Open access • Book Chapter • DOI:10.1007/978-3-642-20161-5_9

Modeling answerer behavior in collaborative question answering systems

Qiaoling Liu, Eugene Agichtein

Institutions: Emory University

Published on: 18 Apr 2011 - European Conference on Information Retrieval

Topics: Question answering

Modeling Answerer Behavior in Collaborative Question Answering Systems

Qiaoling Liu and Eugene Agichtein

Emory University, Atlanta, USA
{qiaoling.liu,eugene}@mathcs.emory.edu

Abstract. A key functionality in Collaborative Question Answering (CQA) systems is the assignment of the questions from information seekers to the potential answerers. An attractive solution is to automatically recommend the questions to the potential answerers with expertise or interest in the question topic. However, previous work has largely ignored a key problem in question recommendation - namely, whether the potential answerer is likely to *accept* and *answer* the recommended questions in a *timely manner*. This paper explores the contextual factors that influence the *answerer behavior* in a large, popular CQA system, with the goal to inform the construction of question routing and recommendation systems. Specifically, we consider *when* users tend to answer questions in a large-scale CQA system, and *how* answerers tend to choose the questions to answer. Our results over a dataset of more than 1 million questions draw from a real CQA system could help develop more realistic evaluation methods for question recommendation, and inform the design of future question recommender systems.

1 Introduction

A key step in Collaborative Question Answering (CQA) systems is matching the posted questions to the best answerers who can contribute the needed information. In the existing systems such as Yahoo! Answers or Naver, this step is performed by the answerers manually, who choose which questions to answer based on widely varying and often subjective criteria[20]. However, this often leads to inefficiencies, redundancies, and often delayed or poor quality answers, which in turn degrades the experience for the question submitters.

An attractive solution is to automatically *recommend* the questions to the potential answerers, usually based on the expertise and/or past performance of these users for similar questions (e.g.,[16,9,5,13]). However, previous work has largely ignored the key problem in question recommendation - namely, whether the potential answerer is likely to accept and answer the questions recommended to them in a timely manner. That is, even if the question is on a topic of past interest to the answerer, they may not have the opportunity or interest in answering the question *at recommendation time*.

This paper addresses this gap by exploring the contextual factors that influence the *answerer behavior* in a large, popular CQA system, with the goal to inform the construction of real-time, online question routing and recommendation

P. Clough et al. (Eds.): ECIR 2011, LNCS 6611, pp. 67–79, 2011.

systems that also take into account the behavior of real answerers. Specifically, we consider the following research questions:

1. *When* do users tend to answer questions in a web-scale CQA system?
2. *How* do users tend to choose the questions to answer in CQA?

Our overall approach is to analyze the real answering behavior of a large group of Yahoo! Answers users, collected for more than 1 million questions over a period of one month in early 2010. Specifically, for the first research question, we analyzed both the *overall* and *user-specific* temporal activity patterns, identifying stable daily and weekly periodicities, as well as not previously observed *bursty* patterns of activity in the individual answer sessions of many users. We exploit this observation to perform a novel *session-based* analysis of the answerer activity. For the second research question, we analyze the factors that may affect the users' decisions of which questions to answer. These factors include the question category (topic), the question position in the list shown to users, and the surface patterns in the question text. We confirmed previous findings that users have "favorite" categories that attract most of their contributions, but interestingly the decisions for most of the users *within* a category are determined more by the *rank position* of the question in the list of available questions, than any other factors such as the text or the provenance of the question itself.

As far as we know, this is the first reported large-scale analysis of *answerer behavior* on session level. Our results identify new temporal patterns of contributor participation in CQA, and shed light on how the participants make minute-by-minute decisions during their online sessions. Our results could help develop more accurate answerer behavior and prediction models; allow the development of more realistic evaluation methodology for question recommendation; and inform the design of future question recommender systems.

In the next section, we overview the related work. Then, Section 3 describes our dataset in more detail, and explores the temporal patterns on *when* users tend to answer questions. Section 4 then discusses our findings of *how* the answerers tend to choose which questions to answer. Section 5 summarizes our results and concludes the paper.

2 Related Work

Collaborative Question Answering (CQA) systems are attracting increasing research effort in information retrieval and HCI communities (e.g., [2,14,1,20,16,7]). A key to the success of CQA systems is to provide askers with efficient and helpful service, by minimizing the time that askers need to obtain good answers for their questions. There are generally three ways to better achieve this goal. First, the large volumes of existing content in CQA systems can be reused to satisfy an asker's information need, based on effective retrieval of relevant questions and answers to the information need [14,7,6,4,22].

