

A Hybrid Near Real-time Question Answering System with Crowdsourcing

Denis Savenkov

Emory University
dsavenk@emory.edu

Eugene Agichtein

Emory University
eugene@mathcs.emory.edu

Abstract

Modern search engines have made dramatic progress in the answering of many user's questions about facts, such as those that might be retrieved or directly inferred from a knowledge base. However, many other questions that real users ask are more complex, such as asking for opinions or advice for a particular situation, and are still largely beyond the competence of the computer systems. As conversational agents become more popular, QA systems are increasingly expected to handle such complex questions, and to do so in (nearly) real-time, as the searcher is unlikely to wait longer than a minute or two for an answer. One way to overcome some of the challenges in complex question answering is crowdsourcing. We explore two ways crowdsourcing can assist a question answering system that operates in (near) real time: by providing answer validation, which could be used to filter or re-rank the candidate answers, and by creating the answer candidates directly. In this paper we describe a hybrid near real-time question answering system for complex informational tasks, that incorporates a crowdsourcing module. Our experiments show that even within a one minute time limit, crowd workers can produce reliable answer ratings and generate new candidate answers, which significantly improves the performance of the QA system. Our findings can be useful for developing hybrid human-computer systems for automatic question answering and conversational agents.

Congratulations on having a paper selected for inclusion in an AAAI Press proceedings or technical report! This document details the requirements necessary to get your accepted paper published using L^AT_EX. If you are using Microsoft Word, instructions are provided in a different document. If you want to use some other formatting software, you must obtain permission from AAAI Press first.

The instructions herein are provided as a general guide for experienced L^AT_EX users who would like to use that software to format their paper for an AAAI Press publication or report. If you are not an experienced L^AT_EX user, do not use it to format your paper. AAAI cannot provide you with support and the accompanying style files are **not** guaranteed to work. If the results you obtain are not in accordance with the specifications you received, you must correct your source file to achieve the correct result.

These instructions are generic. Consequently, they do not include specific dates, page charges, and so forth. Please consult your specific written conference instructions for details regarding your submission. Please review the entire document for specific instructions that might apply to your particular situation. All authors must comply with the following: (Alonso, Rose, and Stewart 2008).

- You must use the latest AAAI Press L^AT_EX macro.
- Download the author kit.
- Complete, sign, and return by the deadline the AAAI copyright form (proceedings authors) or distribution license (technical report authors).
- Read and format your paper source and PDF according to the formatting instructions for authors.
- Submit your electronic files and abstract using our electronic submission form **on time**.
- Submit your copyright form, and any required page or formatting charges to AAAI Press so that they are received by the deadline.
- Check every page of your paper before submitting it.

Introduction

This is our introduction. Currently copied from NAACL.

It has long been a dream to communicate with a computer as one might with another human being using natural language speech and text. Nowadays, we are coming closer to this dream, as natural language interfaces become increasingly popular. Our phones are already reasonably good at recognizing speech, and personal assistants, such as Apple Siri, Google Now, Microsoft Cortana, Amazon Alexa, etc., help us with everyday tasks and answer some of our questions. Chat bots are arguably considered “the next big thing”, and a number of startups developing this kind of technology has emerged in Silicon Valley and around the world¹.

Question answering is one of the major components of such personal assistants. Existing techniques already allow users to get direct answers to their factoid questions. However, there is still a large number of more complex questions, such as advice or accepted general opinions, for which users

¹<http://time.com/4194063/chatbots-facebook-messenger-kik-wechat/>

have to dig into the “10 blue links” and extract or synthesize answers from information buried within the retrieved documents. To cater to these informational needs, community question answering (CQA) sites emerged, such as Yahoo! Answers and Stack Exchange. These sites provide a popular way to connect information seekers with answerers. Unfortunately, it can take minutes or hours, and sometimes days, for the community to respond, and some questions are left unanswered altogether.

To facilitate research on this challenge, TREC LiveQA shared task² was started in 2015, where automated systems attempt to answer real users’ questions within a 1 minute period. This task was successful, with the winning system able to automatically return a reasonable answer to more than half of the submitted questions, as assessed for TREC by the trained judges from NIST. Nevertheless, many questions were unable to be answered well by any of the participating systems.

In this work we explore two ways common crowdsourcing can be used to help an automated system answer complex user questions in near real-time scenario, e.g., within a minute. More specifically, we study if crowd workers can quickly and reliably judge the quality of the proposed answer candidates, and if it is possible to obtain reasonable written answers from the crowd within a limited amount of time. Our research questions can be stated as:

1. RQ1. Can crowdsourcing be used to judge the quality of answers to non-factoid questions under a time limit?
2. RQ2. Is it possible to use crowdsourcing to collect answers to real user questions under a time limit?
3. RQ3. How does the quality of crowdsourced answers to non-factoid questions compare to original CQA answers, and to automatic answers from TREC LiveQA systems?

System design

Here we will describe the system design.

Experiments

Here we describe our results.

Analysis and Discussion

Here we analyze the results.

Conclusion

With this we conclude the paper.

Thank you for reading these instructions carefully. We look forward to receiving your electronic files!

References

Alonso, O.; Rose, D. E.; and Stewart, B. 2008. Crowdsourcing for relevance evaluation. *SIGIR Forum* 42(2):9–15.

²www.trec-liveqa.org