Soc. Netw. Anal. Min. (2016)6:5
DOI 10.1007/s13278-015-0313-x

ORIGINAL ARTICLE

Detecting experts on Quora: by their activity, quality of answers, linguistic characteristics and temporal behaviors

Sumanth Patil[1] · Kyumin Lee[1]

Received: 19 June 2015 / Revised: 2 December 2015 / Accepted: 10 December 2015

Abstract Quora is a fast growing social Q&A site where users create and answer questions, and identify the best answers by upvotes and downvotes with crowd wisdom. Unfortunately, little is known about properties of experts and non-experts and how to detect experts in general topics or a specific topic. To fill the gaps, in this manuscript we (1) analyze behaviors of experts and non-experts in five popular topics; (2) propose user activity features, quality of answer features, linguistic features and temporal features to identify distinguishing patterns between experts and non-experts; and (3) develop statistical models based on the features to automatically detect experts. Our experimental results show that our classifiers effectively identify experts in general topics and a specific topic, achieving up to 97 % accuracy and 0.987 AUC.

1 Introduction

Social Q&A sites are becoming more and more popular because people can post questions, get answers, and befriend with experts. Various platforms have emerged—from general-purpose social Q&A platforms such as Quora and Yahoo Answers to specialized platforms such as Stack Overflow (for programming) and Super User (for computer). Answers are evaluated by upvotes and downvotes from the crowd. These interactions naturally reveal the best answer for a question. Sometimes, new questions in these Q&A sites stimulate answerers to disseminate curated knowledge which may not be available in other websites or it may take time for a user to find, understand and summarize relevant information from other sites. For example, some raw information may be spread across several websites. It would take time for a user to search and understand these pages. Or the information may not be available on the Web. In this case, people may visit a social Q&A site and post a question, expecting experts would give them answers.

As social Q&A sites have become popular with the number of users, people have desire to quickly identify experts in general topics or a specific topic. New users may not be familiar with the community, but they want to find experts who could give them relevant answers. Also, expert identification can be used for an expert recommendation service in a social Q&A site. Unfortunately, little is known about properties of experts and non-experts, and how to detect experts in general topics or a specific topic. Hence, in this manuscript we choose Quora—a fast growing social Q&A site and the 200th most popular site (Alexa 2015). Quora is different from the other Q&A services because it includes social features (e.g., following a user or a topic), which require specific processing. In this manuscript, we answer the following questions: Do experts and non-experts behave differently? Do they change their behaviors over time? Do answers of experts and non-experts contain their linguistic characteristics? Can we measure quality of answers? Based on this analysis and the corresponding observations, can we automatically detect experts in general topics and a specific topic? Will adding temporal (dynamic) properties of experts and non-experts improve the success rate of expert detection?

To answer these questions, we make the following contributions in this manuscript:

✉ Kyumin Lee
kyumin.lee@usu.edu

Sumanth Patil
sumanth.reddy@aggiemail.usu.edu

1 Department of Computer Science, Utah State University, Logan, UT 84322, USA

- First, we collect user profiles from popular topics on Quora, and analyze the properties of experts and non-experts.
- Second, we extract and analyze user activity features, quality of answer features, linguistic features, and temporal features.
- Third, we develop statistical models based on the proposed features to detect experts in general topics and a specific topic. We evaluate what types of classifiers produce the best result.
- Finally, we study whether adding additional features extracted from a user's external accounts such as social media accounts would improve the performance of expert detection.

2 Dataset

To analyze behaviors of experts and non-experts on Quora, the first step is to collect user information. Since there were no publicly available Quora dataset and no official APIs, we developed our own crawler which collected user information on Quora. Our crawling strategy is to first collect five popular topic pages each of which consists of a list of user profile URLs. Users on Quora can follow any topic that they like. Following a topic indicates that they are interested in the topic. We chose five topics such as Mathematics, Business, Politics, Sports and Technology. From each topic page, we randomly selected users, and the crawler collected these users' profiles consisting of information related to user activities and answers posted by the user. Figure 1 shows an example of a Quora user profile which consists of user activity related information such as the number of followees, the number of followers and the number of answers, and a list of answers that the user have posted. By running our crawler, we collected 3720 profiles of users who were interested in one of the five topics in October 2013.

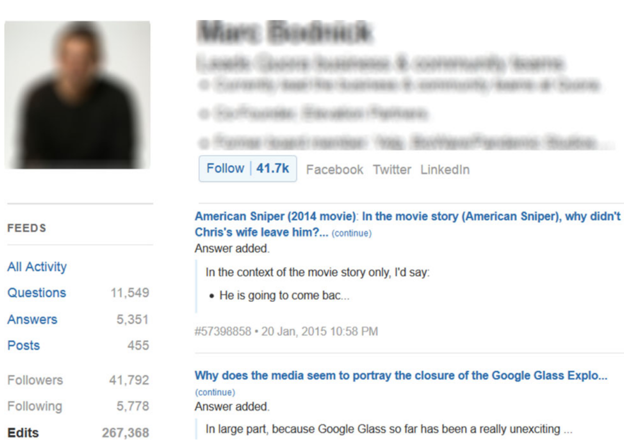


Fig. 1 An example of Quora user profile

Table 1 Dataset consisting of expert and non-expert profiles collected from five topics