Secondly, an attractive approach to improve the answer quality for CQA askers is to route the new questions to experts on the topic of the question, which has been an active area of research. For example, Jurczyk et al. [16] formulated a graph structure for CQA systems and applied a web link analysis algorithm to discover authoritative users in topical categories. Liu et al. [19] cast the expert finding problem as an IR problem, by viewing a new question as a query to retrieve the user profiles as "documents", and tested several language models for expert ranking. Bouguessa et al. [5] focused on automatically discriminating between authoritative and non-authoritative users by modeling users' authority scores as a mixture of gamma distributions for each topic. Beyond the CQA context, there has also been extensive work on expert finding in online forums such as [23,24].

To further reduce the response time to new questions in CQA, additional question routing methods that consider user interests have been proposed [9,21,17,13,24]. For example, Guo et al. [9] developed a probabilistic generative model to obtain latent topics for questions and users, and incorporated both the topic-level and term-level information for recommending new questions to potential answerers. Qu et al. [21] applied PLSA to capture user interests in terms of topics based on their answering history, and Liu et al. [17] employed an integration of the language model and the Latent Dirichlet Allocation model for measuring the relationship between an answerer and a question, which also considered user activity and authority information. Horowitz et al. [13] addressed the question routing problem in a real-time CQA system by considering both user interest and social connectedness.

A third way to reduce the waiting time for CQA askers is simply to encourage more answerers, which depends on better understanding of the answerer behavior. Some previous work has been done on understanding user behavior in CQA [1,20,11,18,10,3]. For example, Adamic et al. [1] analyzed the content properties and user interaction patterns across different Yahoo! Answers categories. Gyöngyi et al. [11] studied several aspects of user behavior in Yahoo! Answers, including users' activity levels, interests, and reputation. Guo et al. [10] studied the patterns of user contributions in knowledge sharing-oriented Online Social Networks including a question answering social network. Nam et al. [20] investigated the motivation of top answerers in Naver - a popular Korean CQA system, including altruism, learning, competence and points. Liu et al. [18] explored the effects of an answerer's Web browsing context on the effectiveness of CQA systems. Aperjis et al. [3] studied the speed-accuracy tradeoff faced by CQA askers, i.e., maximizing the information accuracy while minimizing the waiting time.

As far as we know, ours is the first analysis of the session-level patterns in the answerer behavior - which could provide useful information for improving all of the question recommendation methods above.

3 Temporal Patterns in Answerer Behavior

This section first describes the CQA dataset used throughout this paper. It then shows the aggregate patterns of *when* answerers tend to answer questions, which confirms previous findings and validate our dataset construction. Then, we focus on the novel contribution of this paper, namely modeling the *individual* answerer

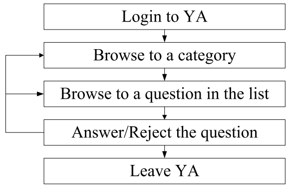


Fig. 1. Basic question answering process in Yahoo! Answers

Table 1. Dimension of the Yahoo! Answers dataset. The USER20 dataset focuses on answerers with at least 20 answers, which is used in the rest of the paper.

	Questions	Answers	Best Answers	Answerers	Askers	Users
ALL	1,056,945	4,734,589	1,056,945	433,902	466,775	726,825
USER20	933,746 (∼88%)	3,319,985 (∼70%)	751,633 (∼71%)	45,543 (∼10%)	419,395 (∼90%)	437,493 (∼60%)

activity within a single answer session. These analysis could help question recommender systems by suggesting *when* to begin recommending questions to a user in the first place, and *how many* questions to recommend to a user.

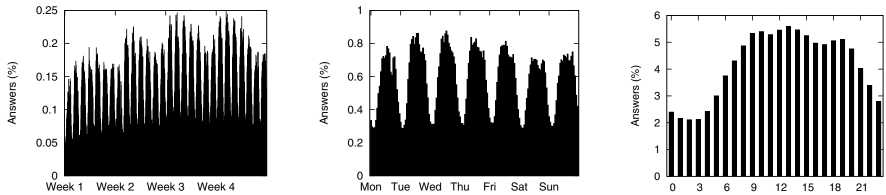
3.1 The Yahoo! Answers Setting and Data

For this study we chose the Yahoo! Answers (YA) website, as a large-scale, popular, and representative example of a CQA system. To clarify our terminology and the subsequent descriptions, we briefly summarize the basic question answering model in YA, which we believe is representative of many other CQA sites. Figure 1 illustrates this process. After logging into the YA site, answerers can choose a category of interest to them to browse (including the root category "All categories"). Then they are shown a list of questions in that category among which they may answer some and skip others. This process is repeated when answerers browse to another category, until they eventually leave the site.

