommend	0.00	10.15
best	0.00	9.81
adjective, superlative*	0.64	9.96
verb, 3rd person singular *	73.28	50.52
does	40.33	19.59
when	14.38	2.75
wh-determiner *	6.74	20.10

* denoting pos tagging features, while the rest are all lexical features

VI. IMPACT OF QUESTION SUBJECTIVITY ON USER BEHAVIOR

A. Subjectivity Detection

In this section, we address our second research objective by understanding the impact of question subjectivity on user's question and answering behavior. In order to do that, we first need to identify the subjectivity orientation of all 25 697 collected questions. However, as we built our classification model as a further step of providing subjectivity indication only after a question has been predetermined as informational, we cannot directly apply it to the entire data set. Therefore, to solve this challenge, we first adopted the text classifier as proposed in [5] and [9] to eliminate all noninformation-seeking tweets from our collection.

First, we implemented the informational/noninformational classification algorithm according to [9], with all its lexical, POS tagging, and meta features included. We did not take in the WordNet features as they have been proven to not be effective in predicting information questions from noninformation-seeking ones, as mentioned in [9]. We applied the classifier on our labeled data set, which included 1327 noninformation-seeking and 1261 information-seeking tweets (536 subjective information-seeking + 725 objective information seeking) and it provided us with a classification accuracy of 80.45%.

To better improve the classification performance, we further combined the features proposed in [5] to the classifier, including whether the question sentence is quoted from other sources, whether the question sentence contains strong feeling, whether there is any strong feeling such as "!" following the question sentence, and whether there is any declarative sentence following the question sentence. With these features added, the classification result improved from 80.45% to 81.66%. We think this result is reasonable compared with the 86.6% accuracy reported in [9], as Replyz has already removed a huge number of noninformational questions based on some

TABLE V
DESCRIPTION OF THE CLASSIFIED DATA SET

Question Type	Non-informational	Informational	
		Subjective	Objective
Questions	15, 311	3,984	6,402
Questioners	4,762	2,267	3,072
Answers	169,690	44,636	57,495
Answerers	87,331	28,190	33,118

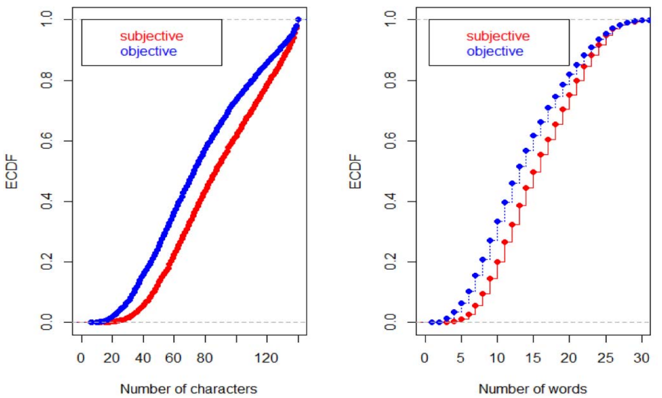


Fig. 2. Distribution of question length on character and word levels across question types.

obvious features, such as whether or not the question contains a link, a phone number, an email address, or a retweet.

The informational/noninformational classifier helped us remove 15 311 noninformation-seeking tweets form the entire collection and left us with 10 386 informational questions for subjectivity detection. We next retrained our classifier on the same subjective and objective training data, and use it to guess the subjectivity orientation of the 10 386 information-seeking tweets. Finally, we detected 6402 objective information-seeking tweets, and 3984 subjective information-seeking ones. We presented the overall statistics of our classified data set in Table V.

B. Characterizing the Subjective and Objective Questions

Based on the classified data set, we first look at the subjective and objective questions that people asked on Twitter. To be more specific, we analyzed the different ways individuals adopted to address their subjective and objective information needs. We adopted a number of statistical tests to assess the differences in question length, phrasing, and posting times across question types.

1) Question Length: Given the positive correlation reported between question length and degree of personalization in [32], we assume that subjective information-seeking questions on Twitter are longer than the objective ones. To examine the difference, we conducted Mann–Whitney U tests across the question types on character and word scales, due to the unequal variances and sample size.

