On the Evolution of the Yahoo! Answers QA Community

Yandong Liu   Eugene Agichtein

Emory University
{yliu49,eugene}@mathcs.emory.edu

ABSTRACT

While question answering communities have been gaining popularity for several years, we wonder if the increased popularity actually improves or degrades the user experience. In addition, automatic QA systems, which utilize different sources such as search engines and social media, are emerging rapidly. QA communities have already created abundant resources of millions of questions and hundreds of millions of answers. The question whether they will continue to serve as an effective source of information for web search and question answering is of vital importance. In this poster, we investigate the temporal evolution of a popular QA community –Yahoo! Answers, with respect to its effectiveness in answering three basic types of questions: factoid, opinion and complex questions. Our experiments show that Yahoo! Answers keeps growing rapidly, while its overall quality as an information source for factoid question-answering degrades. However, instead of answering factoid questions, it might be more effective to answer opinion and complex questions.

Categories and Subject Descriptors

H.3.4 Systems and Software: *Question-answering (fact retrieval) systems*; H.3.3 Information Search and Retrieval: *Search process, Selection process.* H.3.5 Online Information Services: *Web-based services*

General Terms

Measurement, Experimentation, Human Factors

Keywords

Social media, Question answering

1. OVERVIEW

Recently, *community question answering* (CQA) emerged as a popular alternative to finding information online. It has attracted millions of users who post millions of questions and hundreds of millions of answers, producing a huge knowledge repository of all kinds of topics, so many potential applications can be possibly made on top of it. For example, automatic question answering systems, which try to find the information to questions directly instead of giving a list of related documents, might use CQA repositories as a useful information source. However, it is not clear what information needs these CQA portals serve, and how these communities are evolving. In this poster we study a popular CQA portal –Yahoo! Answers– as our test case to quantify relative effectiveness for different types of queries such as factoid question, opinion queries, and complex question. As another motivation for our exploration, we examine how Yahoo! Answers has evolved recently and try to identify some trends in its development.

We analyze the temporal changes of the Yahoo! Answers contents during the years of 2006 and 2007, so that we may get an idea of how it has progressed and to quantify potential uses of the

services. Also we can gain a deeper understanding of community question answering phenomenon as it evolves over time. For this, we initially study CQA performance for simple factoid questions, and evaluate the potential utility for using CQA for opinion and complex questions.

2. EXPERIMENTAL SETUP

2.1 Datasets

Yahoo! Answers: To obtain enough information for a better analysis, we crawled a large portion of Yahoo! Answer, resulting in a local archive of 96,000 questions and 1,150,000 answers, which cover about 125 categories. We divide all the questions and answers equally by posting time, resulting in 4 time periods of 2006.8 ~ 2006.11, 2006.12 ~ 2007.3, 2007.4 ~ 2007.7 and 2007.8 ~ 2007.11.

TREC questions: In order to examine what type of questions Yahoo! Answers is best for, we prepared three datasets: **factoid** questions, **opinion** questions and **complex** questions, all acquired from the NIST TREC website.

- For **factoid** questions, we selected a sample of 1250 questions from the TREC question-answering track (years 1999-2006) that have at least one matching question was found in Yahoo! Answers database. This way we could focus on the quality and relevance of the CQA content instead of the artifacts of the current version of the Yahoo! Answers question retrieval engine. Note that from 1999 to 2003, the TREC factoid questions were independent of each other; however, starting in 2004 the TREC questions were organized in groups with a common target for each group. For the latter, we submit its question text as well as its target.

- **Opinion** questions consist of the 100 query topics from TREC 2006 and 2007 Blog tracks. The original questions were in the form of "find opinion of/about/on xxx". In this case, we removed their prefix and only preserve their content words for later experiments.

- Finally, the **complex** questions were composed from 2006 and 2007 TREC CIQA track, producing a set of 61 questions.

2.2 Methodology

In order to demonstrate Yahoo! Answers capability to provide related knowledge for factoid questions, we submit the factoid questions to Yahoo! Answer's web service and check the answers are correct or not.

Submitting factoid questions and retrieving answers: We submit all the factoid questions to Yahoo! Answers with minimal rewording.

Relevance for factoid questions: TREC makes available "unofficial" answer patterns for factoid questions. We check every answer's content, and if the answer pattern is included by regular expression, we say this answer is 'matched', and is considered relevant for our purposes. For these preliminary experiments we focus only on whether the correct information is contained

somewhere within a page and not on stylistic or ease-of-use issues (e.g., long vs. short answer), which we will explore in the future.

2.3 Metrics

MRR: We use three variants of the standard MRR metric (Mean Reciprocal Rank) to examine the effectiveness of Yahoo! Answers for answering questions:

- Max: the maximum MRR for each of the retrieved question threads. This models the "intelligent" searcher that clicks on the most relevant question thread.
- Strict: Computed by evaluating the answers for each question thread in decreasing order (first, all answers for the top ranked question, then all answers for the next question, etc). This models a "naive" searcher that examines results in order.
- RR: Round-robin evaluation of answers from each retrieved question threads in turn. This models a common meta-search result merging strategy.