To construct the dataset, we crawled about 1M questions and 4.7M answers covering the top 26 categories and 1659 leaf categories in YA, as of May 2010. Since inactive users reveal less information of their answering behavior, our analysis focuses on *active answerers* who posted at least 20 answers during the period of time. This subset, called USER20, includes 45,543 answerers, accounting for about 10% of all answerers but 70% of all answers and best answers. Table 1 presents the statistics of the dataset in more detail.

3.2 Aggregate Temporal Pattern Analysis

First, we analyze the overall temporal patterns of answering activities in Yahoo! Answers with the strategy used in [10]. We bin all the answers by hours, aggregate answers in the same hours by months/weeks/days, and then normalize the number of answers in each hour by the total number of answers. Based on our dataset, the answer activities in YA demonstrate strong monthly, weekly



(a) Monthly Pattern (b) Weekly Pattern (c) Daily Pattern

Fig. 2. Temporal patterns of answer activities in YA, showing the percentage of answers in the same hours aggregated by (a)months; (b)weeks; (c)days

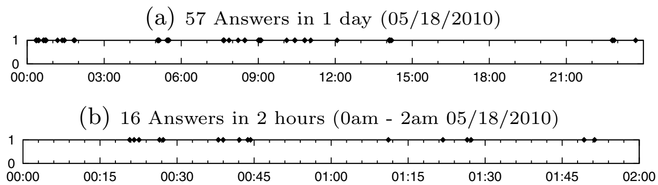


Fig. 3. Example answering behavior for an active user over 1 day (a) and over a period of 2 hours (b)

and daily patterns as shown in Figure 2. From Figure 2(a), we can see that the number of answers across the whole month is increasing, which indicates the growing popularity of YA. Figure 2(b) shows that the number of answers during the weekday is higher than that on the weekends, with Tuesdays and Wednesdays being the most active weekdays. Based on Figure 2(c), there tend to be three peak times in a day for answering questions, 10:00, 13:00, and 19:00 (YA server time). The least active time for answering questions is 2:00-3:00 AM. These results are similar to those described in [10], but with a time shift in the daily pattern possibly due to a different time zone used in their study.

3.3 Burstiness of Individual Answering Activities

We now explore the temporal patterns of answering activity for individual users. We found that users tend to post bursts of answers within short answering sessions, and then "disappear" for relatively long periods of inactivity. For example, Figure 3(a) illustrates the answer activities of an example user. The user answered 57 questions that day; however, the answering was not distributed uniformly, but was concentrated in relatively short *bursts*. To provide a better intuition, we plot the user's answering activities over a period of two hours, shown in Figure 3(b). We can see that some intervals between two successive answers are short (e.g. less than 3 minutes), but others are long (e.g. around 30 minutes), which presumably correspond to breaks between the answer sessions.

There may be two reasons for the long intervals between answers: it could be that it took the user a long time to provide the answer to a difficult question, or

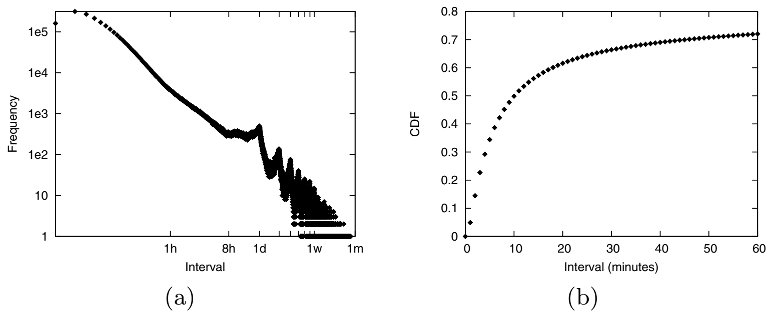


Fig. 4. The (a)Frequency and (b)Cumulative Distribution of the intervals between two successive answers for all active users