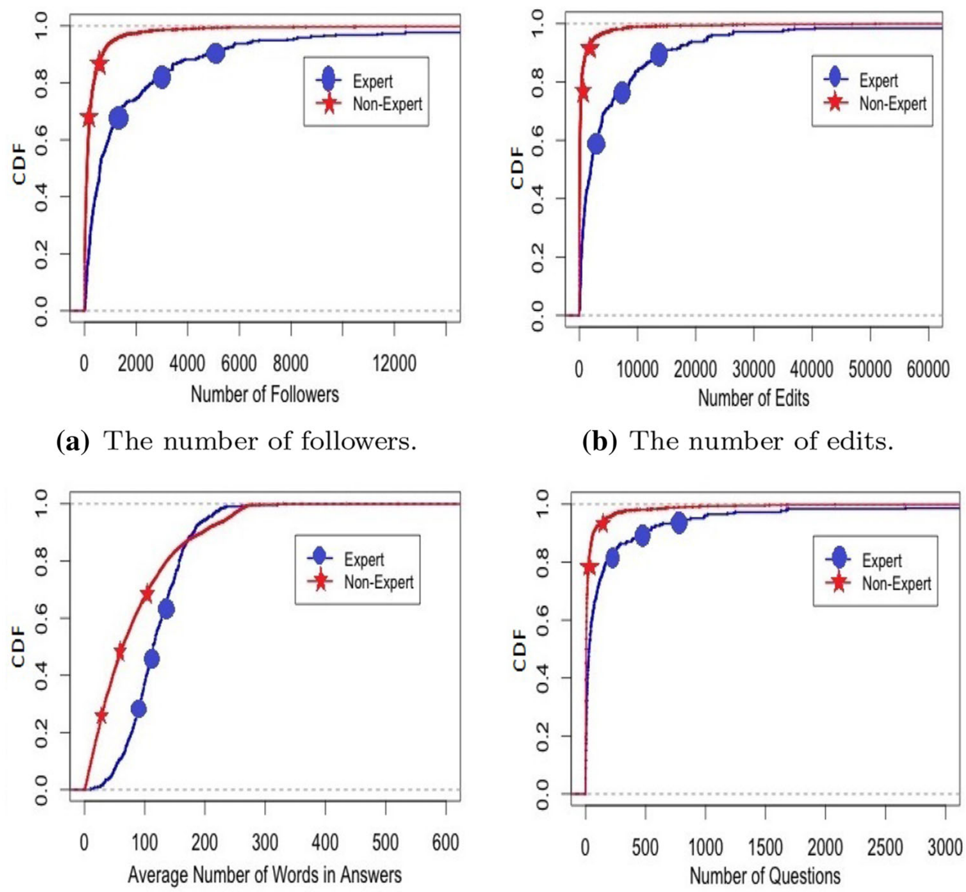
Topic	|Experts|	|Non-experts|	|Users|
Business	74	698	772
Mathematics	94	683	777
Politics	114	742	856
Sports	82	533	615
Technology	68	632	700
Combined dataset	432	3288	3720

Next step is to label the dataset to get the ground truth (i.e., which profile is an expert's profile or a non-expert's profile). In general, an expert is a person who has a comprehensive and authoritative knowledge of or skill in a particular area. Two annotators reviewed each user's profile including answers, activities, the user's profession and so on. Then the annotators labeled each user as an expert or a non-expert to get the ground truth. If they have a conflict result for a user, a third annotator reviewed the user's profile and labeled it. Table 1 shows the statistics of our dataset. Each topic-based data contained between 615 and 856 user profiles. Overall, there were 432 experts and 3288 non-experts in the combined dataset. Note that Quora staff manually select only a few hundred "Top Writers" each year based on recent contributions, overall contributions, and topic expertise (Quora 2014). To get an objective ground truth, we also compared our ground truth with a list of the Top Writers. All the top writers (15 % of the experts) were labeled as experts in our dataset. Even though the remaining 85 % experts in the dataset were not recognized as top writers by Quora, they actively contributed to the community and had expertise in at least one topic.

3 Analyzing behaviors of experts and non-experts: by activity and linguistic characteristics

In the previous section, we presented the collected dataset consisting of expert profiles and non-expert profiles. In this section, we analyze behaviors of experts and non-experts on Quora. First we compare four activities of experts and non-experts.

How many followers do experts and non-experts have? Quora provides followee and follower features like Twitter does. A user can control the number of followees, but cannot control the number of followers. An interesting research question is "will experts have a larger number of followers than non-experts?" Figure 2a presents a cumulative distribution function (CDF) of the number of followers between experts and non-experts. The number of followers of experts is greater than the number of followers



(a) The number of followers. **(b)** The number of edits.

(c) The average number of words in answers. **(d)** The number of questions.

Fig. 2 Cumulative Distribution Functions of the number of followers, edits, words in answers and questions between experts (*blue line with circles*) and non-experts (*red line with stars*)

of non-experts. Since we are analyzing users on a social Q&A site, we conjecture that users tend to follow experts who have posted high quality answers. This is an interesting phenomenon compared with following celebrities on social media sites like Twitter and Facebook. Specifically, users in twitter and Facebook tend to follow celebrities or well-known people to get their live updates including pictures and video, while users in Quora tend to follow experts to get knowledge. These experts may not be well-known people.

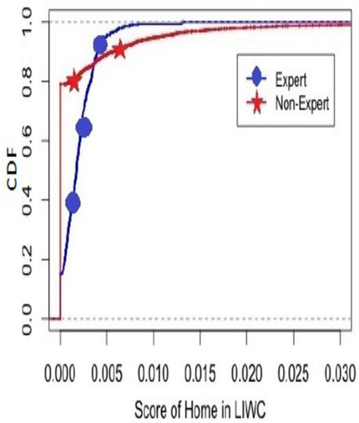
How many edits have experts and non-experts made? A user profile contains the number of edits which means how many times a user edited postings (e.g., editing answers, editing questions). This number would indicate how active a user is on Quora. Figure 2b shows that experts have made a larger number of edits than non-experts, indicating that experts are users who were more active than non-experts. This is an interesting observation. Naturally following questions are “Have experts posted longer answers than non-experts?” and “how many questions have experts and non-experts posted?”.

Have experts posted longer answers than non-experts? To answer this question, we counted the average number of words in answers created by experts and non-experts. Figure 2c shows CDFs of the average number of words. Until reaching to 0.9 in *y*-axis (i.e., 90 % of experts and non-experts), experts have posted longer answers than non-experts. But, some non-experts (the above 0.9 in *y*-axis value) have posted longer answers. more practices to make their answers better in terms of clarity and quality.

