ASK: A Taxonomy of Accuracy, Social, and Knowledge Information Seeking Posts in Social Question and Answering

Zhe Liu
IBM Almaden Research Center, San Jose, CA, USA. E-mail: liuzh@us.ibm.com

Bernard J. Jansen
Social Computing Group, Qatar Computing Research Institute, Doha, Qatar. E-mail: jjansen@acm.org

Many people turn to their social networks to find information through the practice of question and answering. We believe it is necessary to use different answering strategies based on the type of questions to accommodate the different information needs. In this research, we propose the ASK taxonomy that categorizes questions posted on social networking sites into three types according to the nature of the questioner's inquiry of accuracy, social, or knowledge. To automatically decide which answering strategy to use, we develop a predictive model based on ASK question types using question features from the perspectives of lexical, topical, contextual, and syntactic as well as answer features. By applying the classifier on an annotated data set, we present a comprehensive analysis to compare questions in terms of their word usage, topical interests, temporal and spatial restrictions, syntactic structure, and response characteristics. Our research results show that the three types of questions exhibited different characteristics in the way they are asked. Our automatic classification algorithm achieves an 83% correct labeling result, showing the value of the ASK taxonomy for the design of social question and answering systems.

Introduction

The dramatic rise in the use of social networking sites has introduced a new way for people to communicate and stay connected with each other online. Hundreds of millions of new posts are made and shared daily on social networking sites (SNS) among users (Twitter, 2011). This makes the social networking platform a good place for both information broadcasting and seeking (Jansen, Zhang, Sobel, & Chowdury, 2009; Kwak, Lee, Park, & Moon, 2010). Although not intentionally designed for questioning and answering, people enjoy expressing their information needs in natural language questions on SNS (Paul, Hong, & Chi, 2011b), referred to as social questioning and answering (social Q&A). By making use of online interactions (Evans & Chi, 2008), social Q&A can surpass traditional information-seeking techniques (e.g., search engine and online databases, etc.) in certain contexts with a more simplistic and personalized search experience.

Because of its potential impact, social Q&A has generated interest among academic researchers. Research has revealed the diverse information needs in social Q&A and has made significant contributions to the theoretical understanding of information seeking within the social context (Efron & Winget, 2010; Lampe, Gray, Fiore, & Ellison, 2014; Liu & Jansen, 2012; Morris, Teevan, & Panovich, 2010b; Paul et al., 2011b). Considering the varied information needs demonstrated, we believe that it is necessary to use different strategies to get answers for different types of questions asked in social Q&A. For instance, retrieving similar answers through archived question-answer pairs may be effective and efficient in answering questions such as "How do I update to IOS 8?" However, for questions similar to "What should I say when asking her out for a meal?" computer-generated answers obviously may not satisfy the questioner's expectations.

To better understand the nature of questions posted in social Q&A and to provide corresponding answering strategies, in this research we propose a taxonomy that differentiates questions into three primary types based on the underlying informational intent of the questioner. As a memory aid, we propose the acronym ASK for this taxonomy to clarify and categorize questions within the social media domain.

Received February 2, 2015; revised August 26, 2015; accepted August 27, 2015

 • Published online 2 March 2016 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/jasist.23655

- ***A**ccuracy* questions, in which people ask for *facts* or *common sense* responses (e.g., "When is the game of Clippers and Heat?" "I lost my phone how to restore my game?");
- ***S**ocial* questions, in which people ask for *coordination* or *companion* responses that are focused on the views of other people in some joint context (e.g., "Who wants to join me at the #gym today?" "Who's going to North Haledon day?");
- ***K**nowledge* questions, in which people ask personal *opinions* or *advice* responses (e.g., "I'll be in Cape Town in time for supper, which restaurant should I try out?" "What should I do for my birthday?").

To automatically decide which answering strategy to use for a given question, we develop a predictive model to differentiate these three categories of questions using features derived from the content of the question, including lexical, topical, contextual, and syntactic information, as well as features extracted from the answers received. We carry out the experiment of our classification model using questions posted on Twitter.

In total, approximately 25,000 question tweets are collected for this study. From those 25,000 questions, we randomly select 3,000 and manually label each of them as accuracy, social, and knowledge. We implement and evaluate multiple classification algorithms with a combination of proposed features. Our classification model proved to be reliable in distinguishing the three types of ASK questions with a classification accuracy of 83.20%.

We consider this research important for both practitioners and academics in the area of social Q&A. First, we introduce the ASK taxonomy for questions posted in social Q&A to deal with the varied needs of the information seekers on SNS. We believe that the ASK taxonomy serves as a step in designing and developing better social Q&A tools via adopting different answering strategies for different types of questions. Second, we propose an automatic method for determining the types of questions asked on SNS. By automatically differentiating questions according to the nature of the questioner's inquiry, ASK can be adopted to decide whether the question can be responded to with computer-generated answers as well as to determine the most appropriate responses.

Literature Review

Social Q&A

Defined as the process of discovering information online with the assistance of social resources (Morris, Teevan, & Panovich, 2010a), social Q&A lies between the boundaries of technical and human-powered information-seeking models. Examples of actual social Q&A questions include: "Anyone knows how to fix blinking monitor?"; "Can anyone recommend any good places to go for afternoon tea in central London?"; and "#healthadvice Twitter I need help - how can I kick a cold/flu illness quickly?" By making use of all possible social interactions online, social Q&A rivals traditional information-seeking techniques (e.g., search engine and online databases, etc.) with more personalized search experience. Jansen et al. (2009), 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, whereas 18.1% were information seeking. Li, Si, Lyu, King, and Chang (2011) revealed that about 11% of general tweets contained questions, which is similar to the 13% reported in Efron and Winget (2010). So it appears that about 10% of social media posts related to information seeking.

To better understand the underlying motivation behind social Q&A, Morris et al. (2010a) confirmed that seeking information on social networks can provide more personalized responses and with higher answer quality. Morris et al. (2010b) also surveyed 624 social network users' reasons for choosing social networks as the platform for Q&A. Their results indicated that people search socially primarily because of their trust in friends versus strangers. Also, beliefs in weak search engine performances and nonurgent information needs also accounted for the reasons of why people seek information on social networks. Teevan, Collins-Thompson, White, Dumais, and Kim (2013) analyzed a large-scale of query logs and examined how individuals value time when searching via user surveys and found that speed is important in online information seeking and categorized individual's information needs according to task urgency. Through their survey on individuals' social Q&A behavior across countries, Yang, Morris, Teevan, Adamic, and Ackerman (2011) report that the main reasons for engaging in social Q&A were that information obtained from one's social network was, again, more trusted, of higher quality, and more personal. Respondents also indicated that culture played a role in people's motivations to ask their social networks.

As the popularity of social Q&A has grown, several corresponding applications have been built to help responding to questions posed on SNS. Bozzon, Brambilla, and Ceri (2012) described CrowdSearcher, which works by transforming natural language queries into logical paradigms, such as, like, recommend, tag, rank, merge, and correct, and so on. Relying on the power of human suggestions from social platforms, CrowdSearcher improves the quality of search results in complex, exploratory information-seeking tasks. Horowitz and Kamvar (2010) presented a social search engine called Aardvark. To attempt to understand the user's information need, the system runs a set of classifiers to determine both the quality and the topic of the question. Based on the similarity of the inquiry topic and the interests of the potential answerers, Aardvark automatically routes the question to the person who is most likely to answer it. Hecht, Teevan, Morris, and Liebling (2012) developed a prototype system, SearchBuddies, that responds to questions posted on Facebook with algorithmic search results. SearchBuddies deploys two approaches to respond to people's questions: (a) Investigator, which returns direct answers found via major search engines, and (b) Social Butterfly, which identifies topics in a status message question and finds friends of the question asker who have expertise on the

DOI: 10.1002/asi

mentioned topics. Yan, Kumar, and Ganesan (2010) presented an accurate real-time image search system for mobile phones called CrowdSearch that combines automated image search with human validation of search results using Amazon Mechanical Turk. By automatically predicting the response delay of each validation task, CrowdSearch determines which results need to be validated and when and how to validate them.

Taxonomy of Questions in Social Q&A

In addition to motivations, other research has been conducted to understand the taxonomy of questions asked on SNS. Through an analysis of 100 question tweets, Efron and Winget (2010) reported that Twitter users use social networks to satisfy their information needs by asking both factual and personal opinion questions to their friends online. Evans and Chi (2008) conducted their study using the transactional, navigational, and informational taxonomy of traditional search, also used in Jansen, Booth, and Spink (2008) for automatic query classification. Kim, Oh, and Oh (2007) classified questions from Yahoo!Answers into four categories: information, suggestion, opinion, and other. They pointed out that the criteria of selecting the best answer differed across categories. In their study, Morris et al. (2010b) 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. Differently, Paul et al. (2011b) noticed more rhetorical (42%) questions on Twitter, followed by the categories of factual knowledge (16%) and polls (15%). Ellison, Gray, Vitak, Lampe, and Fiore (2013, p. 5) labeled a set of 20,000 status updates on Facebook and presented multiple types of mobilization requests, defined as "requests for actions related to provisions of social, informational, or other forms of support or assistance." Among the previous works on question typology in a social context, Harper, Weinberg, Logie, and Konstan (2010) proposed a comprehensive classification of questions. By coding questions drawn from three popular community Q&A sites, the authors developed a typology of question types that fall into three types of rhetoric, including deliberative, which is future focused (e.g., "What's the next band you want to see get a Rock Band 'special edition'?"); epideictic, which is present focused (e.g., "What's your favorite Beatles song?"); and forensic, which is past focused (e.g., "Will my controllers for the Wii version of 'Guitar Hero' also work on the Wii version of 'The Beatles: Rock Band'?). Gazan (2011) also proposed a need for more appropriate categories of questions that maximize the likelihood of receiving answers, and he suggested that the early introduced taxonomic structures and categorization strategies for social Q&A could be more refined by testing them against users' actual categorization behavior.