In our data set, information-seeking questions asked on Twitter had an average length of 81.47 characters and 14.78 words. With the empirical cumulative distribution function of the question length plotted in Fig. 2, we observed that both the number of characters and words differ across question

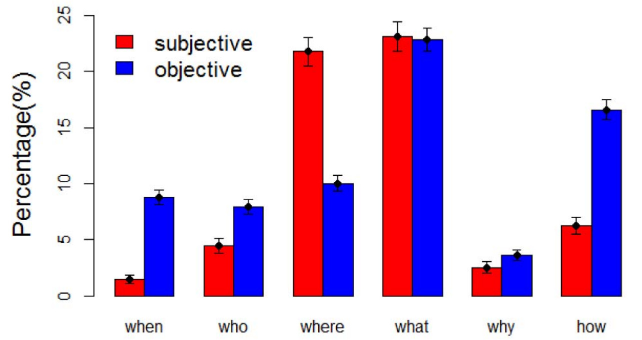


Fig. 3. Question word usage across question type with 95% confidence intervals.

subjectivity categories. Consistent with our hypothesis, in general subjective information-seeking tweets ($M_{\text{sc}} = 87$, $M_{\text{sw}} = 15.95$) contain more characters and words than the objective ones ($M_{\text{oc}} = 73$, $M_{\text{ow}} = 14.05$). Mann–Whitney U tests further proofed our findings with statistically significant p-values less than 0.05 ($z_c = -17.39$, $p_c = 0.00 < 0.05$; $z_w = -15.75$, $p_w = 0.00 < 0.05$). Through our further investigation on the content of questions, we noted that subjective questions tended to use more words to provide additional contextual information about the questioner's information needs. Examples of such questions include *So after listening to @wittertainment and the Herzog interview I need to see more of his work but where to start? Some help @KermodeMovie?, Thinking about doing a local book launch in #ymm any of my tweeps got any ideas?*

2) Question Phrasing: To explore the content difference between subjective and objective information-seeking tweets, we analyzed the question word usage in each question type and listed the results in Fig. 3, from which we saw that the question words, including when, who, and how, had high presence in objective information-seeking tweets, while question word where appeared more than twice as often in subjective questions as it did in objective ones. This result is consistent with our findings as shown in Table IV.

3) Question Posting Time: In addition to analyzing the length and phrasing differences between subjective and objective information-seeking tweets, we also examined the temporal pattern of question posting regarding its subjectivity orientation. By splitting the day time into 24 groups, one for each hour, we calculated the percentage of questions being asked in each group. The distribution of such percentage is shown in Fig. 4 (top). One can see that both objective and subjective questions were being asked the most during the normal working hours (from 8 A.M. to 5 P.M.) and the least from midnight to the early morning. We also observed that individuals asked more objective questions during free time hours (6 P.M. to 12 P.M.), while more subjective questions during normal working hours. Fig. 4 (bottom) further illustrated such temporal distribution differences between subjective and objective questions.

C. Characterizing the Subjective and Objective Answers

So far, we have only examined the characteristics of subjective and objective information-seeking questions posted

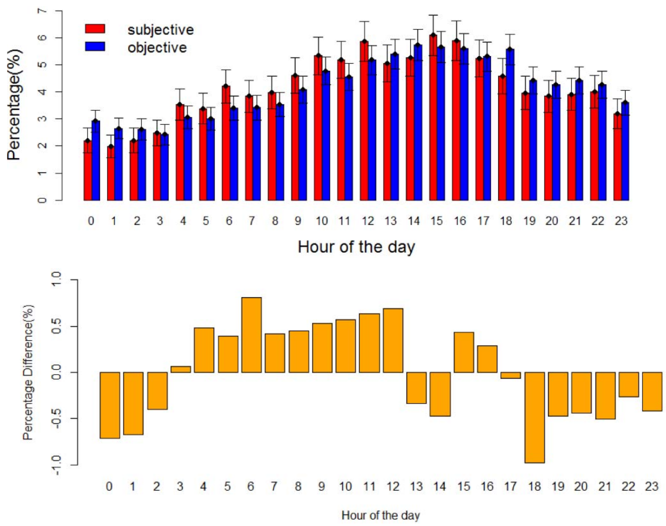


Fig. 4. Posting time distribution across question types with 95% confidence intervals.

