Answering Queries using Humans, Algorithms and Databases

Aditya Parameswaran  Neoklis Polyzotis
Stanford University  University of California
at Santa Cruz

1. Summary

The objective of this paper is to identify the research opportunity presented by crowdsourcing (CrSS) to answer queries. The use of databases is emphasized as a key component of the CrSS approach. They set several examples in which questions that are very easy to answer for humans are very complicated to be answered by a computer. However, several complications arise when doing crowdsourcing, like obtaining several answers when only one is needed, human mistakes, subjectivity or maliciousness. Other difficulties include the need of setting a proper query for humans and analyzing the information obtained. The authors also propose some concrete research directions.

2. Positive/Strong Points

2.1) Crowdsourcing presents a lot of opportunities but there are still many challenges to overcome.

2.2) Different options are presented for when dealing with several responses, including a probability function to determine what option to use. However, they don’t establish a concrete way of determining the probability parameter $\tau$.

3. Negative/Weak Points

3.1) CrSS can be very costly in terms of time and money. The lower time response is required, the more money it would cost. It is also necessary to have several answers instead of just trusting one or a few answers.

3.2) The authors realize that there are problems with CrSS, including human mistakes, answering queries maliciously wrong. They propose a couple of solutions, but none of them is entirely satisfactory.

3.3) The paper is basically a presentation of what CrSS is, and it also presents some possible research areas on CrSS, but there is no real research or new method introduced. It could be considered as a compilation of other works, but no results are shown or compared.

4. Research Questions and Points for Discussion

4.1) How likely would it be possible to create a software to automate the question design and analysis of responses that could be used to alleviate the difficulties presented in CrSS?

4.2) Setting the value of $\tau$ for the Bernoulli distribution for deciding what response to take might be a difficult task. However, CrSS could also be used for determining this parameter in a meta-CrSS algorithm. Also, probably a multinomial distribution would
be more appropriate in the case when many different responses are presented, instead of having just a parameter $\tau$ and disregarding those answers with probability less than $\tau$.

4.3) Combining the two limited solutions to Spam would create a better solution. In this way, using CAPTAs and then using majority voting would have a more accurate response than using each one of them individually.