Automatic Question Classification

Most of the above-mentioned studies performed the question classification task manually based on handcrafted rules. There are also a few studies that touch on the problem of automatic question classification based on machine-learning techniques.

The first stream of studies in automatic questions classification focuses on distinguishing questions with real information needs from those without. Harper, Moy, and Konstan (2009) successfully distinguished conversational questions, which are defined as questions that are intended simply to start discussion, from the informational ones. As a result of the analysis, the authors claimed that conversational questions typically have much lower potential archival value than the informational ones. Li et al. (2011) proposed a cascade approach to identify questions with real information needs. The authors first relied on both rule-based and learning-based approaches to detect interrogative tweets, with both conversational and informational intents. Then they further distinguished the two types of questions using a set of question- or context-related features. Zhao and Mei (2013) classified question tweets into two categories: tweets conveying information needs and tweets not conveying information needs. They manually labeled 5,000 tweets and built an automatic text classifier based on lexical, part-of-speech (POS) tagging, and meta-features. With the classifier, they investigated the temporal characteristics of those information-seeking questions.

On the other hand, studies have been conducted on the topic of question subjectivity identification. Li, Liu, Ram, Garcia, and Agichtein (2008) labeled 987 resolved questions from Yahoo!Answers and explored a supervised learning algorithm using features extracted from both the questions and the associated answers to predict the subjectivity of a question. Zhou, Si, Chang, King, and Lyu (2012) proposed an approach to automatically collect training data based on social signals, such as like, vote, answer number, etc., in community question and answering (community Q&A) sites. The results of their experiment demonstrated that leveraging social interactions in community Q&A portals could significantly improve the prediction performance. Chen, Zhang, and Mark (2012) classified questions from Yahoo!Answers into subjective, objective, and social. They built a predictive model based on both textual and meta features and cotrained them. Their experimental results showed that cotraining worked better than simply pooling these two types of features together. Aikawa, Sakai, and Yamana (2011) employed a supervised approach in detecting Japanese subjective questions in Yahoo!Chiebukuro. Unlike the other studies, they evaluated the classification results using weighed accuracy, which reflected the confidence of annotation.

Our work is a further step of the above-mentioned studies in the direction of understanding and comprehending the questioner's information needs in social Q&A. Our work differs in that we develop a practical and nuanced taxonomy

DOI: 10.1002/asi

of social Q&A and combine this foundational work with automatic classification, with evaluation.

Research Objectives

With the ultimate goal to better understand the nature of questions in social Q&A to choose the most appropriate answering strategies for different types of questions, we establish two research objectives.

1. Develop a taxonomy of questions posed in social Q&A that could be used to assist in selecting the most appropriate answering strategies.

For our first research objective, we develop the ASK taxonomy to serve as the theoretical groundwork for the types of questions posted in social Q&A based on the nature of the questioner's inquiry. The ASK taxonomy addresses three general types of information needs: Accuracy, Social, and Knowledge. Unlike previous studies (Efron & Winget, 2010; Ellison et al., 2013; Gazan, 2011) that were primarily descriptive, we believe our proposed taxonomy can enhance question answering on SNS because categorizing questions into ASK types can result in higher response probability and quality via selecting different answering strategies for different types of questions by automatically categorizing question types.

The hypothesis of the proposed taxonomic system is that accuracy questions tend to be objective and can be answered by machines (i.e., search engines, archived similar question–answer pairs), whereas in contrast the other two types of questions, social and knowledge, require more diverse replies that rely on personal experiences and opinions and are preferably answered by human respondents. In addition, compared with the accuracy and knowledge questions, social questions tend to be more "acquaintance oriented." So to answer social questions, the social Q&A system needs to count more on the questioner's own friends or followers on SNS. For knowledge questions, the system can route the inquiries to a larger audience base for more relevant answers.

2. Implement the proposed taxonomy by automatically classifying questions into ASK types and measure the effectiveness of the classification.

To achieve the second research objective, we explore the distinction among ASK questions in the way they are being proposed and answered. The features that we use for classification are from both the question and the answer contexts, and they can be divided into five categories, as listed below:

- Question Features
  - **Lexical** – the words, expressions, and vocabulary used in the posted question (e.g., word frequency and part of speech tagging)
  - **Topical** – the major subject(s) of the posted question
  - **Contextual** – the situational elements related to the proposed question (e.g., spatial location and temporal context)
  - **Syntactic** – the structure of the vocabulary of the proposed question (e.g., question length in clauses, words, and characters, and whether the question contains an image)
- **Answer Features –** the attributes of answers received in response to a question (e.g., such as the number of answers received, the length of answers, the number of unique answerers, and whether an answer contains an URL)

We then build a prediction model that can reliably distinguish the three types of questions using machine learning algorithms. We evaluate our model using a large sample of questions collected from Twitter.

Methods

Data Collection

To answer our proposed research objectives, using Twitter API, we collect tweets written in English that were posted from September 20 to October 1, 2014 containing at least one question mark and any of the 5WHC questions words, including who, where, when, what, why, how, and can. Given the percentage of information-seeking questions on Twitter (6% as reported in Li et al. [2011]) and the scope of this study on informational questions only, we further constrain our search query by including a general set of question-signaling hashtags, as introduced in Rzeszotarski, Spiro, Matias, Monroy-Hernández, and Morris (2014), to filter out as many conversational or non-information-seeking questions (Harper et al., 2009) as possible. Only questions containing at least one of those hashtags were included in our data set. We also remove retweets and questions directed to specific users (tweets with @username) considering the lack of necessity of question routing. This left us with a total of 23,258 "information-seeking" tweets, which counts for 71.94% of the original set (32,330 total tweets originally collected).

Taxonomy Creation

To address our first research objective, we develop a taxonomy that differentiated questions in a way that is beneficial for developing social Q&A tools, such as question routing, answer ranking, and summarization. Our goal is to create a taxonomy that serves two purposes. First, we hope to articulate generalities that pertain to the diversity of Twitter questions. Second, we build a taxonomy that would identify types of questions that would benefit from further analysis, such as information retrieval, visualization, or routing. Thus, we want our taxonomy to be both descriptive and actionable. To be more specific, we create the ASK taxonomy considering the questioner's intent from two dimensions—(a) subjective or objective and (b) personal or impersonal—and categorized the questions into three types—accuracy, social, or knowledge—according to both dimensions. We assume that most accuracy questions seek objective and impersonal information, such as asking for a fact, common knowledge, or how to do something. Knowledge questions look for subjective responses, both personal and impersonal, such as opinions, recommendations, or open discussions. Social questions are

DOI: 10.1002/asi

TABLE 1. Annotation criteria for ASK question types and non-information-seeking questions, along with examples of each question type.

Question type	Annotation criteria	Sample questions
Accuracy	• fact • definition • how-to method	• *When is the debate on UK time? #replytweet* • *Hey gamers. Anyone know how to turn off motion controls for Hyrule Warriors? #HyruleWarriors #wiiu #controllerproblems #help* • *Anyone know how to say "fun smasher" in Spanish? #help*
Social	• invitation • favor • coordination • companion	• *Is there somebody who has an ELLO invitation for me? #dtv #Question* • *Who on lemoyne campus has a 5 charger? #replytweet* • *Um somebody want to teach me what you learned in algebra 2 yesterday? #help* • *Who's going to the game tomorrow? #replytweet*
Knowledge	• recommendation • opinion	• *I hate when the bf has not texted me in five days help guys what should I do? #help* • *How should I get my nails done for homecoming?? #replytweet* http://t.co/Lj6xpsR1IG • *I'm in desperate need of a good book series to read. What would you guys recommend? #replytweet*
Non-Information-Seeking	• rhetorical question • spam	• *I don't really talk to you guys much and it makes me feel bad and antisocial that I don't. Why is that? #ReplyTweet #DMMe* • *How is it possible to have this much workload during the first month of school? #overloadincredits #help* • *Why does this product suck so much? #ReplyTweet*

more personal-oriented with the purposes of seeking private companion and coordination. Social questions can be classified as either subjective or objective in nature. By differentiating questions into those three types, the ASK taxonomy can be used to determine whether the collected answers will be ranked based on authority or summarized for quick digestion (i.e., objective vs. subjective). Also, ASK can be used to route questions to the appropriate respondents according to their relationship to the questioner (i.e., personal vs. impersonal).

Compared with the existing taxonomies of question types in social Q&A, we believe that the ASK taxonomy has the following advantages: First, ASK is more simplified because it contains only three different types of questions, whereas the taxonomy introduced by Morris et al. (2010b) contains eight categories, and the one proposed by Efron and Winget (2010) contains nine distinct question types. Automatically classifying questions according to these existing taxonomies could be difficult. Second, ASK is more applicable to the ultimate goal of classifying questions for routing purpose. As mentioned earlier, we assumed that accuracy, social, and knowledge questions have different subjectivity requirements and distinct personal relevance so that they could be routed to different audience for quicker and better responses.