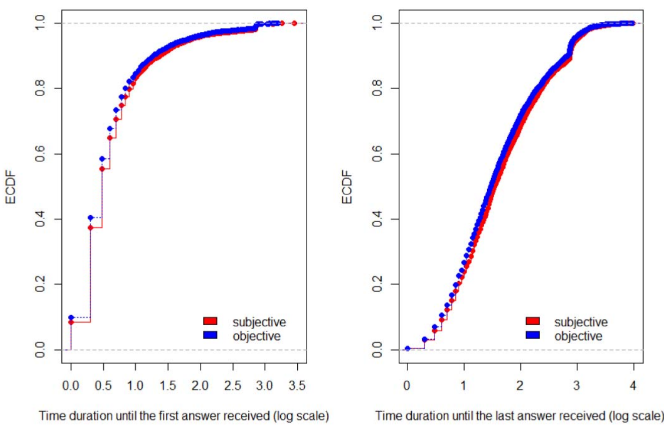


Fig. 5. Distribution of question response time in minutes on log scales across question types.

on Twitter. In this section, we presented how the subjectivity orientation of a question can affect its response.

1) Response Speed: Considering the real-time nature of social Q&A, we first looked at how quickly subjective and objective information-seeking questions received their responses. We adopted two metrics in this study to measure the response speed: the time elapsed until receiving the first answer and the time elapsed until receiving the last answer. In Fig. 5, we plotted the empirical cumulative distribution of response time in minutes using both measurements with log-transformed response time.

In general, in our data set, more than 80% of questions posted on Twitter received their first answer in 10 min or less, no matter their question types (84.60% objective questions and 83.09% subjective ones). Around 95% of questions got their first answer in an hour and almost all questions were answered within a day. From Fig. 5 (right), we observed that it took slightly longer for individuals to answer subjective questions than the objective ones. The Mann–Whitney U test result also revealed significant difference on the arrival time of the first answer between question types ($z = -3.04$, $p = 0.04$), with subjective questions on average being answered in 4.60 min after the question was posted and objective questions being answered in 4.24 min. We assumed that this might be due to the fact that subjective questions were mainly posted during working hours, whereas respondents were more active during free time hours [33]. Even though a half minute time lag may seem short and insignificant, it still has moderate practical importance given that over half of the questions collected in our data set are being answered within 2 min.

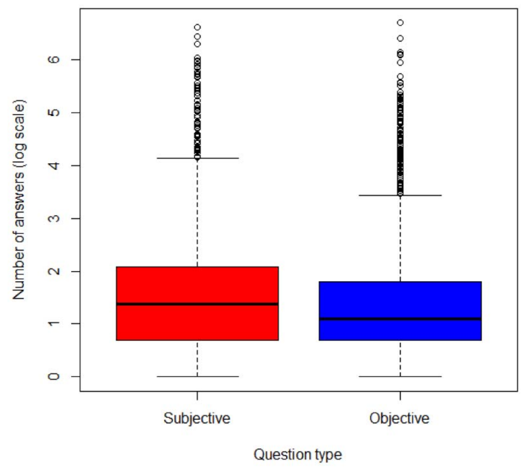


Fig. 6. Distribution of the number of answers received on log scales across question types.