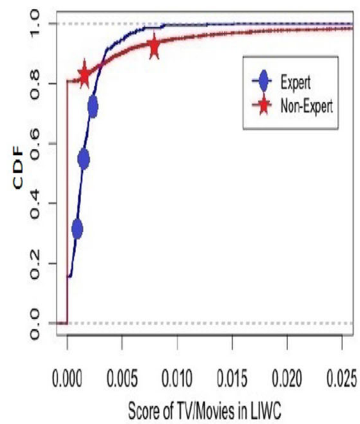
How many questions have experts and non-experts posted? Figure 2d shows CDFs of the number questions that experts and non-experts have posted. People may think experts would be only interested in answering questions. But, surprisingly experts have posted a larger number of questions than non-experts. We conjecture that some experts may be knowledgable in a specific topic, but may be not knowledgable in other topics, so they may post many questions related to other topics.

So far, we have analyzed four activities of experts and non-experts. Next, we study linguistic characteristics of answers posted by experts and non-experts.

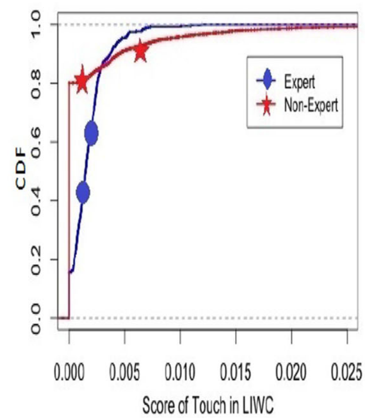
Linguistic Characteristics Do experts create answers with different language use? To answer this question, we used the Linguistic Inquiry and Word Count (LIWC) dictionary, which is a standard approach for mapping text to



(a) Home in LIWC.



(b) TV/Movies in LIWC.



(c) Touch (Feeling) in LIWC.

Fig. 3 Three linguistic characteristics of experts (*blue line with circles*) and non-experts (*red line with stars*)

psychologically meaningful categories (Pennebaker et al. 2001). LIWC-2001 defines 68 different categories, each of which contains several dozens to hundreds of words. Given each user's tweets, we measured his linguistic characteristics in the 68 categories by computing his score of each category based on LIWC dictionary. First we counted the total number of words in his tweets (N). Next we counted the number of words in his tweets overlapped with the words in each category i on LIWC dictionary (C_i). Then, we computed his score of a category i as C_i/N.

Figure 3 shows linguistic characteristics of experts and non-experts in three categories such as Home (e.g., house, kitchen, lawn), TV/Movies (e.g., TV, sitcom, cinema) and Touch (e.g., touch, hold, felt). While experts favor words of the lexical fields, non-experts rarely used these words.

In summary, we have analyzed behaviors of experts and non-experts by their activity and linguistic characteristics. We observed that their behaviors were clearly different. These observations motivated us to do further study.

4 Detection of experts

In the previous section, we have analyzed behaviors of experts and non-experts and found that their activities and linguistic characteristics are different. Based on the observations, in this section we propose features, measure distinguishing power of the features, and then develop and test expert classifiers.

4.1 Features

To build an expert classifier, we need to convert user profile information to meaningful feature values. Based on our previous analysis and observations, we propose 78 features as shown in Table 2 and grouped them into the following 3 categories:

Table 2 Features

Group	Feature
AF	The number of edits
AF	The number of posted questions
AF	The number of followers
AF	The percentage of bidirectional friends: $\frac{\lvert\text{followees} \cap \text{followers}\rvert}{\lvert\text{followees}\rvert}$
QAF	The average number of words in posted answers
QAF	The average number of uppercase words in posted answers
QAF	Subjectivity of answers: the average number of subjective words in posted answers
QAF	Average upvotes: the average number of upvotes in posted answers
QAF	Entropy of answers
QAF	Readability of answers
LF	68 LIWC features, which are Total Pronouns, 1st Person Singular, 1st Person Plural, 1st Person, 2nd Person, 3rd Person, Negation, Assent, Articles, Prepositions, Numbers, Affect, Positive Emotions, Positive Feelings, Optimism, Negative Emotions, Anxiety, Anger, Sadness, Cognitive Processes, Causation, Insight, Discrepancy, Inhibition, Tentative, Certainty, Sensory Processes, Seeing, Hearing, Touch, Social Processes, Communication, Other References to People, Friends, Family, Humans, Time, Past Tense Verb, Present Tense Verb, Future, Space, Up, Down, Inclusive, Exclusive, Motion, Occupation, School, Job/Work, Achievement, Leisure, Home, Sports, TV/Movies, Music, Money, Metaphysical States, Religion, Death, Physical States, Body States, Sexual, Eating, Sleeping, Grooming, Swearing, Nonfluencies, and Fillers

- *Activity Features* (*AF*): These features measure a user's activities on Quora. They consist of the number of edits, the number of questions, the number of followers and the number of bidirectional friends. The first three features indicate activity levels of the user. The last feature measures how many bidirectional friends the user has. In the bidirectional friends feature, we chose the denominator as # of followees because our intent is to measure what percent of followees also followed the user back.

- *Quality of Answer Features* (*QAF*): These features are extracted from a user's aggregated answers. The features consist of the average number of words, the average number of uppercase words, the average number of subjective words in answers, entropy of answers and readability of answers and the average number of upvotes in posted answers. To count the number of subjective words, we used the Subjectivity Lexicon (Riloff and Wiebe 2003) which consists of 8222 subjective words collected from various information sources.
We measured the complexity of answers by the entropy of the words in the answers:

$$\text{entropy}(a_j) = -\sum_{i=1}^{k} P(x_i)\log P(x_i), \qquad (1)$$

where k is the number of distinct words in answers, and $P(x_i)$ is $\frac{\text{frequency of a word } i}{\text{total number of words } n \text{ in answers}}$. A low entropy score indicates that answers contain a few words or repetitive words. A high entropy score indicates that a user's answers contain various words and are complex. In other words, the user with a high entropy score has knowledge to use various words, and know how to present complex or complicated ideas.
The readability of aggregated answers was measured by the following SMOG formula:

$$1.043\sqrt{|\text{polysyllables}| \times \frac{30}{|\text{sentences}|}} + 3.1291 \qquad (2)$$

The SMOG grade estimates the years of education needed to understand a piece of writing (McLaughlin 1969).