To evaluate the proposed ASK taxonomy, we randomly sampled 3,000 questions from our collected data set and recruited two human annotators to perform the labeling task based on our annotation criteria for accuracy, social, and knowledge questions. Each category was mutually exclusive. We selected 3,000 (12.9% of the working data set) as a number to enable manual classification, while also providing a meaningful training set.

To guide the annotation process and to promote continuity between human annotators, the following annotation criteria were adopted:

- **Accuracy Question**: The aim of an accuracy question is to receive answers based on some factual or prescriptive knowledge. The purpose of the question is to receive one or more correct answers instead of responses based on the answerer's personal experience. This type of question usually looks for facts, definitions, and prescriptive methods on how to do something.
- **Social Question**: The aim of a social question is to request either companionship or coordination support from others. It includes questions searching for someone who shares the same agendas or for someone who can provide physical or emotional assistance.
- **Knowledge Question**: The aim of a knowledge question is to receive responses reflecting the answerer's personal opinions, advice, preferences, or experiences. It usually has a survey purpose that encourages the audience to provide personal answers.

In addition, considering there are tweets that appear as questions but do not convey any real information need, we also asked the annotators to tag those questions because they do not belong to the ASK types. These posts are presented as questions with no information-seeking purpose but instead are information providing or rhetorical.

To better illustrate the ASK taxonomy proposed in this study, Table 1 lists a number of sample questions with accuracy, social, and knowledge types. The accuracy questions generally refer to some particular things (e.g., the game questions refer to a particular how-to method, specifically the game controls). The social questions typically have a people focus (e.g., the algebra question refers to the person, asking him or her for a favor). Knowledge questions typically have a blend of both (e.g., the nails questions refers both to act of getting nails done and a person's opinion of doing it at this time).

With the above annotation criteria, the two human annotators worked on the labeling separately. There were 2,621 of 3,000 questions (87.37%) agreed on by the two independent coders, indicating the relatively high reliability of the ASK taxonomy. Among the 2,621 question tweets,

- 1,253 (47.81%) were labeled as having a non-information-seeking intent or mixed-information-seeking intents,

DOI: 10.1002/asi

- 475 (18.12%) as accuracy seeking,
- 112 (4.27%) as social seeking,
- 781 (29.80%) were labeled as knowledge seeking.

We also examined the 379 questions with annotation differences and found that the major cause of such disagreement is that, without knowing the very specific context of a question, annotators interpreted the questioner's intent differently. For instance, for the question, *"Ok as we start to evaluate #Obama legacy who is worse him or #Carter? #QuestionsThatNeedAnswers,"* one annotator tagged it as knowledge question as it surveys the audience about their opinions regarding Obama, whereas the other treated this tweet as sarcasm and tagged it as conversational. For our research questions, we used only the 475 accuracy, 112 social, and 781 knowledge questions.

Question Classification Using ASK Taxonomy

To accomplish our second research objective, in this section, we present our design of a model for the automatic prediction of accuracy, social, and knowledge questions posted on Twitter.

Feature Extraction

First, we introduce the number of features extracted for the purpose of question classification. In total, we identify five different attributes for questions:

- Lexical Attributes
- Question Attributes

Given the different information needs, there should be a different usage of lexical terms among questions of distinct types, so we included lexical features that operate at a word level, including both part-of-speech-tagging-based (POS-tagging) (Schmid, 1994) and n-gram ($n = 1, 2, 3$) (Jelinek, 1990) patterns for each question. POS-tag means to tag each term as a noun, verb, adverb, and so on.

We adopt word-level n-gram features by counting the frequencies of all unigram, bigram, and trigram tokens that appear in the training data, as they have proven useful in previous work (Aikawa et al., 2011; Sadilek, Kautz, & Silenzio, 2012; Zhao & Mei, 2013). Before feature extraction, we lowercased and stemmed all of the tokens using the Porter stemmer (Porter, 1980). We discard rare terms with observed frequencies of less than five to reduce the data's sparsity. This data processing left us with 996 n-gram features.

In addition to the lexical features, we believe that POS-tagging may also help distinguish the three ASK types of questions, as such annotation can provide more context to the words used in the interrogative tweets. To tag the POS of each tweet, we adopt the Stanford tagger (Toutanova, Klein, Manning, & Singer, 2003). Again, we count the frequencies of all unigram, bigram, and trigram POS patterns that appeared in the training data. In total, we extract 664 POS-tagging features. This results in a total of 1660 lexical attributes.

- Topical Attributes

The topical features relate to the subject categorization of each question. The discriminative effect of question topic on questioner's intent has been quantitatively shown by a number of previous studies (Chen et al., 2012; Harper et al., 2009; Li et al., 2008; Liu & Jansen, 2013). To test its power on classifying ASK types of questions, we ask the two human annotators to manually code the 1,368 informational questions into 15 categories obtained from Yahoo! Answers, including "Automobile," "Beauty & Style," "Business," "Family & Relationship," "Education," "Entertainment," "Food," "Health," "Home & Garden," "Pets," "Society & Culture," "Sports," "Technology,"Travel," and "Word & Reference." We derive the 15 categories by combining the similar topical groups such as "Dining Out" and "Food," "Pregnancy & Parenting" and "Family & Relationship," "Computers & Internet" and "Consumer Electronics," etc., together. We also add the topical group "Word & Reference" for questions such as "How should I pronounce 'Skadi'?" The interrater agreement between the two annotators was 86.11%, which is quite high given 15 possible categories.

- Contextual Attributes

For the contextual features, we focus on the two aspects of temporal and spatial attributes of a question. Liu and Jansen (2012) indicated the high percentage of questions containing either explicit temporal or spatial indicators and claimed that contextual features play important roles in social Q&A. Other studies (Liu, Alexandrova, & Nakajima, 2013; Liu & Jansen, 2013; Wang, Chen, Hou, & Chen, 2014) further verified the feasibility and advantages of building real-time social Q&A services on top of microblogging platforms.

Extracting Spatial Attributes. To identify the spatial indicators of a question tweet, we use a service called Alchemy API (http://www.alchemyapi.com/), which is a text mining platform based on natural language processing and machine learning algorithms. The Alchemy API allows one to retrieve an entity's information, such as people, companies, organizations, geographic names, and son on from a paragraph of input text. It has been adopted in a number of previous studies for text mining tasks (Batool et al., 2012; Quercia, Capra, & Crowcroft, 2012). For the purpose of spatial extraction, we only extract entities with the types of "City," "Continent," "Country," "Facility," and "Region."

Extracting Temporal Attributes. For temporal extraction, we adopt the Stanford temporal tagger (SUTime) (Chang & Manning, 2012). SUTime is a library for recognizing and extracting date, time, and duration values. It is rule-based

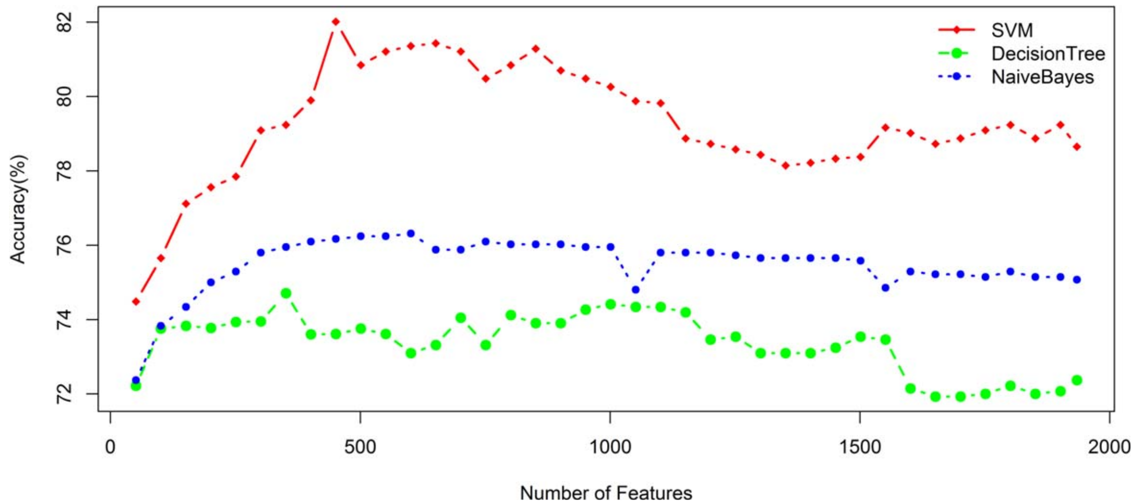

FIG. 1. Lexical feature selection for ASK taxonomy using SVM, Decision Tree, and Naïve Bayes. [Color figure can be viewed at wileyonlinelibrary.com]

and extracts explicit temporal expressions based on regular expressions. SUTime exhibits stronger performance compare with other temporal taggers, like HeidelTime (Strötgen & Gertz, 2010) for English text (Sil & Cucerzan, 2002). With SUTime, we identified a list of temporal expressions, such as "today," "this week," "2:30am," "Halloween," "tomorrow night," etc., in our data set.