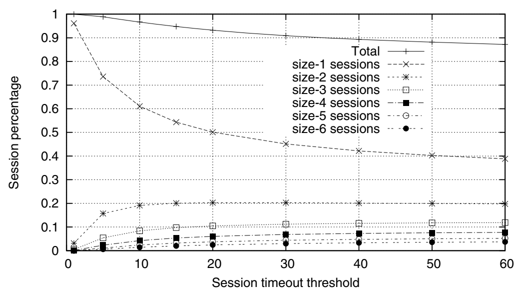


Fig. 5. The change of session percentage with different session timeout thresholds

that the user left Yahoo! Answers to do something else (a more likely scenario). Therefore, we define the continuous answer activities of a user as an *answer session* of the user. Understanding the number of questions that a user would answer continuously within a single answer session would be helpful for designing question recommender systems, e.g. how many questions to recommend to a user.

To detect answering session boundaries, we adapt some of the methods proposed to determine Web search session boundaries (e.g., [8]). In our setting, the time gap between the successive answers was chosen as the most intuitive metric. We report the distribution of intervals between two successive answers of a user, shown in Figure 4(a). As we can see, the frequency of intervals less than 8 hours long, forming a roughly power-law-like distribution. However, there are seven secondary peaks, corresponding to intervals of one to seven days. We further "zoom in" to consider the intervals of one hour, shown in Figure 4(b), which shows that for over 70% of the cases, users post the next answer within 1 hour after the current answer.

Based on this observation, we apply a timeout threshold to detect the session boundaries. If the interval of two successive answers is larger than the threshold, they belong to different sessions, and to the same session otherwise. We use the methods in [12] to determine the optimal session timeout threshold, i.e., analyzing the effect of different session timeout thresholds on the proportions

Table 2. Answering session statistics for varying timeout values

Threshold	Session size	Session size($\geq$2)	Session duration	Answer time	Gap duration
30m	2.89±3.53	4.45±4.16	26.5m±27.7m	7.68m±6.70m	19.1h±33.8h
40m	3.13±3.86	4.69±4.48	32.2m±34.7m	8.73m±8.44m	20.6h±34.8h

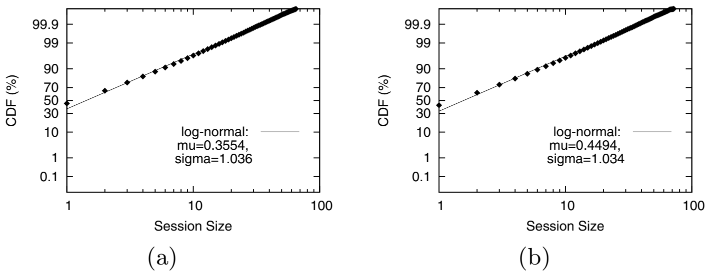


Fig. 6. The Cumulative Distribution (CDF) of session sizes based on the session timeout threshold of 30 minutes (a) and 40 minutes (b). The x axis is in log scale, and the y axis is in normal probability scale. Log-normal functions were fit to each CDF, with the parameters shown on the corresponding plot areas.

of sessions with different sizes. Figure 5 shows the results. The proportion of 1-size sessions decreases quickly with the increase of session timeout threshold until 30 minutes, while the proportion of sessions with size 3-6 increases. After that, increasing timeout threshold has negligible impact on proportions of these sessions, especially when the session timeout threshold is larger than 40 minutes. Therefore, the session timeout threshold should be between 30 and 40 minutes.

Table 2 shows the session statistics computed based on the two thresholds for session detection. As we can see, the average session size is around 3 for both session threshold values. This means that users answer 3 questions in a session on average, providing guidance for designing real-time question recommender systems, e.g. three or more questions can be recommended to a user. To explore the average time that users spend on posting an answer, we also compute the average session duration. A session duration is computed as the time between the posting of the first answer and the last answer in a session. For sessions with size $n \geq 2$ and duration d, we can then compute the average answer time t as $t = \frac{d}{n-1}$. The results are shown in Table 2. As we can see, the average answer time appears to be about 8 minutes for both session threshold values (which also includes the time needed to choose the next question to answer).

To understand better the properties of answer session size, we show the distribution of session sizes in Figure 6. Regardless of the specific session timeout value, the distribution agrees well with the log-normal distribution, which is a line on the "normal probability" (y axis) vs log (x axis) plot. The CDF of a log-normal distribution is $\Phi(\frac{ln\ x-\mu}{\sigma})$, where Φ is the CDF of the standard normal distribution. The best-fit line is specified by the equation $y = \frac{x-\mu}{\sigma}$ with the parameters reported in the plot area.

