
A Taxonomy for Classifying Questions Asked in Social Question and Answering

Zhe Liu

College of Information Sciences and Technology
The Pennsylvania State University
zul112@ist.psu.edu

Bernard J. Jansen

Principal Scientist
Qatar Computing Research Institute
Doha, Qatar
jjansen@acm.org

Abstract

The rapid advancement of Web2.0 technologies has made social networking sites, such as Facebook and twitter, important venues for individuals to seek and share information. As understanding the information needs of users is crucial for designing and developing tools to support their social Q&A behaviors, in this

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). Copyright is held by the author/owner(s).

CHI'15 Extended Abstracts, April 18–23, 2015, Seoul, Republic of Korea.

ACM 978-1-4503-3146-3/15/04.

<http://dx.doi.org/10.1145/2702613.2732928>

paper, we present a new way of classifying questions from a design perspective, with the aim of facilitating the development of question routing systems according to individual's information need. As an attempt to understand the questioner's intent in social question and answering environments, we propose a taxonomy of questions posted on Twitter, called ASK. Our taxonomy uncovers three different kinds of questions: accuracy, social, and knowledge. In addition, to enable automatic detection on these three types of information needs, we measured and reported on the differences in ASK types of questions reflected at both lexical and syntactic levels.

Author Keywords

Social question and answering; social Q&A; social search; social networks; information seeking; Weibo

ACM Classification Keywords

H.3.5 [Information Storage and Retrieval]: Online Information Services; J.4 [Social and Behavioral Sciences]: Sociology

Introduction

The dramatic rise in the use of social networking sites (SNS) in the recent years has made the platform a good place for information broadcasting [1] and seeking [2, 3]. Though many of the past studies

Question Type	Sample Questions
Accuracy	When is the debate on UK time? #replytweet Hey gamers. Anyone know how to turn off motion controls for HyruleWarriors?#help Anyone know how to say "fun smasher" in Spanish? #help
Social	Somebody who has an ELLO invitation for me? #Question Who on lemoyn campus has a 5 charger? #replytweet Who's going to the game tomorrow? #replytweet
Knowledge	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 I'm in desperate need of a good book series to read. What would you guys recommend? #help
Conversational	I don't really talk to you guys much and it makes me feel bad and antisocial that I don't. How are you babes today? #ReplyTweet #DMMe How is it possible to have this much workload during the first month of school? #overloadincredits #help What is the 1 thing you desire the most? #ReplyTweet

Table 1. Annotation criteria and examples of ASK question types.

analyzed the topics and types of questions asked on social platforms [3-5], few of them has directly addressed the value of question classification in building social question answering services. In this paper, we present a new way of classifying questions from a design perspective, with the aim of facilitating the development of question routing systems according to individual's information need.

While analyzing the intentions of questioners in social question and answering (social Q&A), we propose a taxonomy named ASK, which differentiate questions into three types, including: accuracy questions, in which people ask for fact or common sense; social questions, in which people ask for coordination or companion; and knowledge questions, in which people seek for personal opinions or advices. To validate the ASK taxonomy, we asked human annotators to manually work on 3,000 questions randomly sampled from a dataset containing over 25,000 questions from Twitter. In addition, to enable automatic detection on these three types of information needs, we measured and reported on the differences in ASK types of questions reflected at both lexical and syntactic levels.

We argue that our proposed taxonomy ASK can serve as the first step in designing and developing social Q&A tools as: first, by automatically differentiating questions according to their subjectivity, ASK can be adopted to decide whether the collected answers will be ranked based on authority or summarized for quick digest. Second, assuming that askers of social questions expect "friend-replies", whereas accuracy questions can be answered by even strangers, ASK can then route questions to appropriate respondents according to their relationship to the questioner

Related Work

Morris et al. [3] manually labeled a set of questions posted on social networking platforms and identified 8 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. [4] in their work noticed more rhetorical (42%) questions on Twitter, followed by the categories of factual knowledge (16%), and polls (15%). Adapting the categorization scheme proposed in Morris's work [3], Ellison et al. [5] labeled a set of 20,000 status updates on Facebook, and presented multiple types of mobilization requests beyond information-seeking attempts.

Among all the previous works on question typology in social context, Harper [6] proposed the most comprehensive classification of questions. By coding questions drawn from three popular community Q&A sites, Harper developed a typology of question types fall into three "species" of rhetoric, including: deliberative, epideictic, and forensic. Furthermore, the authors also proposed some potential alternative taxonomies for future research, such as: objective and subjective, past, present and future, advice, opinion and factual, and conversational and informational. Gazan in his work [7] also proposed a need for more appropriate categorization of questions with the consideration of maximizing the likelihood of receiving answers.