For the contextual attributes, we replace the extracted exact temporal expressions with <TIME> and spatial expressions with <LOCATION> in our data set and rebuild the classification model with these more general contextual factures.

- Syntactic Attributes

The syntactic features measure the writing style of the question at the sentence level and above. The syntactic features that we adopt in this study include the length of tweets in sentences/clauses, words, and characters, and whether or not the tweet contains a picture. To identify tweets containing pictures, we expand all shortened URLs through a website called LongURL (http://longurl.org/).

- Answer Attributes

As mentioned earlier, different types of questions may lead to answers with distinct characteristics, so we include the answer features that record the total number of answers received, the unique number of answerers, and the answer length in words for each question.

In addition, we hypothesize that questions of different types require different ranges of people to answer. For instance, we believe that even strangers can answer most of the accuracy questions, whereas social questions usually are implicitly restricted to online friends/followers only; therefore, we retrieve the relationships between the questioners and answers and include whether or not a stranger answers the question as a feature in our classification model. We also analyze features including: the length of the answer, the number of overlapped terms in all answers, the arrival time of the first answer, the arrival time of the last answer, and whether or not an answer contains an URL, which are all features that indicate the interaction aspects.

Classification Algorithms and Evaluation Metrics

With the above features, we next build a multiclass classifier to automatically label questions into the three ASK types. We train and test our model using a number of classification algorithms implemented in Weka (a data mining software package) (Hall et al., 2009), including Naïve Bayes (John & Langley, 1995), Support Vector Machines (SVM) (Cortes & Vapnik, 1995), and Decision Trees (Quinlan, 1987). More precisely, we implement SVM using the Sequential Minimal Optimization (SMO) (Platt, 1999) and Decision Trees with J48 (Quinlan, 2014) using 10-fold cross-validation for each classification experiment.

For evaluation purposes, we use the traditional metrics, including: precision, recall, F1 (a metric that measure both recall and precision), and accuracy as they have also been adopted in many other studies (Castillo, Mendoza, & Poblete, 2011; Liu, Bian, & Agichtein, 2008; Zhou et al., 2012).

To evaluate the proposed classifier, we compare our model with a majority vote baseline (Vossen, 2012). The majority vote method assigns any newly observed data point (i.e., question in our case) to the type that is most common, yielding a predictive accuracy of 57.09% in our case. The proposed classifier is expected to score at least higher than this baseline.

Results

In this section, we address our second research question by understanding the classification results of question types.

DOI: 10.1002/asi

TABLE 2. Classification performance for SVM, Decision Tree, and Naïve Bayes classifiers using lexical features.

Method	Precision	Recall	Accuracy	F1
SVM	0.823	0.820	0.820	0.821
Decision tree	0.748	0.747	0.747	0.747
Naïve Bayes	0.768	0.763	0.763	0.765

TABLE 3. Top 10 n-gram features of each ASK question type along with the information gain associated with each n-gram feature.

Word	Accuracy (%)	Social (%)	Knowledge (%)	InfoGain
Who	7.98	90.90	6.26	0.21
I	12.18	10.00	48.34	0.12
What	45.59	2.73	49.49	0.06
Should	0.63	0.00	13.68	0.06
Good	0.00	6.36	11.64	0.05
Who is going	0.21	15.45	0.00	0.04
Where	7.77	1.82	22.25	0.04
What time	6.72	0.00	0.00	0.04
Me	4.20	29.09	5.63	0.03
Want	0.63	18.19	3.20	0.03

Classification Results Using Lexical Features

Because of the large number of lexical features extracted, we evaluate the classification accuracies along the number of features selected. Figure 1 shows the feature selection results using the algorithm of information gain (Yang & Pedersen, 1997) implemented in Weka. Information gain measures the variation in entropy when a feature is present versus when it is absent and is frequently employed as a feature selection method in the field of machine learning.

As seen in Figure 1, either too few or too many features result in a decrease of prediction accuracy. In addition, SVM outperforms the other two methods in the question classification process using lexical features with 0.820 accuracy, 0.823 precision, 0.820 recall, and 0.821 F1-measurement. We also note that the accuracy of 0.820 was much better than the majority class baseline of 0.571, which validated the possibility of automatically detecting question types using lexical features only. Table 2 presents the classification results of all three methods.

We adopt the method of information gain to identify the most informative and relevant features of each question type. Table 3 shows the top 10 discriminative word features. From Table 3, we notice that about half of the accuracy and knowledge tweets contain the question word "what"; whereas 90% of the social questions started with the word "who." Example accuracy and knowledge questions containing "what" and social questions containing "who" include "What does #tsibip mean? #twoogle," "what is a good music downloader app? #replytweet," and "who is going to the eagles tomorrow and wants tailgate? #ReplyTweet."

We also discover that, when compared with the other two types of questions, knowledge-seeking tweets ask more about locations. These questions tend to include more

TABLE 4. Topical categories, percentage for informational questions that have been classified, and the fraction of ASK questions within each topical category.

Question topic	Total (%)	Example question	Accuracy (%)	Social (%)	Knowledge (%)
Technology	22.15	*Can anyone help me figure out why Im not able to load any links from a tweet? A "loading error" keeps popping up #help*	43.23	3.30	53.47
Entertainment	16.45	*Does anyone know when Vampire Diaries starts?!?! #replytweet*	33.33	10.23	56.44
Beauty & style	13.96	*Ok ladies I need help!! Navy dress…what color shoes? I have red, fuchsia, silver, gold…no nude though…#replytweet*	5.76	4.71	89.53
Education	9.72	*Anyone in my law class know what part three of the assignment is???? #replytweet*	53.38	18.05	28.57
Sports	7.75	*What time is the soccer game tonight?? #replytweet*	63.21	19.81	16.98
Food	6.29	*Am hungry but i do not know where to go!! Any suggestions?? #help*	11.63	3.49	84.88
Society & culture	4.02	*What is y'all's opinion on breast feeding in public? #AskTwitter*	67.27	1.82	30.91
Travel	3.66	*Really want to visit the #MiddleEast next year, as a #Brit, where would people recommend? #UAE #SaudiArabia #Dubai #Qatar #Kuwait? #help*	24.00	20.00	56.00
Family & relationship	3.58	*How do guys usually act around a girl they're interested in? Trying to decode someone. #ReplyTweet*	0.00	0.00	100.00
Business	3.29	*How do I get a job at Ukhozi FM? #Twoogle*	37.78	17.78	44.44
Health	3.07	*What should I take for a sore throat & a runny nose? #replytweet*	4.76	2.38	92.86
Word & reference	2.41	*What is #LRT? #twoogle*	96.97	0.00	3.03
Home & garden	2.19	*How do you clean washing machines? Mine smells a bit musky! #housework #help*	13.33	6.67	80.00
Pets	0.88	*I am adopting myself a dog… where is the best place to get a rescue? #asktwitter*	41.67	0.00	58.33
Automobile	0.58	*Does any1 in Tampa kno where to get their car detailed?? ? #replytweet*	12.50	12.50	75.00

DOI: 10.1002/asi

TABLE 5. Classification performance for SVM, Decision Tree, and Naïve Bayes classifiers using topical features.

Method	Precision	Recall	Accuracy	F1
SVM	0.607	0.667	0.667	0.622
Decision tree	0.604	0.664	0.664	0.620
Naïve Bayes	0.607	0.667	0.667	0.622

contextual information, using the word "I" in 48.34% of 782 cases. Typical knowledge questions are: "If I ever were to replace my beloved Pentax K200D (and I will have to), what should I get next? Another Pentax? Canon? Nikon? #question."

Classification Results Using Topical Features

In addition to looking at the lexical features, we also examine the topical categories annotated by the two human coders. We divide the 1,368 informational-seeking questions into 15 categories according to their topics. Table 4 shows the 15 topics and the percentage of questions belonging to each category. The topical categorization results are consistent with previous studies (Liu & Jansen, 2012; Morris et al., 2010b), suggesting a predominance of "Technology," "Entertainment," and "Education" questions; however, we also notice a large percentage of "Beauty & Style" related questions in our data set with the questioners asking their friends for opinions on their appearance that were not discussed in previous work. This is consistent with the findings in Shah's work (Shah, 2011), as these are common life concerns and more reasonably directed to a person than possibly other question topics.

While analyzing the percentage of accuracy, social, and knowledge questions within each topical category, we identify that the majority of "Beauty & Style" (89.53%), "Food" (84.88%), "Health" (92.86%), "Family & Relationship" (100%), and "Home & Garden" (80.00%) questions sought other's opinions and advice, whereas 96.97% of "Word & Reference," 63.21% of "Sports," 67.27% of "Society & Culture," and 53.38% of "Education" questions were of an objective nature.

Table 5 illustrates the classification results using topical categories. We notice that although the topic-based classifier outperformed the baseline, its performance was much weaker than the lexical-based model.

TABLE 6. Distribution of ASK question types with and without contextual mention.

With or without contextual mention	Accuracy	Social	Knowledge
With location	27 (5.68%)	15 (13.39%)	65 (8.32%)
Without location	448 (94.32%)	97 (86. 61%)	716 (91.68%)
With time	64 (13.47%)	38 (33.93%)	107 (13.70%)
Without time	411 (86.53%)	74 (66.07%)	674 (86.30%)
With both	3 (0.64%)	7 (6.25%)	15 (1.92%)

TABLE 7. Top five question topics by percentage of location mention.