In addition to the first reply, we also adopted the arrival time of the last answer to imply the temporality of each question. Defined in [12], question temporality is a measure of how long the answers provided on a question are expected to be valuable. Overall 67.79% of subjective and 69.49% of objective questions received their last answer in an hour. More than 96% of questions of both types closed in a day (96.68% objective questions and 96.16% subjective ones). Again, the Mann–Whitney U test result demonstrated significant between-group difference on the arrival time of the last answer ($z = -10.13$, $p = 0.00$), with subjective questions on average being last answered in 44 min after the question was posted and objective questions being answered in 38 min. Examples of objective questions with short temporal durations include *Hey, does anyone know if Staples & No Frills are open today?* and *When is LFC v Valarenga?*

2) Response Informativeness: The previous study [36] suggested that the amount of unique information in a question positively impact its informativeness. We assumed that the same argument applies to the question response as well. Therefore, in this section, we focus on evaluating the informativeness of responses offered in social Q&A by measuring: 1) the number of distinct answers and 2) the similarity of an answer compared with the others provided to each question. For the second measurement, only questions with more than two responses were taken into consideration.

In Fig. 6, we deployed the box plot for the number of answers received across question types. We observed that subjective questions received significantly more responses than

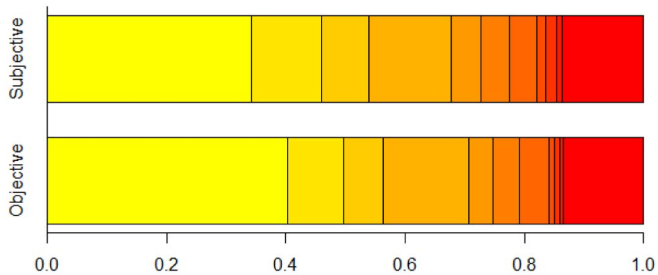


Fig. 7. Distribution of the proportion of similar answers across question types.

objective ones ($z = -7.84$, $p = 0.00$), with on average each subjective question received 9.34 responses and objective questions received 7.24 answers. This is in line with the survey nature of the subjective questions.

For the second measurement, we adopted the term frequency–inverse document frequency (tf-idf) cosine similarity in this study as the similarity measurement between answer pairs, as the method has been widely used in the field of information retrieval. Since the tf-idf cosine similarity leverages the vector space model, we used a bag of words approach, with stemming, to represent each answer using a vector of tf-idf weights. To be more specific, the cosine similarity with tf-idf weights is calculated as

$$\mathrm{sim}(A_i, A_j) = \sum_{t \in A_i \cap A_j} w_{A_i}(t) w_{A_j}(t) \quad (1)$$

where A_i and A_j are answers from distinct users to question Q_k. Here, for each distinct answerer of question Q_k, we only adopted his/her first answer provided in the calculation of response similarity. We did this to avoid the bias toward the more informal chit-chat in the later answering process. $w_{A_i}(t)$ and $w_{A_j}(t)$ are the normalized tf-idf weights for each common word t in answer A_i and A_j, respectively. The normalized tf-idf weights for a word t in a given answer A is defined as

$$w_{A_i}(t) = \frac{tf(t, A_i) idf(t)}{\sqrt{\sum_{t' \in A_i} (tf(t', A_i) idf(t'))^2}} \quad (2)$$

where $tf(t, A_i)$ denotes the frequency of word t in answer A_i, normalized by the total number of words in A_i, and $idf(t)$ is the logarithm of the total number of answers collected divided by the number of answers that containing word t.

Next, we set a threshold parameter $T = 0.75$ to evaluate the redundancy of the answers, such that only answers with tf-idf cosine similarity larger than 0.75 are considered similar. We then calculated the proportion of similar answers for each question in our data set and plotted their distributions in Fig. 7.

From Fig. 7, we observed that objective questions received more unique answers compared with the subjective ones. On average, 30.76% of all objective questions in our data set contained answer pairs with tf-idf cosine similarity larger than 0.75, which was 2.47% less than the subjective questions. The t-test result also revealed significant difference between subjective and objective questions on response similarities ($t = 3.03$, $p = 0.00$). This is consistent with the findings in [37], with objective questions attracting less but more informative responses.