Research Objectives

To address the gaps as mentioned in the literature review section, we propose three overarching research objectives:

Method	Prec	Rec	Accu	F1
SMO	0.82	0.82	0.82	0.82
Decision Tree	0.75	0.75	0.75	0.75
Naïve Bayes	0.77	0.77	0.76	0.77

Table 2. Performance of the lexical classifier

1. Propose a taxonomy of question types asked on Twitter with the consideration of maximizing the response probability.
2. Understand the differences of ASK types of questions on both lexical and syntactic levels.
3. Automatically classify into ASK types.

Data Collection

To answer our proposed research questions, we collected all English written tweets posted during September 20th to October 1st, 2014 that containing at least one question mark and any of the questions words, including "who", "where", "when", "what", "why", "how", and "can" using Twitter API. Given the low percentage of information-seeking questions on Twitter [4] and the scope of this study on informational questions only, to filter out as many conversational questions as possible, we further constrained our search query by including a general set of question-signaling hashtags as introduced in Rzeszotarski et al.'s work [8]. Only questions containing at least one of those hashtags has been included in our dataset. We also removed retweets and questions that were 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.

Taxonomy Creation

To answer our first research question, we develop a taxonomy named ASK, which differentiated questions in

a way that can benefit for further development of social Q&A tools, such as question routing, answer ranking, and summarization. To be more specific, we created the taxonomy considering both the subjectivity of the question, and the scope of potential respondents, and categorized them into: *accuracy*, *social*, and *knowledge*. By differentiating questions into those three types, ASK can be used to decide whether the collected answers will be ranked based on authority or summarized for quick digest. Also, ASK can route questions to appropriate respondents according to their relationship to the questioner.

To test the proposed taxonomy ASK, we randomly sampled 3,000 questions from our collected dataset, and recruited two human annotators to work on the labeling task based on our annotation criteria on accuracy, social, and knowledge questions. In order to guide the annotation process and to promote continuity between human annotators, we in this section present the annotation criteria adopted.

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

Word	Accuracy (%)	Social (%)	Knowledge (%)
who	7.98	90.90	6.26
I	12.18	10.00	48.34
what	45.59	2.73	49.49
should	0.63	0.00	13.68
good	0.00	6.36	11.64
who is going	0.21	15.45	0.00
where	7.77	1.82	22.25
what time	6.72	0.00	0.00
me	4.20	29.09	5.63
want	0.63	18.19	3.20

Table 3. Top 10 lexical features of ASK question types

- *Knowledge Question*: The intent of a knowledge question is to receive responses reflecting the answerer's personal opinions, advices, preferences, or experiences. It is usually with a "survey" purpose, which encourages the audience to provide their personal answers.

Accuracy, social, and knowledge questions are all of information-seeking purpose. However, considering that not all questions on Twitter are informational, in our annotation criteria we also adopted the typology of informational and conversational questions from [9], although differentiating these two types are not of our interest in this study.

- *Conversational Tweet*: The conversational tweet is usually with social or conversational purpose, and does not convey any real information need.

To better illustrate our taxonomy proposed in this study, in Table 1 we listed a number of sample questions with accuracy, social, knowledge, or conversational intents.

With the above annotation criteria, the two human annotators worked on the labeling separately. Finally, 2,621 out of 3,000 questions (87.37%) received agreement on their types from the two coders, indicating the relatively high reliability and generalizability of our proposed taxonomy. Among the 2,621 question tweets, 1,253 (47.81%) were labeled as with conversational intent, 475 (18.12%) as accuracy-seeking, 112 (4.27%) as social-seeking, and the rest 781 (29.80%) as knowledge-seeking.

Question Classification

Feature Engineering

Lexical Features - We assumed that given the different information needs behind, there should be a different usage of lexical terms among questions of distinct types. So, we included lexical features which operate at word level, including n-gram ($n = 1, 2, 3$) and POS-tagging patterns of each question. We first adopted word-level n-gram features by counting the frequencies of all unigram, bigram, and trigram tokens that appeared in the training data. Before feature extraction, we lowercased and stemmed all the tokens using the Porter stemmer. We discarded rare terms with observed frequencies of less than 5 to reduce the sparsity of the data. This leaves us with 996 n-gram features. In addition to the lexical features, we believed that POS-tagging may also help in distinguishing the three 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 adopted the Stanford tagger. Again, we counted the frequencies of all unigram, bigram and trigram POS that appeared in the training data. In total, we extract 664 POS-tagging features.

Syntactic Features - The syntactic features measured the writing style of the question at the sentence level. The context features that we adopted in this study include: the length of tweets in sentences / clauses, words, and characters, and whether or not the tweet contains a picture. In order to identify tweets containing pictures, we expand all shortened URLs through a website called LongURL (<http://longurl.org/>).

Classification Methods

With the above features, we next built a multi-class classifier to automatically label questions into three types: accuracy, social, and knowledge. We trained and

tested our model using a number of classification algorithms implemented in Weka, including: Naïve Bayes, and SVM (SMO) and decision tree (J48), using 10-fold cross-validation. We also adopted the majority induction algorithm, which simply predicts the majority class in the dataset, as a baseline model to interpret our classification results. On our data set, with this approach we got a baseline accuracy of 0.571, as 781 tweets were tagged as knowledge-seeking among 1,368 informational questions.