Question topic	With location	Without location
Travel	28 (56.0%)	22 (44.0%)
Society and culture	9 (16.4%)	46 (83.6%)
Food	12 (14.0%)	74 (86.0%)
Business	4 (8.9%)	41 (91.1%)
Entertainment	17 (7.6%)	208 (92.4%)

Classification Results Using Contextual Features

In total, with the Alchemy API and SUTime, we identify 107 (7.82%) questions in our data set with explicit spatial mentions, and 209 (15.28%) questions with temporal constraints, for a total of 316 context attributes. Only 25 (1.83%) questions contain both spatial and temporal constraints. In Table 6, we depict the distribution of these questions according to their implied nature. We find that compared with accuracy and knowledge questions, the social category contains significantly more location (13.39%) and temporal-specific (33.93%) inquiries. Examples of such questions include "WHO wants to come with me to see Bassnectar at Madison Square Garden on oct. 4?????? #replytweet," and "Who wants to go to Peru with me next summer and climb Machu Picchu? #seriousquestion http://t.co/I80Dd0emo3."

Tables 7 and 8 report the top five question topics results, including location and temporal mentions. We observe that among all topical categories, the travel-related questions contained the most temporal and spatial constraints, followed by the topics "food," "business," and "entertainment." We also notice that more than 30% of sports-related questions are characterized with temporal mentions, such as, "who do we play Friday? #replytweet" and, "Anyone know who is going to start for the October 10th game? #ColombiavsSalvador #replytweet." Last, 16.4% of "society & culture" questions are targeted at certain location or area. Examples include "Alright, how to get a marriage certificate attested in Dubai? #Help," and "When immigrating to Canada do they transfer your driver's license over or is there a mandatory test first? #askingforafriend #makeitharder."

By replacing the spatial and temporal expressions with the corresponding annotations (<TIME> and <LOCATION>), we rebuilt the classification model using the word level n-gram ($n = 1, 2, 3$) features. Again, we remove words and phrases with frequencies less than five to reduce feature sparsity. Table 9 compares the classification performance of the n-gram features with and without replacement of location and

TABLE 8. Top 5 question topics by percentage of temporal mention.

Question topic	With time	Without time
Travel	16 (32.0%)	34 (68.0%)
Sports	32 (30.2%)	74 (69.8%)
Business	10 (22.2%)	35 (77.8%)
Food	18 (20.9%)	68 (79.1%)
Entertainment	46 (20.4%)	179 (79.6%)

DOI: 10.1002/asi

TABLE 9. Classification performance for SVM, Decision Tree, and Naïve Bayes classifiers using contextual features.

With or without contextual features	Method	Precision	Recall	Accuracy	F1
With temporal and spatial replacement	SVM	0.805	0.803	0.803	0.804
	Decision Tree	0.765	0.765	0.765	0.765
	Naïve Bayes	0.768	0.765	0.765	0.766
Without temporal and spatial replacement	SVM	0.805	0.801	0.801	0.802
	Decision Tree	0.764	0.764	0.764	0.764
	Naïve Bayes	0.769	0.765	0.765	0.766

TABLE 10. Classification performance for SVM, Decision Tree, and Naïve Bayes classifiers using syntactic features.

Method	Precision	Recall	Accuracy	F1
SVM	0.327	0.572	0.572	0.416
Decision Tree	0.558	0.621	0.621	0.577
Naïve Bayes	0.563	0.577	0.577	0.561

TABLE 12. Classification performance for SVM, Decision Tree, and Naïve Bayes classifiers using all question features.

Method	Precision	Recall	Accuracy	F1
SVM	0.836	0.832	0.832	0.833
Decision Tree	0.772	0.770	0.770	0.771
Naïve Bayes	0.767	0.763	0.763	0.765

temporal mentions. We notice that the general contextual features improved the performance relative to just the n-gram features, although not significantly.

Classification Results Using Syntactic Features

Table 10 illustrates the classification results using syntactic features, including number of clauses, words, characters, and whether the tweet contains a picture, as a form of information seeking and sharing (Jansen, 2008). Again, the syntactic-based classifier outperformed the majority-voted baseline, although its predictive power is limited.

Upon further review of question length (as shown in Table 11), we observe that on average, knowledge questions were significantly longer than the accuracy and social ones on all three levels. Through our subsequent investigation on the content of questions, we note that knowledge questions tend to use more words to provide additional contextual information about the questioner's information needs. Examples of such questions include "Any ideas where I can get some keepsake trunks from? Want something special to store memorable bits for each member of the family. #help," "What kind of laptop should I get for college work and possibly some online gaming with B? #replytweet #help #laptop #gaming."

In total, we extract 67 (4.90%) questions from our data set containing pictures. Through our analysis, we find that 4.8% of accuracy, 1.8% of social, and 5.4% of knowledge questions include pictures that are related to the post's content; however, based on our analysis, no statistical difference was noted across the three question categories, χ^2 (2, N = 1,368) = 2.64, $p = 0.27 > 0.05$.

TABLE 11. Question length across ASK question types.

	Num clauses		Num words		Num characters	
Question type	Mean	*SD*	Mean	*SD*	Mean	*SD*
Accuracy	1.321	0.696	11.286	5.244	66.496	28.792
Social	1.318	0.753	11.009	4.276	64.055	25.363
Knowledge	1.669	0.991	14.670	5.565	83.349	30.589

Ensemble Classification Results Using Question Attributes

Until this point, we perform the classification tasks along each individual question dimension, one at a time. In this section, we construct an ensemble classifier by using all of those features simultaneously. We report the ensemble classification results in Table 12, from which we see an improvement over all of the aforementioned models using features from just a single dimension. The ensemble classifier correctly classifies 81.2% of the accuracy questions, 78.6% of the social questions, and 85.1% of the knowledge ones, achieving a weighted average accuracy of 83.2%.

Classification Results Using Answer Features

In addition to the question features, we also study the interaction and social patterns using answer attributes. In total, we find that among all 1,368 questions, 574 questions receive at least one answer. The descriptive statistics for answered questions is shown in Table 13. Consistent with previous studies (Liu & Jansen, 2013; Paul, Hong, & Chi, 2011a), less than half of questions received replies. On

TABLE 13. Descriptive Statistics of answered questions across ASK question types.

		Avg #answers		Avg #unique answerers	
Question type	#Answered (%)	Mean	*SD*	Mean	*SD*
Accuracy	182 (38.32%)	4.39	3.94	1.71	1.43
Social	44 (39.29%)	3.95	2.65	1.57	1.07
Knowledge	348 (44.56%)	5.16	4.38	2.01	1.63

DOI: 10.1002/asi

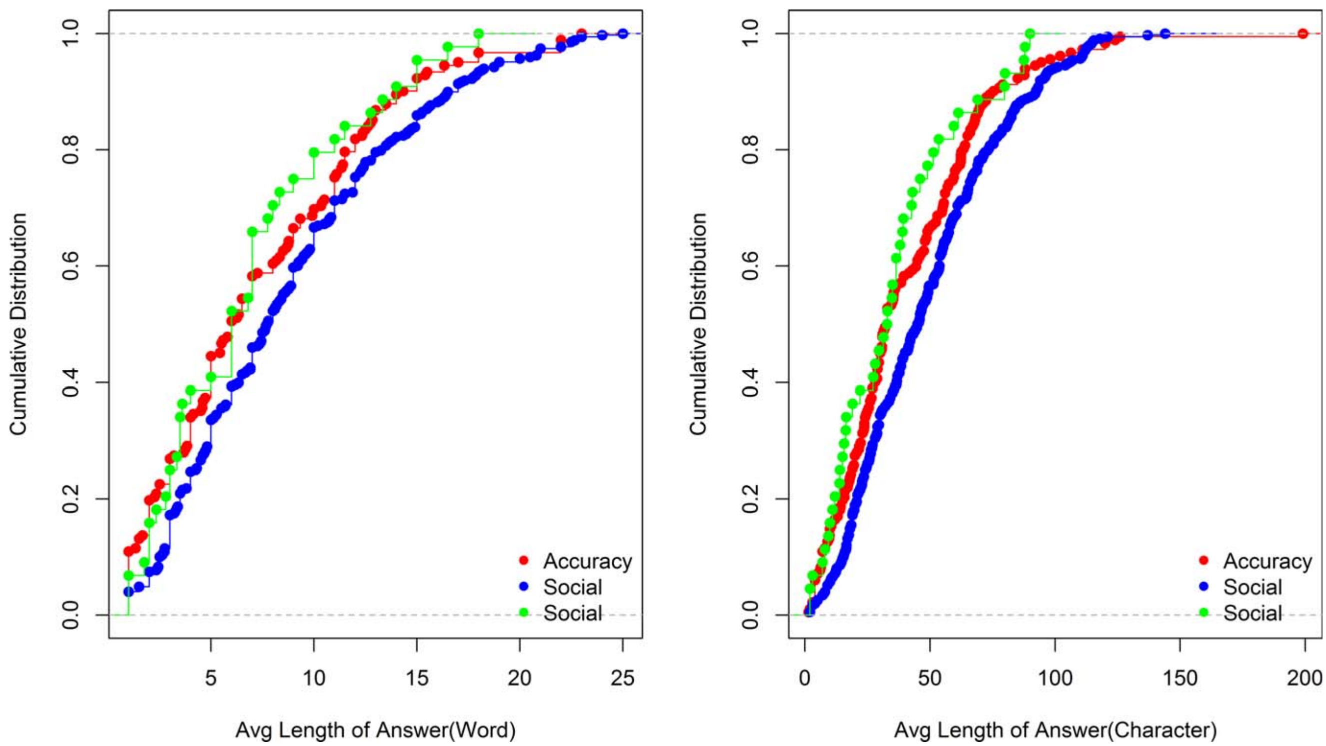

FIG. 2. Distribution of answer length on word and character levels across ASK question types. [Color figure can be viewed at wileyonlinelibrary.com]

average, knowledge questions get the highest number of answers from the greatest number of unique respondents.