TABLE VI
LOGISTIC REGRESSION ANALYSIS OF VARIABLES ASSOCIATED WITH QUESTION ANSWERING BEHAVIOR ACROSS QUESTION TYPES

Predictor	Odds Ratio	p-value
Number of followers	1.00	0.24
Number of friends	1.00	0.07
Daily tweet volume	0.99	0.00*
Friendship	1.04	0.03*

3) Characteristics of Respondents: In addition to the above two measurements, we were also interested in understanding whether the characteristics of a respondent affect his/her tendency to answer a subjective or objective question on Twitter. In order to do so, we proposed a number of profile-based factors, including the number of followers, the number of friends, and daily tweet volume, which is measured as the ratio of the total count of status to the total number of days on Twitter, and the friendship between the questioner and the respondent. Here, we only categorized questioner–answerer pairs with reciprocal follow relations as friends, while the rest as strangers.

We crawled the profile information of all respondent in our data set as well as their friendships with the corresponding questioners via Twitter API. Since our data set spanned from March 2010 to February 2014, 2998 out of 59 856 unique users in our collection have either deleted their Twitter accounts or have their accounts set as private. Therefore, at last, we were only able to collect the follow relationship between 95% (78 697) of the unique questioner–answer pairs in our data set.

We used logistic regression to test whether any of our proposed factors were independently associated with the respondent's behavior of answering subjective or objective questions on Twitter. The results of our logistic regression analysis were shown in Table VI.

From Table VI, we observed that among all four variables, the respondent's daily tweet volume and friendship with the questioner were significantly associated with his/her choice of answering subjective or objective questions in social Q&A. To better understand those associations, we further performed post hoc analyses on those significant factors.

First, as for the friendship between the questioner and the respondent, among all 78 697 questioner–answerer pairs in our data set, 22 220 (28.23%) of the follow relations were reciprocal, 24 601 (31.26%) were one way, and 31 871 (40.51%) were not following each other. The number of reciprocal-following relations in our collection is relatively low, comparing with the 70%–80% and the 36% rates as reported in [34] and [35]. We think this is because Replyz has created another venue for people to answer other's questions, even if they were not following each other on Twitter, and this enabled us to better understand how strangers in social Q&A select and answer questions.

Besides the overall patterns described, we also conducted chi-square test to examine the dependency between the questioner–respondent friendship and the answered question type. As shown in Table VII, the chi-square cross

TABLE VII
QUESTIONER–ANSWERER FRIENDSHIP ACROSS ANSWERED QUESTION TYPES

Question Type	Friendship Type	
	Friends	Strangers
Subjective	23.9% (n = 6359)	25.3% (n = 20229)
Objective	76.1% (n = 8234)	74.7% (n = 24355)

tabulations revealed a significant trend between the two variables ($\chi^2 = 13.96$, $p = 0.00 < 0.05$). We found that in real-world settings strangers were more likely to answer subjective questions than friends. This was unexpected given that [3] showed that people claimed in survey that they prefer to ask subjective questions to their friends for tailored responses. One reason for this could be that compared with objective questions, subjective questions require less expertise and time investment, so that could be a better option for strangers to offer their help.

In addition, to examine the relationship between the respondent's daily tweet volume and his/her answered question type, a Mann–Whitney U test was performed. The result was significant ($z = -7.87$, $p = 0.00 < 0.05$) with respondents to the subjective questions having more tweets posted per day ($M = 15.07$) than the respondents of the subjective questions ($M = 13.24$). This result further proved our presumption in the previous paragraph that individuals with more time spent in social platforms are more willing to answer more time-consuming questions (in our case, the objective ones).

VII. DISCUSSION AND CONCLUSION

In this paper, we investigate the subjectivity orientation of questions asked on Twitter. We proposed a predictive model based on features constructed from lexical, syntactical, and contextual perspectives using machine learning techniques. Our method achieved satisfactory performance with a classification accuracy of 84.9%. While previous work existed on similar topics [17], [19], [20], to our knowledge, our work is the first to identify question subjectivity in social context.