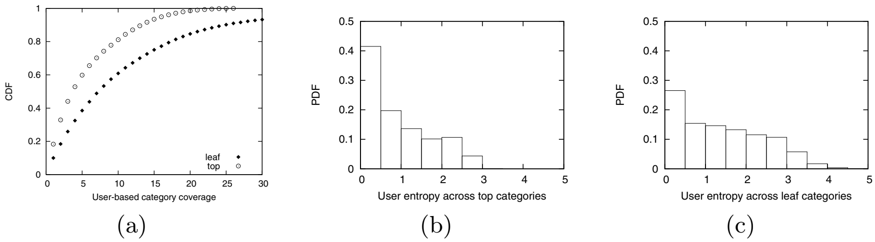


Fig. 7. The Cumulative Distribution (CDF) of user-based category coverage, which is the number of categories in which a user has posted answers across the entire dataset duration. The hollow circles represent user-based category coverage for top categories, and solid diamonds represent the leaf categories (a); The distribution of user entropy across all top (b) and leaf (c) categories: lower entropy indicates user activity focused on fewer categories.

In summary, the analysis above first focused on the answerer behavior in the aggregate (weekly and daily), and largely confirms previous findings, thus validating our data collection method. We then considered session-based behavior of *individual* answerers, and identified a novel *bursty* behavior of the answerers.

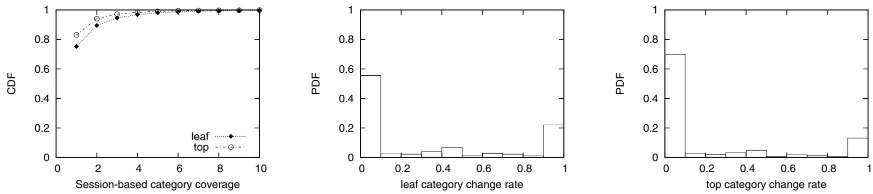
4 Understanding How Answerers Choose Questions

Having analyzed *when* users would like to answer questions, we now explore *how* they tend to choose the questions to answer. Based on the simplified answering process shown in Figure 1, we explore several factors that may affect the users' decisions of which questions to answer, including question category (topic), the question's rank in the list shown to users, question text, and the users' previous answering history profile.

4.1 Question Category Effects

Browsing a category is a first step of an answering process in YA. Users can choose any category to browse, from top categories to leaf categories. If a category is not chosen explicitly, the root category "All categories" is used by default. To explore the effect of question category on users' choices of which question to answer, we first compute the category coverage of users. The category coverage of a user is the number of different categories in which the user has posted answers. The results shown in Figure 7(a) confirm that some users answer questions in more than one leaf categories within the same top category. Moreover, we can see that more than 90% of users post answers in less than 30 leaf categories (out of 1659 leaf categories present in our dataset).

Next, we explore how focused the answers are across different categories, using the entropy measurement introduced by Adamic et al. [1]. The entropy of a user is defined as $H = -\sum_i p_i log(p_i)$, where each i means a category covered by answers of the user and p_i means the percentage of answers of the user in that



(a) (b) (c)

Fig. 8. (a)The Cumulative Distribution Function of session-based category coverage, which is the number of categories in which a user has posted answers in a single answer session. The hollow circles (solid diamonds) represent session-based category coverage for top (leaf) categories. (b)(c)The Probabilistic Distribution Function of session-based category change rate for leaf(b) and top(c) categories, which is the percentage of two successive answers in different categories posted by a user in a single answer session. Note that the session timeout threshold of 30m is used here.

category. The results are shown in Figure 7(b) and 7(c). We can see that users tend to be relatively focused to answer questions primarily on a handful of topics.

For real-time question recommender systems, it is also very important to know whether a user would like to answer questions in different categories within a single session. Therefore, we also compute the session-based category coverage of users. The session-based category coverage of a user in a session is the number of different categories in which the user has posted answers in the session. The results are reported in Figure 8(a). As we can see, for around 70% of cases, the users post questions in just one leaf category in a single session.

To explore more about how users would change categories during his single answer session, we also compute the change rate of categories, shown in Figure 8(b) and 8(c). We can see that in most cases they tend not to change throughout an answer session; however, in some cases they change at every chance. Understanding the user preference on category changes in a single session can be very helpful for improving the user experience in question recommender systems.