As for the analysis on the reciprocal relationship between the questioner and the respondent on Twitter (i.e., following each other, unfollowing each other, and only one following the other), among all 1,171 unique questioner-answerer pairs in our data set, 886 (75.66%) of the follow relations were reciprocal; 108 (9.22%) were one way, and 143 (12.21%) were not following each other. The number of reciprocal following relations in our collection is relatively high, which is consistent with the 79.5% as reported in Zhang's study (2012). We only treat individuals with reciprocal followship as "friends" and the rest as "strangers." Chi-square test examining the dependency between the questioner-respondent friendship and the answered question type reveal a significant trend between the two variables, χ^2 (2, $N = 574$) = 7.74, $p = 0.02 < 0.05$. We find that "strangers," interestingly, had a higher probability (29.6%) of answering knowledge questions than friends, perhaps supporting of the concept of the theory of weak ties (Granovetter, 1983). However, this is unexpected given that previous work (Morris et al., 2010b) showed that people claimed that they preferred to ask subjective questions to their friends for more personalized and trustworthy responses, indicating that the stated preference is not based on actual performance. The observed behavior in our data set does not support this claim. One reason for this could also be that compared to accuracy (20.3%) and social questions (15.9%), knowledge questions, such as "Need new phone what should I get? #help," require less expertise and time investment, so that could be a better option for strangers to offer their help.

In addition to examining the relationship between the type of followship and the answered question type, we perform Kruskal-Wallis tests on the average answer length. The results were significant with answers to the knowledge questions containing the most words ($M = 8.73$, $SD = 5.48$) and characters ($M = 48.58$, $SD = 29.24$), and answers to the social questions containing the least ($M = 6.82$, $SD = 4.53$; $M = 34.96$, $SD = 24.69$) ($p = 0.00 < 0.05$). We plot the cumulative distribution of answer length across question types in Figure 2.

Considering the real time nature of social Q&A, we also look at how quickly the three different types of questions received responses. We adopt two metrics in this study to measure the response speed: (a) the time elapsed until receiving the first answer and (b) the time elapsed until receiving the last answer. In Figure 3, we plot the empirical cumulative distribution of response time in seconds using both measurements. We logarithmically transform the response time given its logarithmic distribution.

In general, about 70% of questions in our data set posted on Twitter received their first answer within an hour, no matter their question types (73.10% of accuracy questions, 63.64% of social questions, and 67.53% of knowledge questions). From Figure 3, we notice that it took slightly longer for individuals to answer knowledge questions than the accuracy and social ones. The Kruskal-Wallis result also reveals a significant difference in the arrival time of the first

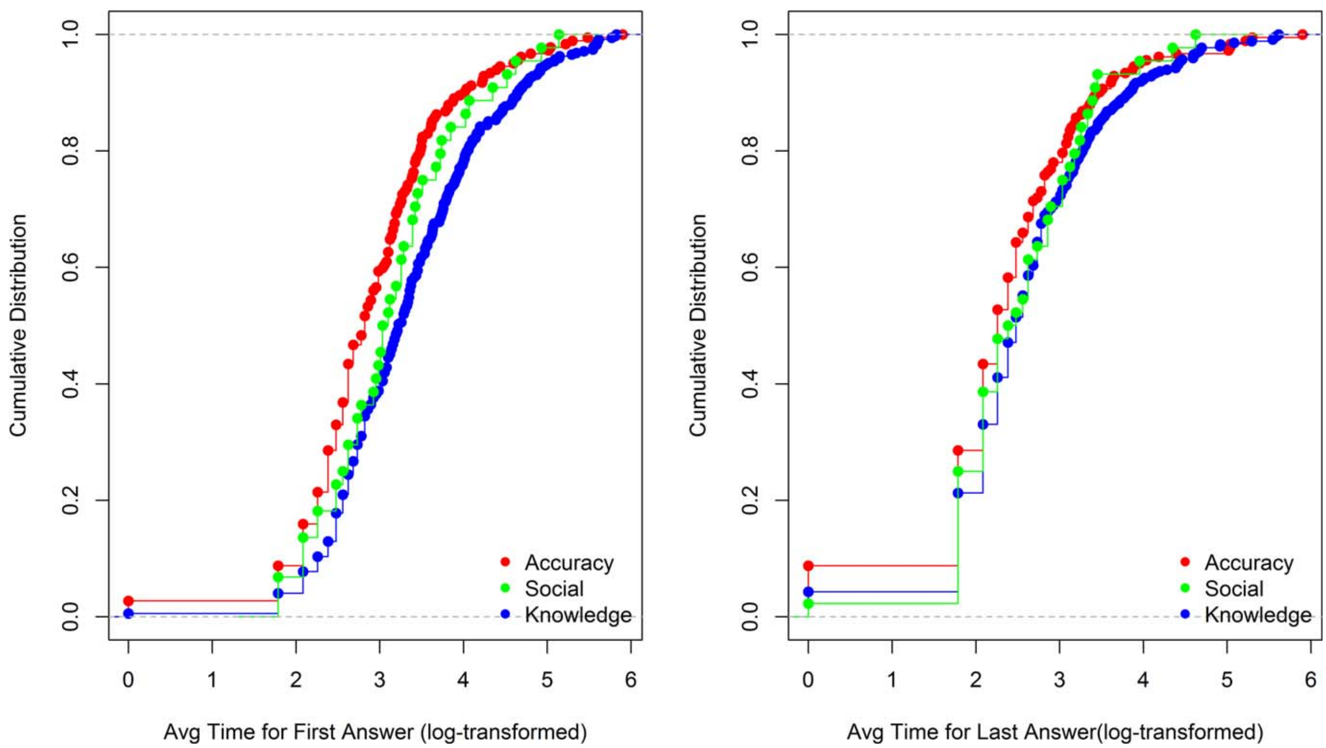

FIG. 3. Distribution of question response time in log (seconds) across ASK question types. [Color figure can be viewed at wileyonlinelibrary.com]

answer across question types ($p = 0.00 < 0.05$). In addition to the first reply, we also adopt the arrival time of the last answer to imply the temporality of each question. As defined by Pal, Margatan, and Konstan (2012, p. 2), question temporality is "a measure of how long the answers provided on a question are expected to be valuable." We chose this measure of temporality, assuming that individuals would not waste their time answering out-of-date questions. Again, the Kruskal-Wallis result demonstrates significant between-group differences on the arrival time of the last answer ($t = 3.76$, $p < 0.05$), with knowledge questions having the longest temporality and the accuracy ones having the least. An example of accuracy questions with short temporal durations is: "When is the game today? #replytweet." At last, concerning the use of URLs in answering others' questions, the Kruskal-Wallis result was also significant ($p = 0.00 < 0.05$) in that almost no social question contained any URL that pointed to external sources.

Table 14 illustrates the classification results using the interaction features from the answer's perspective.

Discussion and Design Implications

From literature on identifying information-seeking questions in social Q&A (Efron & Winget, 2010; Li et al., 2011; Zhao & Mei, 2013), we conduct this research to investigate the nature of the questioner's inquiry in social Q&A. Unlike taxonomies proposed in previous studies, ASK was developed to cope with information-seeking intents that can be answered with different strategies. According to their different subjectivity requirements and personal relevance, we assume that social questions are targeted mainly at one's online friends or followers rather than strangers, as they are more personal oriented; knowledge questions are for survey purposes and require as many subjective responses as possible; whereas computer generated responses could satisfy accuracy-seeking questioners because most of these questions are objective in nature. The ASK taxonomy proves to be reliable with a relatively high interannotator agreement.

In addition, our study also shows the feasibility of automatically classifying questions into accuracy, knowledge, and social types. We assess the effectiveness of our classifier and demonstrate its reliability in distinguishing the ASK types of questions with a classification accuracy of 83.20%. Through more in-depth analyses of all three types of questions, we find that, first, regarding the phrasing of questions, our results imply that contextual restrictions (e.g., time and location) were imposed more often on knowledge and social questions. In addition, our results reveal that knowledge

TABLE 14. Classification performance for SVM, Decision Tree, and Naïve Bayes classifiers using answer features.