Classification Results

Lexical Features - Due to the large number of lexical features extracted, we evaluated the classification accuracies along the number of features selected. We conducted feature selection using the algorithm of information gain implemented in Weka.

While trying different number of lexical features, we found that either too few or too many features would result in a decrease of prediction accuracy. In addition, SMO outperformed the other two methods in the question classification process using lexical features, with 0.820 accuracy (Accu), 0.823 precision (Prec), 0.820 recall (Rec) and 0.821 F1-measurement (F1). We also noted that the accuracy of 0.820 was much better than the majority class baseline of 0.571, which proofed the possibility of automatically detecting question types using lexical features only. Table 2 demonstrated the classification results of all three methods.

We adopted 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 the table, we noticed that about half of the accuracy and knowledge tweets containing

the question word “what”, whereas 90% of the social 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 found that, while comparing with the other two types of questions, knowledge-seeking tweets asked more about locations. Besides, these questions tend to include more contextual information, with the word “I” used in 48.34% of 782 cases. Typical knowledge questions are like: “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”

Syntactic Features - Table 4 illustrates the classification results using syntactic features, including: number of clauses, words, characters and whether or not the tweet contains a picture. Again, the syntactic-based classifier outperformed the majority-voted baseline, although its predictive power is limited.

We further look at the features of question length as shown in Table 5, and notice that knowledge questions on average were significantly longer than the accuracy and social ones on all three levels. Through our further investigation on the content of questions, we noted that knowledge questions tended 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”

Method	Prec	Rec	Accu	F1
SMO	0.33	0.57	0.57	0.42
Decision Tree	0.56	0.63	0.63	0.58
Naïve Bayes	0.56	0.58	0.58	0.56

Table 4. Performance of the syntactic classifier

Discussion and Implication

We conducted this study to investigate the intent of questions asked on Twitter. We proposed a taxonomy called ASK, which differentiated questions into three types: accuracy, social, and knowledge, with the aim to benefit for further development of social Q&A tools, such as question routing, answer ranking, and summarization. Based on ASK, we also build a predictive model based on features constructed from lexical and syntactic perspectives using machine learning techniques. Using our taxonomy, the inter-annotator agreement in this study was 86.27%.

We assessed the effectiveness of our classifier and proved its reliability in distinguishing the three types of questions, with a classification accuracy of 0.82. In terms of design implications, we believe that our work contributes to the social Q&A field in two ways: First, our proposed taxonomy ASK can be adopted as a theoretical groundwork for future studies on social Q&A. Second, our analysis results allow the practitioners to understand the distinct intentions behind all three types of 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 knowledge 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 accuracy 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.

Num Clauses		
	Mean	StanDev
Accuracy	1.321	0.696
Social	1.318	0.753
Knowledge	1.669	0.991
Num Words		
	Mean	StanDev
Accuracy	11.286	5.244
Social	11.009	4.276
Knowledge	14.670	5.565
Num Characters		
	Mean	StanDev
Accuracy	66.496	28.792
Social	64.055	25.363
Knowledge	83.349	30.589

Table 5. Question length across types.

Reference

- [1] Jansen, B.J., et al., Twitter Power: Tweets as Electronic Word of Mouth. *Journal of the American Society for Information Science and Technology*, 2009. 60(11): p. 2169-2188.
- [2] Liu, Z. and B.J. Jansen, *Almighty Twitter, What are People Asking for?* Proceedings of the American Society for Information Science and Technology, 2012. 49(1): p. 1-10.
- [3] Morris, M.R., J. Teevan, and K. Panovich. *What Do People Ask Their Social Networks, and Why?: A Survey Study of Status Message Q&A Behavior*. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2010. ACM.
- [4] Paul, S.A., L. Hong, and E. Chi, *What Is a Question? Crowdsourcing Tweet Categorization*. Proceedings of the SIGCHI conference on Human Factors in Computing Systems. 2011. ACM.
- [5] Ellison, N.B., et al. *Calling All Facebook Friends: Exploring Requests for Help on Facebook*. in ICWSM. 2013.
- [6] Harper, F.M., et al., Question Types in Social Q&A Sites. *First Monday*, 2010. 15(7).
- [7] Gazan, R., Social Q&A. *Journal of the American Society for Information Science and Technology*, 2011. 62(12): p. 2301-2312.
- [8] Rzeszotarski, J.M., et al. *Is Anyone Out There?: Unpacking Q&A Hashtags on Twitter*. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2014. ACM.
- [9] Harper, F.M., D. Moy, and J.A. Konstan. *Facts or Friends?: Distinguishing Informational and Conversational Questions in Social Q&A Sites*. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2009. ACM