- *Linguistic Features* (*LF*) Researchers have found that word usage in one's writings is related to one's personality or linguistic characteristics (Fast and Funder 2008; Gill et al. 2009; Mahmud et al. 2013). Instead of creating the whole lexicon by TF-IDF (i.e., extracting general text features such as unigram, bigram and trigram), we measured 68 linguistic features by LIWC (Pennebaker et al. 2001) as shown in Table 2 so that we can know how experts and non-experts are different in terms of their personalities. Each feature is a word category which contains up to hundreds of words selected by psychologists. We extract these features from answers posted by a user. Detailed information regarding how we calculated these features was described in the previous section.

4.2 Feature selection and analysis

Before building classifiers, we conduct feature selection to make sure only using features having positive distinguishing power between experts and non-experts. To measure discriminative power of our proposed 78 features, we computed the χ^2 value (Yang and Pedersen 1997) of each of the features. Our χ^2 test results showed that all features had positive discriminative power, though with different relative strengths.

Next, we measured the Mean Decrease Accuracy (MDA) of Random Forests which is another method to measure importance features. The larger its mean decrease accuracy is, the more important a feature is. We measured MDA of all features in the combined dataset in Table 1. Figure 4 shows top 5 important features—the number of edits, the number of followers, entropy of answers, TV/Movies and Grooming in LIWC.

While we were measuring MDA of features in the combined dataset, an interesting question was raised. "Does each topic-based dataset have a different set of top features?" To answer this question, we measured MDA of all features in each topic-based dataset presented in Table 1. Figure 5 shows the experiential results. Interestingly, important features varied across the topic-based datasets. A commonly important feature was the number of edits. Some LIWC features like TV/Movies and Home and Touch in Sports, and Grooming and Family in Politics were considered as important features. We conjecture that while experts answered sports-related questions, they might express what they watched (e.g., NFL games) on TV with some feelings like some sadness and joy because these experts wanted to deliver detailed information and feelings regarding the sports.

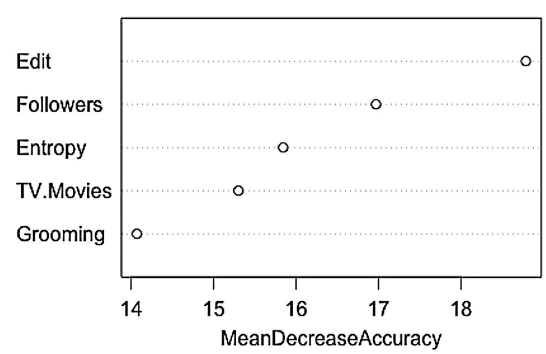
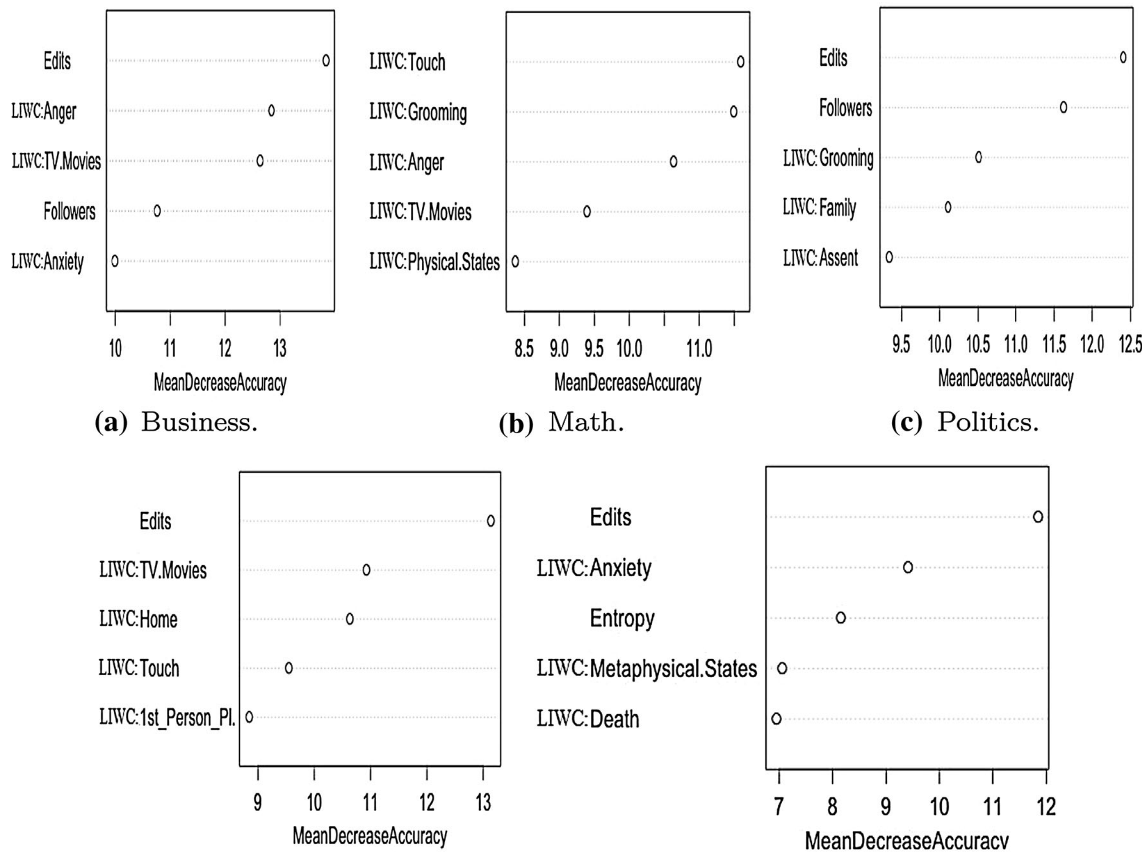


Fig. 4 Top 5 features in the combined dataset



(a) Business. **(b)** Math. **(c)** Politics.

(d) Sports. **(e)** Technology.