The above analysis shows that categories play an important role in deciding users' choices of which questions to answer, as they tend to be focused rather than diverse regarding the topics. In a single session, most users prefer to answer questions in only very few categories and to answer the next question in the same category.

4.2 Question Rank Effects

According to the basic answering process shown in Figure 1, after choosing a category, the users will see a list of questions – by default, arranged in the order of most recent arrival. Then, the user will answer one or more questions in the list. We posit that the users tend to examine the questions in order of listing and answer them in order of the examination. This examination hypothesis has extensive support from the web search result examination literature.

Therefore, we propose the following simple, yet surprisingly accurate model of answerer behavior that simply follows the order of the posted questions.

Ordered Question Examination Model (OQE): *The Answerer repeatedly examines the questions in the order presented in the Category list (normally, in reverse order of arrival, most-recent first), and answers one of the top-K questions in the list - and then goes back and repeats.*

To verify this OQE model, we need to know the questions and their order that the answerers saw before choosing a question and posting an answer. However, it is difficult to recreate the exact list of questions that the answerers saw, so we *approximate* the list based on the known characteristics of the YA site and the externally available data.

First, we represent an answer by a tuple $A(u_A, q_A, c_A, t_A)$, which means the answer A is posted by user u_A for question q_A in category c_A at time t_A. Similarly, we represent a question by a tuple $Q(u_Q, c_Q, t_Q)$, which means the question Q is posted by user u_Q in category c_Q at time t_Q. Then, for each answer $A(u_A, q_A, c_A, t_A)$ of the user u_A, we create a ranked list of questions in the category c_A that are posted before the time t_A, ordered by their recentness, most recent first. More formally, the list with respect to $A(u_A, q_A, c_A, t_A)$ can be represented as

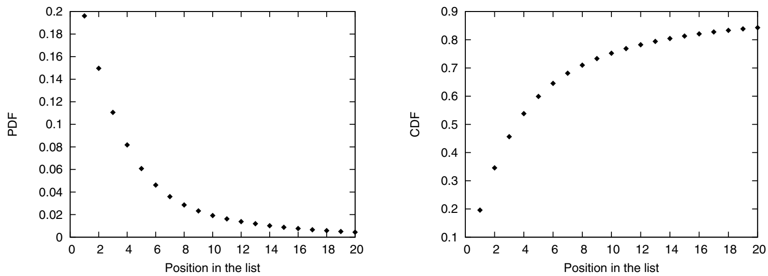
$$L_A = [Q_i(u_{Q_i}, c_{Q_i}, t_{Q_i}) \mid c_{Q_i} = c_A \ \wedge \ t_{Q_i} < t_A \ \wedge \ \forall j > i, t_{Q_j} < t_{Q_i}]$$

Note that in real scenarios, the answerers may browse any category from top to leaf. However, for simplicity, we just assume that answerers always browse to leaf categories before they answer questions. Also, we do not count in the user's time for submitting the answer A which will shift the estimated questions in L_A slightly, compared to the actual list. In addition, considering that YA shows 20 questions by default, and many answerers do not bother to click to the next page, we focus on the sublist $L_{A,20}$ containing the top 20 questions in the estimated list L_A. Then, based on the list $L_{A,20}$ for answer $A(u_A, q_A, c_A, t_A)$, we want to check whether the question q_A is in $L_{A,20}$. If yes, we are also interested in the rank of the hit, that is the i such that $Q_i = q_A$. This indicates that after the user u_A browsed to the category c_A, she chose to answer the question ranked at the ith$(1 \leq i \leq 20)$ position in the list shown to her.

Figure 9 shows the distribution of the rank positions of the chosen questions. As we can see, the higher a question is ranked, the greater probability it is answered. In addition, while only top 20 questions in the list are considered, the OQE model achieves a recall of 0.84. This means that for 84% of the cases, users just choose questions from the first page they see to answer.