KL-Divergence: We also want to measure the similarity of the CQA content to known classified queries (e.g., factoid questions) to better characterize the content and its evolution. Specifically, we use KL-Divergence, computed as:

$$D_{\mathrm{KL}}(P\|Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

The lower the KL-Divergence value, the more similar are two distributions P and Q. In our setting we compare the language models for, say, TREC factoid questions (P), and Yahoo! Answers questions (Q). We use a simple word unigram language model with estimated likelihood smoothing ($\lambda = 0.2$).

3. RESULTS

We report the MRR, and the KL-divergence for factoid questions in Figure 1 (a), for varying time periods. **Retrieved Yahoo! questions** (*retrieved Y! Q* in the figure) are the questions returned from Yahoo! Answers that were triggered by TREC queries (keywords) while **Matched Yahoo questions** (*matched Y! Q* in the figure) refer to those that have at least one *correct* answer that matched against TREC patterns. The KL-Divergence values were calculated the between the sets *retrieved Y! Q* and *matched Y! Q,* and the language model build over all TREC Factoid questions. We also report the MRR values for the same time periods.

As we can see, for the first three time periods, KL-Divergence initially increases, but drops during the final period. The MRR values appear to slightly decrease. This suggests that the questions posted on Yahoo! Answers increasingly diverge from the typical TREC factoid questions, and the CQA content is becoming less valuable as a source of factoid information.

Next we compare whether Yahoo! Answers content is evolving to be more similar to TREC factoid questions, opinion questions, or complex questions. These were obtained by calculating the KL-Divergence between the TREC factoid, opinion, and complex questions respectively, and the Yahoo! Answers snapshots for the different time periods. Figure 1 (b) shows that the Yahoo! Answers corpus of questions is far more similar to the TREC opinion and complex questions, than to TREC factoid questions

Last, we investigate the site-level evolution of the Yahoo! Answers community (Figure 1 (c)). The three bars indicate the numbers of newly posted questions, answers, and the number of votes per each time period. For number of votes, we use the sum of both positive votes and negative votes. The number of newly produced questions and answers during each time period remained steady, though their speed of growth decreases. It is worth noting that the number of votes increases faster than the content. This implies that instead of contributing more answers, Yahoo! Answers participants are becoming more interested rating the existing answers.

4. CONCLUSIONS

Yahoo! Answers is rapidly growing. Our experiments suggest that searchers increasingly turn to Yahoo! Answers for opinion and complex questions. We can also see as the size grows, the overall quality in terms of answering factoid questions degrades. Finally, it appears that the form of participation in the Yahoo! Answers community is becoming more passive: users increasingly vote on other users' content and provide fewer answers. Finally, our findings suggest that while Yahoo! Answers is not optimal for factoid QA, it is becoming the destination of choice for complex information needs such as opinion or advice.

5. REFERENCES

[1] E. Agichtein, C. Castillo, D. Donato, A. Gionis, and G. Mishne, Finding High Quality Content in Social Media, with an Application to Community-Based Question Answering, in Proc. of WSDM 2008

[2] J. Lin and B. Katz, Question Answering from the Web Using Knowledge Annotation and Knowledge Mining Techniques, in Proc. of CIKM 2003

[3] A. Sugiura and O. Etzioni, Query routing for Web search engines: architecture and experiments, Computer Networks, Volume 33, Issues 1-6, 2000

[4] E. M. Voorhees, Overview of the TREC 2003 Question Answering Track, in Proc. of TREEC 2003
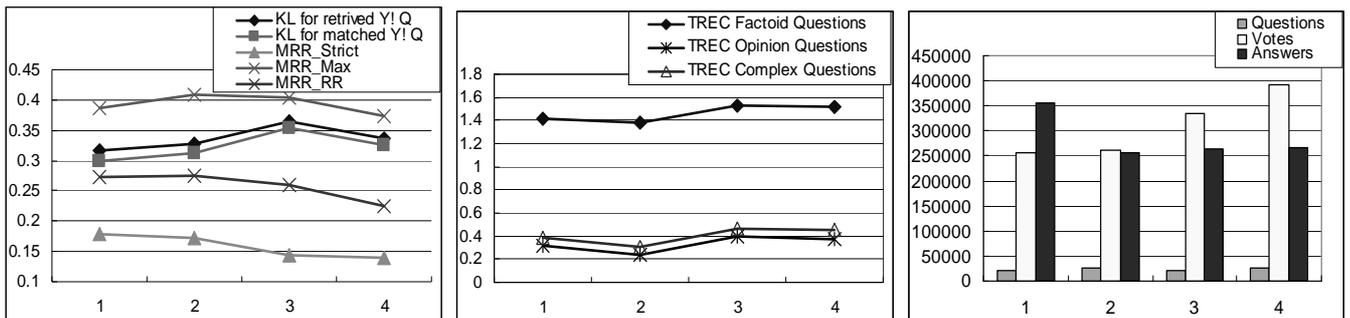
Figure 1: Statistics for varying time periods. (a) MRR and KL-Divergence of CQA vs. TREC Factoid questions. (b) KL-Divergence of CQA vs. TREC factoid, complex, and opinion topics; (c) Number of CQA questions, answers, and votes for each time period.