Using our predictive model, we extracted and analyzed 6402 objective and 3984 subjective information-seeking tweets from both the perspectives of question asking and answering. We found that contextual restrictions (e.g., time, location, and preference) were imposed more often on subjective questions, and thus made them normally longer in length than the objective ones. Moreover, through our analyses on posting and responding times, we observed that subjective questions experienced longer time lags in getting their initial answers, whereas it took shorter time for the objective questions to receive all their responses. One interpretation of this finding could be that many of the objective questions asked on Twitter were about real-time content (e.g., when will a game start or where to watch the election debates) and were sensitive to real-world events [9], so answers to those questions tended to expire in shorter durations [12]. Another possible explanation was that since answers to the objective questions were supposed to be less diverse, individuals would quickly stop providing responses after they saw a satisfactory number of answers already existing to those questions. The second interpretation is in line with our findings on less but more unique responses received by objective questions. But of course, both speculations need support from future detailed case studies. At last, in assessing the preferences of friends and strangers on answering subjective or objective questions, we demonstrated that even though individuals prefer to ask subjective questions to their friends for tailored responses [3], it turned out that in reality subjective questions were being responded more by strangers. We thought this gap between the ideal and reality imposed a design challenge in maximizing the personalization benefits from strangers in social Q&A.

In terms of design implications, we believe that our work contributes to the social Q&A field in two ways.

1) Our predictive model on question subjectivity enables automatic detection of subjective and objective information-seeking questions posted on Twitter and can be used to facilitate future studies on large scales.
2) Our analysis results allow the practitioners to understand the distinct intentions behind subjective and objective questions and to build corresponding tools or systems to better enhance the collaboration among individuals in supporting social Q&A activities. For instance, we think that given the survey nature of subjective questions and stranger's interests in answering them, one could develop an algorithm to route those subjective questions to appropriate respondents based on their locations and past experiences. In contrast, considering the factorial nature and short duration of objective questions, they could be routed to either search engines or individuals with equivalent expertise or availability.

In summary, our work is of good value to both research community and industrial practice.

We are aware of certain limitations that may restrict the ability to generalize our conclusions. One limitation is that our study is based on only one SNS, Twitter, so it may not be representative of the question asking behaviors demonstrated on other platforms. In addition, in this study, we only recruited two annotators to produce the ground truth and this may seem insufficient, even though with relatively promising inter-rater agreement. Therefore, we certainly plan to collect our ground truth data in the future with the help of crowdsourcing services, such as Amazon MTurk or CrowdFlower. For future work, we will rely on the different characteristics found in this study to automatically identify potential respondents to subjective and objective questions, respectively. In this way, after identifying a question's subjective orientation, we could then route it to the appropriate respondents to ensure its response rate and quality.

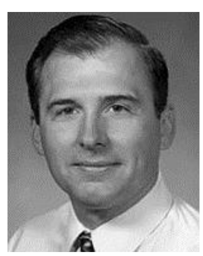
REFERENCES

[1] B. J. Jansen, M. Zhang, K. Sobe, and A. Chowdury, "Twitter power: Tweets as electronic word of mouth," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 60, no. 11, pp. 2169–2188, Nov. 2009.
[2] D. Zhao and M. B. Rosson, "How and why people Twitter: The role that micro-blogging plays in informal communication at work," in *Proc. ACM Int. Conf. Supporting Group Work*, 2009, pp. 243–252.