4.3 Question Text Effects

Next, we try to explore beyond the question rank, to understand how the question text affects users' choices of which questions to answer. So we performed an experiment that learns to find the target question q_A in the list of $L_{A,20}$ based on question text features. We use the learning-to-rank framework for this task. Given a list of questions $L_{A,20}$ seen by a user, we derive features representing the



(a) (b)

Fig. 9. The Probability Distribution Function(a) and Cumulative Distribution Function(b) of the positions in the list seen by a user, containing a question that was selected by the user to answer

Table 3. Features (50 total) used in the experiment

Position Information (4 total): ∗ The position where the question is in the list $L_{A,20}$; ∗ The delay of the question since it was posted until seen by the user. ∗ The deviation of the above 2 feature values from the average values of the user.
Similarity (5 total): ∗ The similarities between the question and user profile against the 4 fields and the whole profile.
Visual Quality (16 total): ∗ The length of question subject/content. ∗ The punctuation, capitalization and spacing density of question subject/content. ∗ The deviation of the above 8 feature values from the average values of the user.
History (4 total): ∗ The number of prior answers for the question seen by the user. ∗ The number of prior questions asked by the question asker. ∗ The deviation of the above 2 feature values from the average values of the user.
Keywords (21 total): ∗ A vector of length 20 representing whether this question contains the 20 most frequent terms in popular questions (i.e. questions with more than 20 answers). ∗ The number of 1s in the above vector.

associated information (e.g. question text, user's answering history) to predict which question will be answered by the user.

Guided by reference [2], we design features according to five layers: position information about questions, question-user similarities, visual quality of questions, popular keywords in questions, and history information about the questions. The complete list of features is shown in Table 3.

To make our experiments feasible, we randomly selected 1000 out of the 45,543 active users to build the dataset for this experiment. For each user, the first half of her answers is used to build her user profile. Then, we use the next 1/4 of her answers as training data, and the last 1/4 of her answers as testing data for training and testing the ranker. The resulted dataset contains 15,226 answers and 304,434 questions for training, and 14,721 answers and 294,361 questions for testing.

We use Lucene[1] in our experiment to compute the similarity features between a user profile and the quetion. Each user profile is indexed as a document with

[1] `http://lucene.apache.org/`

Table 4. The performance of learning-to-rank approach for predicting the chosen question. A * indicates significance at $p<0.05$, and ** indicates significance at $p<0.01$. The p-values are computed using paired t-tests (one-tailed distribution).

	P@1	Improvement over baseline
Baseline(whether the question is at position 1)	0.2445	0%
Pos(4)+Sim(5)	0.2496	+2.1%**
Pos(4)+Vis(16)	0.2438	−0.3%
Pos(4)+His(4)	0.2427	−0.8%
Pos(4)+Key(21)	0.2461	+0.6%
Pos(4)+Sim(5)+Vis(16)	0.2498	+2.1%*
Pos(4)+Sim(5)+His(4)	0.2512	+2.7%**
Pos(4)+Sim(5)+Key(21)	0.2539	+3.8%**
Pos(4)+Sim(5)+Key(21)+Vis(16)	0.2526	+3.3%**
Pos(4)+Sim(5)+Key(21)+His(4)	0.2533	+3.6%**
Pos(4)+Sim(5)+Key(21)+His(4)+Vis(16)	0.2542	+4.0%**

four fields: the content of her answers; and the title, content, and category of the questions she answered. Then we treat a question (including the title, content and category) as 5 queries against the 4 different fields as well as the whole user profile. The 5 scores returned are used as the question-user similarity features. All the features are normalized by linear scaling to unit range. After computing all the features, we then apply a learning-to-rank algorithm, SVM^{rank}[15], to rank the questions.

Since our goal is to find the target question q_A in the list of $L_{A,20}$, we evaluate the results by P@1. The results are shown in Table 4. As we can see, the baseline by only checking whether the question is ranked at position 1 achieves the P@1 of 0.24. This means around one fourth of cases, users answer questions ranked top 1 in the list they see. Although position features dominate the performance, including the additional text features provides a slight, but statistically significant improvement of 4% ($p<0.01$) over the simple position-only OQE model. Therefore, while the question text does affects answerers for choosing questions, the effect of the question text is not as large as that of category and rank.

5 Conclusions and Future Work

This paper explored the contextual factors that influence the *answerer behavior* in a large, popular CQA system, with the goal to inform the construction of real-time, online question routing and recommendation systems. Specifically, we considered *when* users tend to answer questions in a large-scale CQA system, and *how* answerers tend to choose the questions to answer. Our analysis could help develop more realistic evaluation methods for question recommendation, and provide valuable insights into answerer behavior.