Fig. 5 Top 5 features in each of the five topic-based datasets

4.3 Experiments

So far, we have learned that all of the proposed features have positive discriminative powers, and each topic-based dataset has had a different order of important features. Based on this analysis and observation, now we turn to develop classifiers to see whether automatically detect experts in the *combined dataset* (i.e., containing general topics—multiple topic-based datasets) is possible. Further, we develop topic-specific classifiers to test whether we can detect experts in each *topic-based dataset*.

Evaluation Metrics To evaluate a classifier, we compute accuracy and area under the ROC curve (AUC).

In the confusion matrix, Table 3, *a* represents the number of correctly classified experts, *b* represents the number of experts misclassified as non-experts, *c* represents the number of non-experts misclassified as experts, and *d* represents the number of correctly classified non-experts. The accuracy means the fraction of correct classifications and is $(a + d)/(a + b + c + d)$.

Baselines We propose two baselines, and compare them with our approach in terms of accuracy and AUC. Baseline 1 is always to predict a user's class as the majority case (i.e., non-expert). Baseline 2 is to build a classifier based on only the best feature identified at the feature selection step. We apply the baselines to combined dataset and topic-based datasets.

Combined Dataset We chose three classification algorithms such as Random Forest, J48 and SMO (SVM) to compare how their classification performances are different and which one is the best. We used Weka (Hall et al. 2009), a machine learning toolkit consisting of implementations of these algorithms. First, we converted the *combined dataset* consisting of profiles of 432 experts and 3288 non-experts in Table 1 to feature values, and ran tenfold cross-validation for each classification algorithm. Table 4 shows classification results of the combined dataset by measuring AUC and accuracy of each classification algorithm. Random Forest classifier outperformed J48 and SMO classifiers, achieving 0.979 AUC and 95.94 % accuracy.

Next, we compared the performance of our approach against the two baselines. Note that baseline 2 and our approach used Random Forest algorithm. Table 5 shows experimental results of two baselines and our approach. Baselines 1 and 2 achieved 88.3 % and 93.1 % accuracy while our approach achieved 95.9 % accuracy. Overall, our

Table 3 Confusion matrix

Actual	Predicted	
	Expert	Non-expert
Expert	*a*	*b*
Non-expert	*c*	*d*

Table 4 Classification results of the combined dataset

Classifier	AUC	Accuracy (%)
J48	0.843	94.00
Random forest	**0.979**	**95.94**
SMO	0.509	88.52

Bold indecates an approach achieving the best result/performance

Table 5 Two baselines vs. our approach in the combined dataset

Classifier	AUC	Accuracy (%)
Baseline 1		88.3
Baseline 2	0.775	93.1
Our approach	**0.979**	**95.9**

Bold indecates an approach achieving the best result/performance

approach improved accuracy of two baselines by 8.6 % ($=\frac{95.9\times100}{88.3}-100$) and 3 % ($=\frac{95.9\times100}{93.1}-100$). Our approach also improved AUC of baseline 2 by 26.3 % ($=\frac{0.979\times100}{0.775}-100$). The experimental results confirmed that our approach outperformed two baselines.

Topic-based Datasets As we described in our data collection strategy in Sect. 2, we intentionally collected five topic-based datasets—Technology, Politics, Sports, Mathematics and Business. Users in each dataset were interested in the topic and followed the topic. For example, experts were interested in the topic and wanted to observe what kind of questions had been posted in this topic thread. An interesting research question is "can we detect a topic-specific experts by using the classification approach"? As we observed in the previous subsection, importance of features varied in each topic-based dataset. Developing topic-specific classifiers would make sense. To answer the research question, we developed three classifiers in each topic-based dataset.

Table 6 shows 15 classifiers' experimental results after running tenfold cross-validation. Overall, Random Forest classifier outperformed J48 and SMO classifiers in all five topic-based datasets. Random Forest classifiers achieved 96.37, 94.20, 96.61, 96.91 and 95.42 % in Business, Mathematics, Politics, Sports and Technology, respectively. Especially, topic-specific classifiers for Sports and Technology achieved higher accuracies than the classifier built based on the combined dataset. These results show that building statistical models in each topic is possible, and the models work well in detecting topic-specific experts.

Table 6 Classification results of the five topics-based datasets

Topic	Classifier	AUC	Accuracy (%)
Business	J48	0.803	95.07
	Random forest	**0.976**	**96.37**
	SMO	0.520	90.80
Mathematics	J48	0.777	91.89
	Random forest	**0.972**	**94.20**
	SMO	0.500	87.90
Politics	J48	0.83	93.22
	Random forest	**0.983**	**96.61**
	SMO	0.512	86.79
Sports	J48	0.811	92.52
	Random forest	**0.987**	**96.91**
	SMO	0.529	87.15
Technology	J48	0.767	91.71
	Random forest	**0.951**	**95.42**
	SMO	0.499	90.14

Bold indecates an approach achieving the best result/performance

Next, we compared our approach with two baselines. Table 7 shows experimental results of two baselines and our approach. Overall, our approach outperformed the two baselines regardless of a topic.

In summary, we have thoroughly analyzed our proposed features, and developed two types of classifiers—(i) a universal classifier to detect experts in general topics (containing multiple topics); and (ii) topic-based classifiers to detect topic-specific experts. Both types of classifiers worked well, and achieved up to 96 % accuracy and 0.979 AUC.

5 Detecting experts with temporal features

In the previous section, we developed classifiers based on static features extracted from a snapshot of user profiles. In this section, we are interested in studying temporal behaviors of experts and non-experts. Do they have clearly different temporal patterns? If yes, can we use these temporal patterns to improve the performance of expert classifiers?