Method	Precision	Recall	Accuracy	F1
SVM	0.368	0.606	0.606	0.458
Decision Tree	0.527	0.592	0.592	0.541
NaiveBayes	0.500	0.516	0.516	0.500

DOI: 10.1002/asi

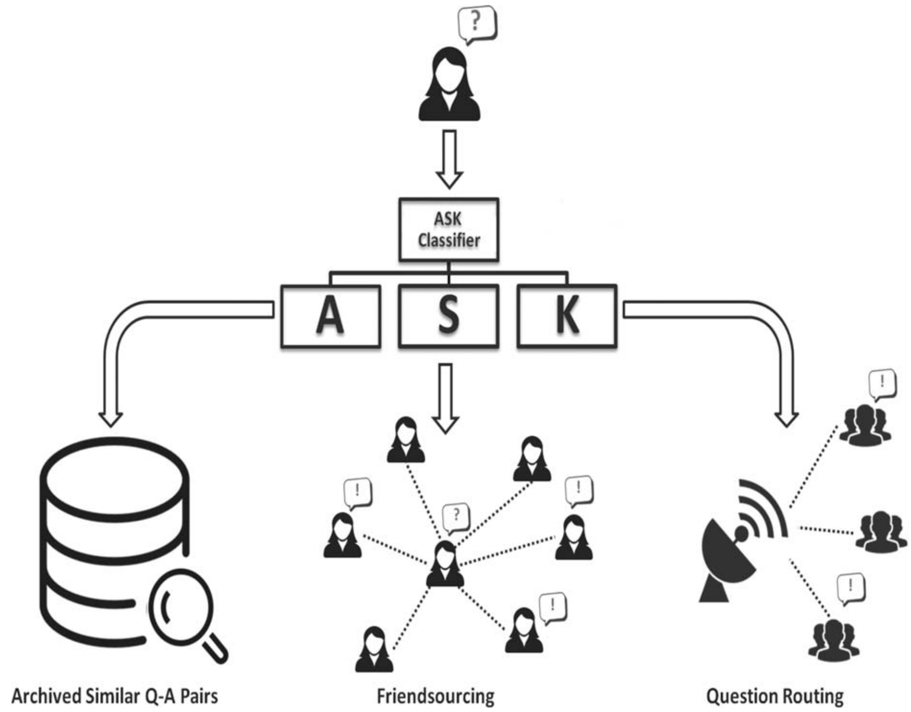

FIG. 4. Hybrid social Q&A system leveraging ASK taxonomy for question routing locating for social Q&A questions.

questions experience longer time-lags in getting their initial answers and also tend to expire in shorter durations. Last, we demonstrate that even though individuals prefer to ask subjective questions to their friends for tailored responses (Morris et al., 2010b), it turns out that, in reality, knowledge questions were being responded to more by strangers than accuracy and social questions.

We think this gap between the ideal and reality imposed a design challenge in maximizing the personalization benefits from strangers in social Q&A. Although routing the knowledge questions to the asker's friends or followers could, to some extent, increase the question response probability, it could also limit the variety of answers received, due to the similar backgrounds and knowledge among SNS friends and followers (Liu & Weber, 2014; Weng, Lim, Jiang, & He, 2010). For example, it would be inappropriate to route questions such as *"I'll be in Cape Town in time for supper, which restaurant should I try out?"* to the asker's followers on Twitter unless they have been to Cape Town before. One potential approach to fulfilling the diverse expectations of the questioners with knowledge information needs would be via routing the questions to a larger number of audiences. This would be accompanied by measuring the similarity between the asker and a number of strangers with relevant experiences, as well as ranking the potential answerers according to their relation to the asker, or their reputation (Chen & Fong, 2010).

Our study on the ASK types of questions has importance for both researchers and practitioners for several reasons. From the theoretical point of view, our taxonomy, ASK, can be adopted as a conceptual framework for future studies on social Q&A. Also, the predictive model enables automatic identification of question types and can be used to facilitate future social Q&A studies on a larger scale. In terms of design implications, the ASK taxonomy allows practitioners to understand the distinct intentions behind all three types of questions, and then to design and develop better social Q&A systems to support question answering on SNS.

We believe that results in this study reveals the inadequacy of using one single answering strategy (i.e., searching through the archived question-answer pairs, question routing) to serve all social Q&A questions, and points to a hybrid approach in future social Q&A systems. In Figure 4, we illustrate a possible design of such a hybrid Q&A system relying on the ASK taxonomy. Under that condition, after questions are asked, the ASK social Q&A classifier automatically categorizes the questions into one of the three ASK types. Considering the factual nature and short duration of those accuracy questions, the system generates replies based on an archive of similar question–answer pairs, without human intervention; additionally, given the nature of knowledge questions and stranger's interests in answering them, one could potentially develop an algorithm to route those subjective questions to appropriate respondents based on their locations and past experiences. Last, considering the "acquaintance-oriented" nature of certain social questions, the system leaves them for one's online friends or followers to answer.

For future research, the most immediate goal is to implement the above mentioned hybrid social Q&A system so that we could conduct user studies to evaluate the

DOI: 10.1002/asi

effectiveness of the ASK taxonomy. In addition, we aim to conduct more detailed analysis of the ASK types of questions to learn more granular attributes of user intent. With the analysis results, we could later expand the ASK taxonomy depth-wise to include more granular categories at lower levels. For instance, for knowledge-seeking questions, we could further divide them into emotion questions, purchasing questions, and so on. Also, to validate the generalizability of our classification model, it would be good to replicate this study using data collected at different intervals or periods of time.

Conclusion

In this research, we present the ASK taxonomy that differentiates social Q&A questions into three types, accuracy, social, and knowledge, according to their different scope of targeted respondents and different requirements for answers. Based on the ASK taxonomy, we develop and implement a predictive model based on features constructed from lexical, topical, contextual, syntactic, and interaction perspectives using machine learning techniques. The automated method proves to be effective in classifying ASK types of questions proposed in social Q&A, with a high classification accuracy of 0.83. We further discuss design implications relying on the ASK taxonomy and associated automated classification method, proposing a hybrid social Q&A system with different routing schemes based on ASK question types. We believe that our ASK taxonomy could be used as a theoretical framework for characterizing user's information needs in social Q&A. Also, with the help of the automated classification method that we implemented, different types of questions can be routed to the most appropriate types of respondents to increase the question response rate.

References

Aikawa, N., Sakai, T., & Yamana, H. (2011). Community QA question classification: Is the Asker looking for subjective answers or not? IPSJ Online Transactions, 4, 160–168.

Batool, R., Khan, W.A., Hussain, M., Maqbool, J., Afzal, M., & Lee, S. (2012). Towards personalized health profiling in social network. Paper presented at the 6th International Conference on New Trends in Information Science and Service Science and Data Mining (ISSDM 2013), Macau.

Bozzon, A., Brambilla, M., & Ceri, S. (2012). Answering search queries with crowdsearcher Proceedings of the 21st International Conference on World Wide Web (pp. 1009–1018). Lyon, France: ACM.

Castillo, C., Mendoza, M., & Poblete, B. (2011). Information credibility on Twitter. Paper presented at the Proceedings of the 20th International Conference on World Wide Web (WWW 2011), Hyderabad, India. New York: ACM.

Chang, A.X., & Manning, C.D. (2012). SUTime: A library for recognizing and normalizing time expressions. Paper presented at the 8th International Conference on Language Resources and Evaluation (LREC 2012), Istanbul, Turkey.

Chen, L., Zhang, D., & Mark, L. (2012). Understanding user intent in community question answering. Paper presented at the 21st International Conference Companion on World Wide Web (WWW 2012), Lyon, France.

Chen, W., & Fong, S. (2010). Social network collaborative filtering framework and online trust factors: A case study on Facebook. Proceedings of the 5th International Conference on Digital Information Management (pp. 266–273). Thunder Bay, ON, Canada: IEEE.

Cortes, C., & Vapnik, V. (1995). Support-vector networks. Machine Learning, 20(3), 273–297.

Efron, M., & Winget, M. (2010). Questions are content: A taxonomy of questions in a microblogging environment. Paper presented at the Annual Conference of American Society for Information Science and Technology (Vol. 47, pp. 1–10). Pittsburgh, PA: John Wiley & Sons.

Ellison, N.B., Gray, R., Vitak, J., Lampe, C., & Fiore, A.T. (2013). Calling all Facebook friends: Exploring requests for help on Facebook. Paper presented at the 7th International AAAI Conference on Web and Social Media ICWSM (ICWSM 2013), Washington, DC.

Evans, B.M., & Chi, E.H. (2008). Towards a model of understanding social search. Paper presented at the The 2008 ACM Conference on Computer Supported Cooperative Work (CSCW 2008), San Diego, CA.

Gazan, R. (2011). Social Q&A. Journal of the American Society for Information Science and Technology, 62(12), 2301–2312.

Granovetter, M. (1983). The strength of weak ties: A network theory revisited. Sociological Theory, 1(1), 201–233.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I.H. (2009). The WEKA data mining software: An update. ACM SIGKDD Explorations Newsletter, 11(1), 10–18.

Harper, F.M., Moy, D., & Konstan, J.A. (2009). Facts or friends?: Distinguishing informational and conversational questions in social Q&A sites. Paper presented at the 26th SIGCHI Conference on Human Factors in Computing Systems (CHI 2009), Florence, Italy.

Harper, F., Weinberg, J., Logie, J., & Konstan, J. (2010). Question types in social Q&A sites. First Monday, 15(7).

Hecht, B., Teevan, J., Morris, M.R., & Liebling, D.J. (2012). SearchBuddies: Bringing search engines into the conversation. Proceedings of the 6th International AAAI Conference on Weblogs and Social Media (Vol. 12, pp. 138–145). Dublin, Ireland: AAAI.