[3] M. R. Morris, J. Teevan, and K. Panovich, "What do people ask their social networks, and why?: A survey study of status message q&a behavior," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2010, pp. 1739–1748.
[4] Z. Liu and B. J. Jansen, "Almighty Twitter, what are people asking for?" *Proc. Amer. Soc. Inf. Sci. Technol.*, vol. 49, no. 1, pp. 1–10, 2012.
[5] B. Li, X. Si, M. R. Lyu, I. King, and E. Y. Chang, "Question identification on Twitter," in *Proc. 20th ACM Int. Conf. Inf. Knowl. Manage.*, 2011, pp. 2477–2480.
[6] M. Efron and M. Winget, "Questions are content: A taxonomy of questions in a microblogging environment," *Proc. Amer. Soc. Inf. Sci. Technol.*, vol. 47, no. 1, pp. 1–10, Nov./Dec. 2010.
[7] S. A. Paul, L. Hong, and E. H. Chi, "What is a question? Crowdsourcing tweet categorization," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2011, pp. 578–581.
[8] N. B. Ellison, R. Gray, J. Vitak, C. Lampe, and A. T. Fiore, "Calling all Facebook friends: Exploring requests for help on Facebook," in *Proc. 7th Int. Conf. Weblogs Social Media*, 2013, pp. 155–164.
[9] Z. Zhao and Q. Mei, "Questions about questions: An empirical analysis of information needs on Twitter," in *Proc. 22nd Int. Conf. World Wide Web*, 2009, pp. 1545–1556.
[10] F. M. Harper, D. Moy, and J. A. Konstan, "Facts or friends?: Distinguishing informational and conversational questions in social Q&A sites," in *Proc. SIGCHI Conf. Human Factors Comput. Syst.*, 2009, pp. 759–768.
[11] S. Kim, J. S. Oh, and S. Oh, "Best-answer selection criteria in a social Q&A site from the user-oriented relevance perspective," *Proc. Amer. Soc. Inf. Sci. Technol.*, vol. 144, no. 1, pp. 1–15, 2007.
[12] A. Pal, J. Margatan, and J. Konstan, "Question temporality: Identification and uses," in *Proc. ACM Conf. Comput. Supported Cooperat. Work*, 2012, pp. 257–260.
[13] T. Wilson *et al.*, "OpinionFinder: A system for subjectivity analysis," in *Proc. HLT/EMNLP Interact. Demonstrations*, 2005, pp. 34–35.
[14] M. Jiang and S. Argamon, "Exploiting subjectivity analysis in blogs to improve political leaning categorization," in *Proc. 31st Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2008, pp. 725–726.
[15] P. Biyani, C. Caragea, and P. Mitra, "Predicting subjectivity orientation of online forum threads," in *Proc. Comput. Linguistics Intell. Text Process.*, 2013, pp. 109–120.
[16] P. Biyani, S. Bhatia, C. Caragea, and P. Mitra, "Using non-lexical features for identifying factual and opinionative threads in online forums," *Knowl.-Based Syst.*, vol. 69, pp. 170–178, Oct. 2014.
[17] B. Li, Y. Liu, A. Ram, E. V. Garcia, and E. Agichtein, "Exploring question subjectivity prediction in community QA," in *Proc. 31st Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2008, pp. 735–736.
[18] T. C. Zhou, X. Si, E. Y. Chang, I. King, and M. R. Lyu, "A data-driven approach to question subjectivity identification in community question answering," in *Proc. AAAI*, 2012, pp. 164–170.
[19] L. Chen, D. Zhang, and L. Mark, "Understanding user intent in community question answering," in *Proc. 21st Int. Conf. Companion World Wide Web*, 2012, pp. 823–828.
[20] N. Aikawa, T. Sakai, and H. Yamana, "Community QA question classification: Is the asker looking for subjective answers or not?" *IPSJ Online Trans.*, vol. 4, pp. 160–168, Jul. 2011.
[21] M. F. Porter, "An algorithm for suffix stripping," *Program, Electron. Library Inf. Syst.*, vol. 14, no. 3, pp. 130–137, 1980.
[22] J. Wiebe and E. Riloff, "Creating subjective and objective sentence classifiers from unannotated texts," in *Proc. 6th Int. Conf. Comput. Linguistics Intell. Text Process.*, 2005, pp. 486–497.
[23] J. Wiebe, T. Wilson, and C. Cardie, "Annotating expressions of opinions and emotions in language," *Lang. Resour. Eval.*, vol. 39, no. 2, pp. 165–210, May 2005.
[24] C. Lin, Y. He, and R. Everson, "Sentence subjectivity detection with weakly-supervised learning," in *Proc. 5th Int. Joint Conf. Natural Lang. Process.*, Nov. 2011, pp. 1153–1161.
[25] A. Agarwal, B. Xie, I. Vovsha, O. Rambow, and R. Passonneau, "Sentiment analysis of Twitter data," in *Proc. Workshop Lang. Social Media*, 2011, pp. 30–38.
[26] Z. Luo, M. Osborne, and T. Wang, "Opinion retrieval in Twitter," in *Proc. 6th Int. Conf. Weblogs Social Media*, 2012, pp. 507–510.
[27] K. Toutanova, D. Klein, C. D. Manning, and Y. Singer, "Feature-rich part-of-speech tagging with a cyclic dependency network," in *Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics Human Lang. Technol.*, vol. 1. 2003, pp. 173–180.
[28] Y. Yang and J. P. Pedersen, "A comparative study on feature selection in text categorization," in *Proc. 14th Int. Conf. Mach. Learn.*, 1997, pp. 412–420.
[29] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The WEKA data mining software: An update," *ACM SIGKDD Explorations Newslett.*, vol. 11, no. 1, pp. 10–18, 2009.
[30] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on Twitter," in *Proc. 20th Int. Conf. World Wide Web*, 2011, pp. 675–684.
[31] Y. Liu, J. Bian, and E. Agichtein, "Predicting information seeker satisfaction in community question answering," in *Proc. 31st Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr.*, 2008, pp. 483–490.
[32] F. M. Harper, J. Weinberg, J. Logie, and J. A. Konstan, "Question types in social Q&A sites," *First Monday*, vol. 15, no. 7, Jul. 2010.
[33] Z. Liu and B. J. Jansen, "Factors influencing the response rate in social question and answering behavior," in *Proc. Conf. Comput. Supported Cooperat. Work*, 2013, pp. 1263–1274.
[34] S. A. Paul, L. Hong, and E. H. Chi, "Is Twitter a good place for asking questions? A characterization study," in *Proc. 5th Int. Conf. Weblogs Social Media*, 2011, pp. 578–581.
[35] P. Zhang, "Information seeking through microblog questions: The impact of social capital and relationships," *Proc. Amer. Soc. Inf. Sci. Technol.*, vol. 49, no. 1, pp. 1–9, 2012.
[36] V. Kitzie, E. Choi, and C. Shah, "From bad to good: An investigation of question quality and transformation," *Proc. Amer. Soc. Inf. Sci. Technol.*, vol. 50, no. 1, pp. 1–4, 2013.
[37] L. A. Adamic, J. Zhang, E. Bakshy, and M. S. Ackerman, "Knowledge sharing and Yahoo Answers: Everyone knows something," in *Proc. 17th Int. Conf. World Wide Web*, 2008, pp. 665–674.