Data Collection To answer these questions, we collected another dataset presented in Table 8 in November 2014. Our data collection strategy is that first we randomly selected 786 users. Then we collected their profiles once per day during 22 consecutive days. In other words, each day we got one snapshot of each user, in total we collected 22 snapshots of each user. Our intuition is 3 weeks data

Table 7 Two baselines vs. our approach in the five topics-based datasets

Topic	Classifier	AUC	Accuracy (%)
Business	Baseline 1		90.4
	Baseline 2	0.842	93.1
	Our approach	**0.976**	**96.3**
Mathematics	Baseline 1		87.9
	Baseline 2	0.739	91.7
	Our approach	**0.972**	**94.2**
Politics	Baseline 1		86.6
	Baseline 2	0.813	93.4
	Our approach	**0.983**	**96.6**
Sports	Baseline 1		86.6
	Baseline 2	0.800	91.5
	Our approach	**0.987**	**96.9**
Technology	Baseline 1		90.2
	Baseline 2	0.720	93.7
	Our approach	**0.951**	**95.4**

Bold indecates an approach achieving the best result/performance

Table 8 Another dataset containing 22 user profile snapshots of each user

|Experts|	|Non-experts|	|Users|
114	672	786

containing daily snapshot would reveal interesting patterns of experts, and between experts and non-experts like daily patterns and weekly patterns. Using the labeling method in Sect. 2, we got the ground truth. Finally, the dataset consisted of user profiles (22 user profile snapshots of each user) of 114 experts and 672 non-experts.

Analysis of temporal behaviors of experts and non-experts Next, we analyze temporal behaviors of experts and non-experts. First, to understand what types of experts exist on Quora in terms of different temporal behaviors, we calculated two variables—(i) the average value of weekly change of the number of answers; and (ii) standard deviation of weekly change of the number of answers. We calculated these two variable values for each expert. Then, we grouped the experts to three categories as shown in Fig. 6:

- *Fluctuating Experts* (18.4 %): The fluctuating experts have posted different number of new answers each week. For example, these experts posted less number of answers in the first week. Then they posted more in the second week and posted a little less number of answers. We found 21 (18.4 %) fluctuating experts in the dataset.
- *Stable* (*and Active*) *Experts* (54.4 %): The stable experts are very active experts who post almost similar number of answers every week. They constantly provided answers and played a prominent role on disseminating knowledge on Quora. We found 62 (54.4 %) stable experts in the dataset.
- *Idle Experts* (27.2 %): The idle experts have posted far fewer answers constantly every week. They posted high quality answers, but posted far fewer answers recently. We conjecture that they used to be active and posted high quality answers, but might lose passion on posting more answers on Quora. Q&A service providers should think of how to motivate them to become active (stable) experts again. We found 31 (27.2 %) idle experts in the dataset.

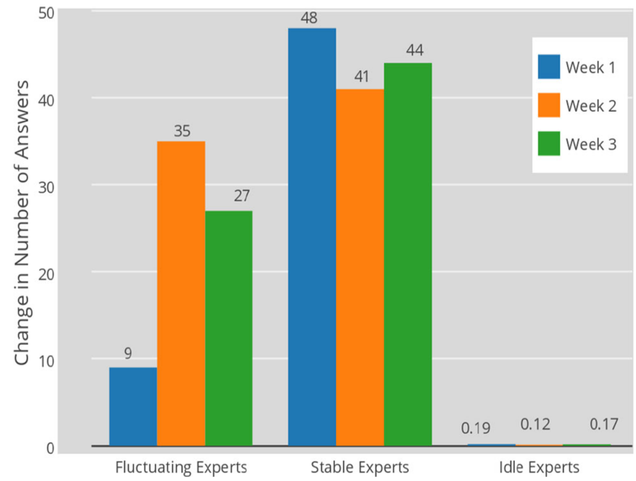
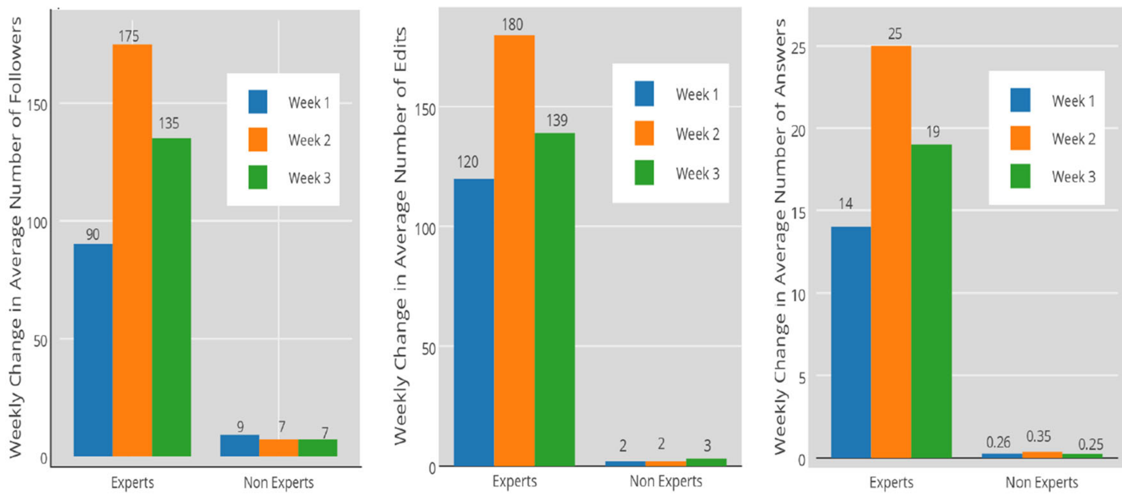


Fig. 6 Three types of experts grouped by their weekly change of the number of answers

Note that we also tried to understand what types of non-experts exist on Quora but we did not find interesting patterns/groups among non-experts.

Second, we analyze how the average number of followers, edits and answers of experts and non-experts had been changed over time. To do this, we measured weekly change of followers, edits and answers. Figure 7 shows weekly change results of experts and non-experts. Experts and non-experts strongly differ in terms of change in the average number of followers as shown in Fig. 7a. Experts had a larger number of weekly change in the average number of followers than non-experts. More number of other users followed experts than non-experts. We conjecture that other users tend to follow experts to get useful information. Weekly change in the average number of edits and answers in Fig. 7b and c also followed similar patterns with weekly change in the average number of followers. Experts were more active than non-experts by making a larger number of edits. Experts posted more answers than non-experts during the period of 22 days. Of course, we observed that some non-experts had increased the number of answers over time. We conjecture that these non-experts have a high chance to become experts in the future. The



(a) Weekly change in the average number of followers. **(b)** Weekly change in the average number of edits. **(c)** Weekly change in the average number of answers.