Horowitz, D., & Kamvar, S.D. (2010). The anatomy of a large-scale social search engine. Proceedings of the 19th International Conference on World Wide Web (pp. 431–440). Raleigh, NC: ACM.

Jansen, B.J. (2008). Searching for digital images on the web. Journal of Documentation, 64(1), 81–101.

Jansen, B.J., Booth, D.L., & Spink, A. (2008). Determining the informational, navigational, and transactional intent of web queries. Information Processing & Management, 44(3), 1251–1266.

Jansen, B.J., Zhang, M., Sobel, K., & Chowdury, A. (2009). Twitter power: Tweets as electronic word of mouth. Journal of the American Society for Information Science and Technology, 60(11), 2169–2188.

Jelinek, F. (1990). Self-organized language modeling for speech recognition. Readings in speech recognition, 450–506.

John, G.H., & Langley, P. (1995). Estimating continuous distributions in Bayesian classifiers. Paper presented at the Proceedings of the 11th Conference on Uncertainty in Artificial Intelligence, Montreal, Quebec, Canada.

Kim, S., Oh, J.S., & Oh, S. (2007). Best-answer selection criteria in a social Q&A site from the user-oriented relevance perspective. Paper presented at the American Society for Information Science and Technology (ASIS&T 2007), Milwaukee, WI.

Kwak, H., Lee, C., Park, H., & Moon, S. (2010). What is Twitter, a social network or a news media? Paper presented at the 19th International Conference on World Wide Web (WWW 2010), Raleigh, NC.

Lampe, C., Gray, R., Fiore, A.T., & Ellison, N. (2014). Help is on the way: Patterns of responses to resource requests on Facebook. Paper presented at the 17th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW 2014), Vancouver, BC, Canada.

Li, B., Liu, Y., Ram, A., Garcia, E.V., & Agichtein, E. (2008). Exploring question subjectivity prediction in community QA. Paper presented at the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2008), Singapore.

Li, B., Si, X., Lyu, M.R., King, I., & Chang, E.Y. (2011). Question identification on Twitter. Paper presented at the 20th ACM International Conference on Information and Knowledge Management (CIKM 2011), Glasgow, United Kingdom.

Liu, Y., Alexandrova, T., & Nakajima, T. (2013). Using stranger as sensors: Temporal and geo-sensitive question answering via social media. Paper presented at the 22nd international conference on World Wide Web (WWW 2013), Rio de Janeiro, Brazil.

Liu, Y., Bian, J., & Agichtein, E. (2008). Predicting information seeker satisfaction in community question answering. Paper presented at the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2008), Singapore.

Liu, Z., & Jansen, B. (2012). Almighty Twitter, what are people asking for? Paper presented at the Annual Conference of American Society for Information Science and Technology (Vol. 49, pp. 1–10). Baltimore, MD: John Wiley & Sons.

Liu, Z., & Jansen, B.J (2013). Factors influencing the response rate in social question and answering behavior. Paper presented at the 16th ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW 2013), San Antonio, TX.

Liu, Z., & Weber, I. (2014). Is twitter a public sphere for online conflicts? A cross-ideological and cross-hierarchical look. Paper presented at the 6th International Conference on Social Informatics (pp. 336–347). Barcelona, Spain: Springer International.

Morris, M.R., Teevan, J., & Panovich, K. (2010a). A comparison of information seeking using search engines and social networks. Paper presented at the 4th Inernational AAAI Conference on Weblogs and Social Media (ICWSM 2010), Washington, DC.

Morris, M.R., Teevan, J., & Panovich, K. (2010b). What do people ask their social networks, and why? A survey study of status message Q&A behavior. Paper presented at the 28th ACM Conference on Human Factors in Computing Systems (SIGCHI 2010), Atlanta, Georgia.

Pal, A., Margatan, J., & Konstan, J. (2012). Question temporality: Identification and uses. Paper presented at the ACM 2012 Conference on Computer Supported Cooperative Work (pp. 257–260). Seattle, WA: ACM.

Paul, S., Hong, L., & Chi, E. (2011a). Is Twitter a good place for asking questions? A characterization study. Paper presented at the 5th International AAAI Conference on Weblogs and Social Media (pp. 578–581). Barcelona, Spain: AAAI.

Paul, S.A., Hong, L., & Chi, E.H. (2011b). What is a question? Crowdsourcing tweet categorization. Paper presented at the 29th ACM Conference on Human Factors in Computing Systems (SIGCHI 2011), Vancouver, BC, Canada.

Platt, J. (1999). Fast training of support vector machines using sequential minimal optimization. In Advances in Kernel Methods (1st ed., pp. 185–208). Cambridge, MA: MIT Press.

Porter, M.F. (1980). An algorithm for suffix stripping. Program: Electronic Library and Information Systems, 14(3), 130–137.

Quercia, D., Capra, L., & Crowcroft, J. (2012). The social world of Twitter: Topics, geography, and emotions. Paper presented at the 6th Inernational AAAI Conference on Weblogs and Social Media (ICWSM 2012), Dublin, Ireland.

Quinlan, J. (1987). Simplifying decision trees. International Journal of Man-machine Studies, 27(3), 221–234. Retrieved from http://dx.doi.org/10.1016/s0020-7373(87)80053-6.

Quinlan, J.R. (2014). C4. 5: Programs for machine learning. Elsevier.

Rzeszotarski, J.M., Spiro, E.S., Matias, J.N., Monroy-Hernández, A., & Morris, M.R. (2014). Is anyone out there? Unpacking Q&A hashtags on Twitter. Paper presented at the 32nd Annual ACM Conference on Human Factors in Computing Systems (SIGCHI 2014), Toronto, ON, Canada.

Sadilek, A., Kautz, H., & Silenzio, V. (2012). Modeling spread of disease from social interactions. Paper presented at the 6th International AAAI Conference on Weblogs and Social Media (pp. 322–329). Dublin, Ireland: AAAI.

Schmid, H. (1994). Probabilistic part-of-speech tagging using decision trees. In International Conference on New Methods in Language Processing (pp. 44–40). Manchester, UK: ACL.

Shah, C. (2011). Measuring effectiveness and user satisfaction in Yahoo! Answers. First Monday, 16(2).

Sil, A., & Cucerzan, S. (2002). Temporal scoping of relational facts based on Wikipedia data. Journal of Natural Language Engineering, 8(4), 4.

Strötgen, J., & Gertz, M. (2010). HeidelTime: High quality rule-based extraction and normalization of temporal expressions. Paper presented at the 5th International Workshop on Semantic Evaluation (SemEval 2010), Uppsala, Sweden.

Teevan, J., Collins-Thompson, K., White, R.W., Dumais, S.T., & Kim, Y. (2013). Slow search: Information retrieval without time constraints. Proceedings of the Symposium on Human-Computer Interaction and Information Retrieval (p. 1). Vancouver, BC, Canada: ACM.

Toutanova, K., Klein, D., Manning, C.D., & Singer, Y. (2003). Feature-rich Part-of-speech tagging with a cyclic dependency network. Paper presented at the Conference of the North American Chapter of the Association for Computational Linguistics on Human Language Technology (HLT-NAACL 2003), Edmonton, Canada.

Twitter. (2011). "#Numbers." Retrieved from http://blog.twitter.com/2011/03/numbers.html

Vossen, D. (2012). Detecting speaker status in a social setting with a single triaxial accelerometer (Doctoral dissertation). University of Amsterdam.

Wang, L., Chen, L., Hou, C., & Chen, G. (2014). What people inquire about locations?: A study on the taxonomy of location-based questions in campus. Paper presented at the CHI'14 Extended Abstracts on Human Factors in Computing Systems, Toronto, Canada.

Weng, J., Lim, E.-P., Jiang, J., & He, Q. (2010). Twitterrank: Finding topic-sensitive influential twitterers. Paper presented at the Proceedings of the 3rd ACM International Conference on Web Search and Data Mining, New York, NY.

Yan, T., Kumar, V., & Ganesan, D. (2010). Crowdsearch: Exploiting crowds for accurate real-time image search on mobile phones. Proceedings of the 8th International Conference on Mobile Systems, Applications, and Services (pp. 77–90). San Francisco, CA: ACM.

Yang, J., Morris, M.R., Teevan, J., Adamic, L.A., & Ackerman, M.S. (2011). Culture matters: A survey study of social Q&A behavior. Paper presented at the 5th International AAAI Conference on Weblogs and Social Media (ICWSM 2011), Barcelona, Spain.

Yang, Y., & Pedersen, J.O. (1997). A comparative study on feature selection in text categorization. Paper presented at the the 29th International Conference on Machine Learning (ICML 2012) Edinburgh, Scotland.

Zhang, P. (2012). Information seeking through microblog questions: The impact of social capital and relationships. Paper presented at the Annual Conference of the American Society for Information Science and Technology (Vol. 49, pp. 1–9). Baltimore, MD: John Wiley & Sons.

Zhao, Z., & Mei, Q. (2013). Questions about questions: An empirical analysis of information needs on Twitter. Paper presented at the 22nd International Conference on World Wide Web (WWW 2013), Rio de Janeiro, Brazil.

Zhou, T.C., Si, X., Chang, E.Y., King, I., & Lyu, M.R. (2012). A data-driven approach to question subjectivity identification in community question answering 26th AAAI Conference on Artificial Intelligence (AAAI 2012). Toronto, Ontario, Canada.

DOI: 10.1002/asi