In future work, we plan to study the answerer behavior in the more proactive experiments (instead of the passive observation performed in this study), and to perform deeper investigation of the answerer behavior, e.g. by analyzing the differences in activity of the extremely active "top contributor" users.

References

1. Adamic, L.A., Zhang, J., Bakshy, E., Ackerman, M.S.: Knowledge sharing and yahoo answers: everyone knows something. In: WWW, pp. 665–674 (2008)
2. Agichtein, E., Castillo, C., Donato, D., Gionis, A., Mishne, G.: Finding high-quality content in social media. In: WSDM, pp. 183–194 (2008)
3. Aperjis, C., Huberman, B.A., Wu, F.: Harvesting collective intelligence: Temporal behavior in yahoo answers (January 2010), `http://arxiv.org/abs/1001.2320`
4. Bian, J., Liu, Y., Agichtein, E., Zha, H.: Finding the right facts in the crowd: factoid question answering over social media. In: WWW, pp. 467–476 (2008)
5. Bouguessa, M., Dumoulin, B., Wang, S.: Identifying authoritative actors in question-answering forums: the case of yahoo! answers. In: KDD (2008)
6. Cao, X., Cong, G., Cui, B., Jensen, C.S., Zhang, C.: The use of categorization information in language models for question retrieval. In: CIKM, pp. 265–274 (2009)
7. Cao, Y., Duan, H., Lin, C.Y., Yu, Y., Hon, H.W.: Recommending questions using the mdl-based tree cut model. In: WWW, pp. 81–90 (2008)
8. Gayo-Avello, D.: A survey on session detection methods in query logs and a proposal for future evaluation. Inf. Sci. 179(12), 1822–1843 (2009)
9. Guo, J., Xu, S., Bao, S., Yu, Y.: Tapping on the potential of q&a community by recommending answer providers. In: CIKM, pp. 921–930 (2008)
10. Guo, L., Tan, E., Chen, S., Zhang, X., Zhao, Y.E.: Analyzing patterns of user content generation in online social networks. In: KDD, pp. 369–378 (2009)
11. Gyöngyi, Z., Koutrika, G., Pedersen, J., Garcia-Molina, H.: Questioning yahoo! answers. In: Proc. of the 1st Workshop on Question Answering on the Web (2008)
12. He, D., Göker, A.: Detecting session boundaries from web user logs. In: Proc. of the BCS-IRSG 22nd Annual Colloquium on Information Retrieval Research (2000)
13. Horowitz, D., Kamvar, S.D.: The anatomy of a large-scale social search engine. In: WWW, pp. 431–440 (2010)
14. Jeon, J., Croft, W.B., Lee, J.H.: Finding similar questions in large question and answer archives. In: CIKM, pp. 84–90 (2005)
15. Joachims, T.: Training linear svms in linear time. In: KDD, pp. 217–226 (2006)
16. Jurczyk, P., Agichtein, E.: Discovering authorities in question answer communities by using link analysis. In: CIKM, pp. 919–922 (2007)
17. Liu, M., Liu, Y., Yang, Q.: Predicting best answerers for new questions in community question answering. In: Chen, L., Tang, C., Yang, J., Gao, Y. (eds.) WAIM 2010. LNCS, vol. 6184, pp. 127–138. Springer, Heidelberg (2010)
18. Liu, Q., Liu, Y., Agichtein, E.: Exploring web browsing context for collaborative question answering. In: IIiX, pp. 305–310 (2010)
19. Liu, X., Croft, W.B., Koll, M.: Finding experts in community-based question-answering services. In: CIKM, pp. 315–316 (2005)
20. Nam, K.K., Ackerman, M.S., Adamic, L.A.: Questions in, knowledge in?: a study of naver's question answering community. In: CHI, pp. 779–788 (2009)
21. Qu, M., Qiu, G., He, X., Zhang, C., Wu, H., Bu, J., Chen, C.: Probabilistic question recommendation for question answering communities. In: WWW (2009)
22. Suryanto, M.A., Lim, E.P., Sun, A., Chiang, R.H.L.: Quality-aware collaborative question answering: methods and evaluation. In: WSDM, pp. 142–151 (2009)
23. Zhang, J., Ackerman, M.S., Adamic, L.: Expertise networks in online communities: structure and algorithms. In: WWW, pp. 221–230 (2007)
24. Zhou, Y., Cong, G., Cui, B., Jensen, C.S., Yao, J.: Routing questions to the right users in online communities. In: ICDE, pp. 700–711 (2009)