Fig. 7 Weekly change in the average number of followers, edits and answers

temporal data analysis presents that experts and non-experts have different temporal behaviors. With the positive observations, next we extract temporal features toward building an expert classifier.

Temporal features In each user profile snapshot, we extracted 5 variable values—the number of followees, followers, edits, questions and answers. By doing this for 22 user profile snapshots of each user, we got 22 time-series values of each variable (e.g., number of followees, number of followers, number of edits). Then, we computed following temporal features for each variable:

- *Average daily change* (*in total* 5 *features*): From 22 time-series values of each variable, we calculated daily change (increase/decrease) between two consecutive days. Then we averaged these values. Finally, we got 5 average daily change features each of which was calculated from each variable.
- *Standard deviation of daily change* (*in total* 5 *feature*): We measured standard deviation of 21 daily changes of each variable.
- *Probability of average daily change on the day of a week* (*in total* 35 *features*): These features capture the day's average change. In this context, the day of a week means Monday, Tuesday, Wednesday, Thursday, Friday, Saturday or Sunday. From each variable, we calculated 7 features (in total, 35 features).

Overall, we extracted 45 temporal features.

Experiments Next, we are interested in testing whether adding these temporal features to the existing static feature set improves the performance of expert detection. To answer this research question, we conducted two experiments—

Table 9 Classification results without/with temporal features

Random forest classifier	AUC	Accuracy (%)
Without temporal features	0.982	95.92
With temporal features	**0.986**	**96.56**

Bold indecates an approach achieving the best result/performance

evaluate (1) the performance of classifiers based on the existing static features without temporal features; and (2) the performance of classifiers based on the existing static features with temporal features. We ran tenfold cross-validation for each classifier. Table 9 shows experimental result of Random Forest classifier without temporal features. Random Forest classifier *without* temporal features achieved 0.982 AUC and 95.92 % accuracy. Then, we added 45 temporal features to the existing feature set and developed another Random Forest classifier. As shown in Table 9, Random Forest classifier *with* temporal feature achieved 0.986 AUC and 96.56 % accuracy. Based on these experiments, Random Forest classifier *with* temporal features outperformed Random Forest classifier *without* temporal features by additionally increasing 0.64 % accuracy.

6 Discussion: adding features extracted from an external source—Twitter

Some people have accounts in multiple sites such as social Q&A sites like Quora and social media sites such as Twitter and Facebook. Sometimes people link their own accounts each other. We came up with interesting research questions. How many people link their external accounts to their Quora profiles? Will collecting a user's information in external sources like social media sites and extracting features from these external sources help improving the performance of expert detection?

To answer these research questions, we analyzed what percent of Quora users in Table 1 linked their accounts in external sites to their Quora profile pages. We found that 60 % of the users linked URLs of their Twitter and Facebook profile pages to their Quora profile pages. Then, we collected their Twitter account information such as user profile, recent 200 tweets, a list of followees and a list of followers. From the Twitter data, we extracted profile features like the number of followees, the number of followers and the number of posted tweets, linguistic features (based on LIWC) extracted from the recent 200 tweets, and the user's Klout score to measure her influence on Twitter network (Klout 2013). We added Twitter features to the Quora feature set, and built Random Forest classifier.

Experimental result showed that adding Twitter features did not improve the performance of expert detection. While Quora is a place where people share their knowledge with detailed information regarding specific questions, Twitter is a place where people share personal thoughts, breaking news, sentiments regarding products or politics, and etc in a brief format. Because of these reasons, we conjecture that adding features extracted from Twitter did not improve the performance of the expert classifier.

7 Related work

Social Q&A sites have been used by many people for years. As people have shared their curiosities, questions, information and knowledge with others in the social Q&A sites, researchers have conducted research to solve various research problems in these sites. In this section, we summarize existing research work by three research problems.

First research problem is to understand what kind of social Q&A sites exists and what people do in these Q&A sites. Harper et al. (2008) have categorized Q&A sites to three types—digital reference services, ask an expert service, community Q&A sites. They further analyzed how these three types of Q&A sites are different in terms of responsiveness about questions. Furtado et al. (2014) have analyzed contributors' activity on the Stack Exchange Q&A platform by clustering contributors' profiles. They have measured the quality and quantity of contribution of these users. Wang et al. (2013) suggested that user-topic model produced user interest in searching and answering questions on Quora.

Second research problem is to measure quality of an answer. Su et al. (2007) examined the quality of answers in Q&A sites. Jeon et al. (2006) built a model for detecting the quality of answer based on features derived from the particular answer being examined. Zhu et al. (2009) examined the quality of answers in Q&A sites. Zhou et al. (2012) proposed a joint learning method to measure quality of an answer. Toba et al. (2014) examined a question type to select a right answer. Paul et al. (2012) studied how people evaluate quality of an answer.

Third research problem is to measure expertise of users or detect experts based on a question. Kao et al. (2010) proposed a hybrid approach to effectively find expertise of users in different categories of the target question in Q&A sites. They used user's reputation, subject relevance and their authority of a category in detecting experts. Bouguessa et al. (2008) model the expertise of users based on the number of best answers in Yahoo Answers. Zhang et al. (2007) proposed Z-score measure to calculate the expertise level of users in Q&A sites. Some researchers measured expertise of users by analyzing their link structure using PageRank and HITS (Jurczyk and Agichtein 2007; Schall and Skopik 2011; Zhou et al. 2014). Other researchers studied how to detect experts based on a question in Q&A sites. Pal et al. (2011), Pal et al. (2012) proposed a probabilistic model that captures the selection preferences of users based on the questions they choose. Liu et al. (2013) proposed a hybrid approach to find experts for the category of a target question. Luo et al. (2014) studied to recommend answerers in an enterprise social Q&A system. The most relevant work was performed by Song et al. (2013). In the study, they collected data from Quora and used limited number of features (e.g., authority, activity and influence features) to build a ranking model for identifying leading users. But, their approach has a drawback, achieving low precision when top K value was increased. For example, they only achieved 80 % precision in top 20 results.