Zhe Liu received the B.S. degree in management information systems from Beijing Language and Culture University, Beijing, China, in 2003, the M.S. degree in management information systems from the University of Arizona, Tucson, AZ, USA, in 2007, and the Ph.D. degree in information sciences and technology from The Pennsylvania State University, State College, PA, USA, in 2015.

She is currently a Research Staff Member with the IBM Almaden Research Laboratory, San Jose, CA, USA. She is interested in harnesses the power of computational methods to study human behavior within social contexts. Her current research interests include social analytics and computing, computer-supported cooperative work, and human–computer interaction.



Bernard J. Jansen received the master's degree in computer science from Texas A&M University, College Station, TX, USA, the master's degree in international relations from Troy University, Troy, AL, USA, and the Ph.D. degree in computer science from Texas A&M University.

He was a Professor of Information Sciences and Technology with The Pennsylvania State University, State College, PA, USA. He is currently a Principal Scientist with the Social Computing Group, Qatar Computing Research Institute, Doha, Qatar. He studies the uses and affordances of the Web for information searching and ecommerce, with a focus on interactions among the person, information, and technology. He is also a Senior Fellow of the Pew Internet and American Life Project with the Pew Research Center, Washington, DC, USA. He has authored or co-authored roughly 250 research publications, with articles in a multidisciplinary and extremely wide range of journals and conferences. His current research interests include Web searching, information retrieval, keyword advertising, online marketing, and online social networking all within the ecommerce domain.

Dr. Jansen has received several awards and honors, including an ACM Research Award and six application development awards, along with other writing, publishing, research, and leadership honors.