Compared with the previous research work, we proposed rich feature set including user activity features, quality of answer features, linguistic features, temporal features and social network features (i.e., Twitter features in this study). In addition, we linked Quora accounts to Twitter accounts to extract additional information, and we tested whether adding social network features would be helpful to improve expert detection rate. We developed statistical models to detect experts for general topics and specific topics. This research will complement the existing research work.

8 Conclusion

In this manuscript, we have presented analysis of behaviors of experts and non-experts on Quora and identified four types of features such as user activity features, quality of answer features, linguistic features based on LIWC and temporal features. Then, we measured what features are top features in general topics and a specific topic. Based on this analysis and the observations, we proposed and developed statistical classification models to automatically identify experts. Our experimental results showed that these models effectively detected experts in general topics with 0.979 AUC and 95.94 % accuracy, and in a specific topic with up to 0.987 AUC and 97 % accuracy. We also studied whether adding temporal features would improve the performance of expert detection. The experimental results showed that adding temporal features further improved the performance of expert detection by additionally increasing 0.64 % accuracy.

References

Alexa (2015) Quora.com site overview—Alexa. http://www.alexa.com/siteinfo/quora.com

Bouguessa M, Dumoulin B, Wang S (2008) Identifying authoritative actors in question-answering forums: The case of yahoo! answers. In: Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining

Fast LA, Funder DC (2008) Personality as manifest in word use: correlations with self-report, acquaintance report, and behavior. J Pers Soc Psychol 94(2):334–346

Furtado A, Oliveira N, Andrade N (2014) A case study of contributor behavior in Q&A site and tags: the importance of prominent profiles in community productivity. J Braz Comput Soc 20(1):5

Gill AJ, Nowson S, Oberlander J (2009) What are they blogging about? personality, topic and motivation in blogs. In: Proceedings of the 3rd International AAAI Conference on Weblogs and Social Media

Hall M, Frank E, Holmes G, Pfahringer B, Reutemann P, Witten IH (2009) The Weka data mining software: an update. SIGKDD Explor Newsl 11(1):10–18

Harper FM, Raban D, Rafaeli S, Konstan JA (2008) Predictors of answer quality in online Q&A sites. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems

Jeon J, Croft WB, Lee JH, Park S (2006) A framework to predict the quality of answers with non-textual features. In: Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval

Jurczyk P, Agichtein E (2007) Discovering authorities in question answer communities by using link analysis. In: Proceedings of the sixteenth ACM conference on Conference on information and knowledge management

Kao WC, Liu DR, Wang SW (2010) Expert finding in question-answering websites: a novel hybrid approach. In: Proceedings of the 2010 ACM Symposium on Applied Computing

Klout (2013) See how it works—Klout. http://klout.com/corp/how-it-works

Liu DR, Chen YH, Kao WC, Wang HW (2013) Integrating expert profile, reputation and link analysis for expert finding in question-answering websites. Inf Process Manag 49(1):312–329

Luo L, Wang F, Zhou MX, Pan Y, Chen H (2014) Who have got answers?: growing the pool of answerers in a smart enterprise social QA system. In: Proceedings of the 19th international conference on Intelligent User Interfaces

Mahmud J, Zhou MX, Megiddo N, Nichols J, Drews C (2013) Recommending targeted strangers from whom to solicit information on social media. In: Proceedings of the 2013 international conference on Intelligent user interfaces

McLaughlin HG (1969) SMOG grading—a new readability formula. J Read 12(8):639–646

Pal A, Farzan R, Konstan JA, Kraut RE (2011) Early detection of potential experts in question answering communities. In: Proceedings of the 19th international conference on User modeling, adaption, and personalization

Pal A, Harper FM, Konstan JA (2012) Exploring question selection bias to identify experts and potential experts in community question answering. ACM Trans Inf Syst 30(2):10:1–10:28

Paul SA, Hong L, Chi EH (2012) Who is authoritative? understanding reputation mechanisms in Quora. In: Collective intelligence

Pennebaker J, Francis M, Booth R (2001) Linguistic inquiry and word count. Erlbaum Publishers, Mahwah

Quora (2014) How are top writers chosen? what are the criteria for being chosen as a top writer? http://www.quora.com/How-are-Top-Writers-chosen-What-are-the-criteria-for-being-chosen-as-a-Top-Writer

Riloff E, Wiebe J (2003) Learning extraction patterns for subjective expressions. In: Proceedings of the conference on empirical methods in natural language processing

Schall D, Skopik F (2011) An analysis of the structure and dynamics of large-scale q/a communities. In: Proceedings of the 15th international conference on advances in databases and information systems (ADBIS)

Song S, Tian Y, Han W, Que X, Wang W (2013) Leading users detecting model in professional community question answering services. In: Proceedings of International Conference on Green Computing and Communications and Internet of Things and Cyber, Physical and Social Computing

Su Q, Pavlov D, Chow JH, Baker WC (2007) Internet-scale collection of human-reviewed data. In: Proceedings of the 16th international conference on World Wide Web

Toba H, Ming ZY, Adriani M, Chua TS (2014) Discovering high quality answers in community question answering archives using a hierarchy of classifiers. Inf Sci 261:101–115

Wang G, Gill K, Mohanlal M, Zheng H, Zhao BY (2013) Wisdom in the social crowd: an analysis of Quora. In: Proceedings of the 22nd international conference on World Wide Web

Yang Y, Pedersen JO (1997) A comparative study on feature selection in text categorization. In: Proceedings of the Fourteenth International Conference on Machine Learning

Zhang J, Ackerman MS, Adamic L (2007) Expertise networks in online communities: structure and algorithms. In: Proceedings of the 16th international conference on World Wide Web

Zhou G, Liu K, Zhao J (2012) Joint relevance and answer quality learning for question routing in community QA. In: Proceedings of the 21st ACM international conference on Information and knowledge management

Zhou G, Zhao J, He T, Wu W (2014) An empirical study of topic-sensitive probabilistic model for expert finding in question answer communities. Knowl Based Syst 66:136–145

Zhu Z, Bernhard D, Gurevych I (2009) A multi-dimensional model for assessing the quality of answers in social Q&A sites. In: Proceedings of 14th International Conference on Information Quality