Towards Building an Automated Fact-Checking System

Jimenez Damian
The University of Texas at Arlington
701 South Nedderman Drive
Arlington, Texas
damian.jimenez@uta.edu

ABSTRACT

Fact-checking has become a hot-topic issue in today's society, yet a system that can fact-check automatically has yet to surface. In this paper we explore the hurdles of implementing such a system and assess the limitations of currently available technologies. The results of this research show that there is still much work to be done in parsing natural language inputs and mapping them to queries that can be used to extract information from knowledge bases. In particular we explore the use of a question generation tool as a means of mapping claims to natural language queries. From this we observe that, while moderately successful, current tools still require substantial improvements to be useful as part of an end-to-end fact-checking system. However, the scaffolding for tackling this problem exists, and ultimately we worked to understand where the shortcomings lie and their nature.

1. PROBLEM AND MOTIVATION

Modern technology has allowed us to have information accessible to us with just a few taps on a phone screen. This accessibility has not come without pitfalls, for there are many places that serve misinformation knowingly and unknowingly. Due to these developments, there have been increased efforts to automate fact-checking, in order to facilitate the digestion of information for an individual without expert domain knowledge. This research stems from those needs.

2. BACKGROUND AND RELATED WORK

While, currently, no such fact-checking system exists, there have been substantial strides made in the effort to identify factual claims within the political domain. One of the more successful efforts comes from ClaimBuster of the University of Texas at Arlington[1]. The ClaimBuster platform has gained traction in the media and has undertaken the challenge of classifying political claims [1]. A large part of this project was motivated by ClaimBuster, and its claim picking process was also used when processing some of the PolitiFact data. The obvious next step, beyond what ClaimBuster was already doing, was to see how feasible an automated system would be to create, from scratch or just using currently existing technologies. My efforts focused on the latter as we wanted to see the limitations of current technologies so that we could learn where our efforts would best make an impact. One integral part of this whole system involved querying knowledge databaes using natural language. In order to do this, we needed a tool that converts sentences into relevant questions. Currently, the most viable, publicly available tool is from Michael Heilman. That program was developed as part of his Ph.D. dissertation, and showed moderate to fairly good success in mapping sentences to relevant questions [2]. Ultimately, our goal was to provide a system that any user could use without any technical expertise in formulating database queries.

3. APPROACH AND UNIQUENESS

The end-to-end system that I constructed consisted of three main components: the question generation system, the natural language database which would be queried using the automatically generated questions, and a method for determining if the claim was true or false (see Figure 1). For the question generation component, as previously mentioned, a publicly available question generator was used [2]. In the end, we decided to use Wolfram as the database to query using natural language. Wolfram was easily accessible via an API which us to quickly develop an end-to-end system to test the limitations of these tools. This particular selection of technologies and approach was taken so as to simplify the inherent difficulties when trying to build such a system. By leveraging the power of knowledge bases that can be queried using natural language, we were able to use the question generation tool to assess the viability of a system that any user without knowledge of how to query a database could use to fact-check claims.

To test the system I used claims extracted from Twitter, PolitiFact[2], Elias Says[3], and Wikipedia. I also used hand-constructed questions from a Wikipedia question-answer dataset provided by Carnegie Mellon University for use in research[4], as well as trivia questions that I found on the web[5]. These

[1]http://idir-server2.uta.edu/claimbuster/

[2]http://www.politifact.com/

[3]http://www.espn.com/blog/elias-says

[4]http://www.cs.cmu.edu/~ark/QA-data/

[5]https://web.archive.org/web/20150418134735/http://www.irc-wiki.org/Category:Trivia_questions

SIGMOD'17 Student Research Competition May 14-19 2017, Chicago, IL, USA

ACM ISBN 978-1-4503-4199-8/17/05.

DOI: http://dx.doi.org/10.1145/3055167.3055176

last two data-sets were used to assess the performance of the question generation system and to an extent Wolfram's breadth of knowledge (i.e., since these questions tended to be easier). I then tried to assess the veracity of the input claim by comparing the answer that was fed with the claim and the answer that was returned from Wolfram.

4. RESULTS AND CONTRIBUTIONS

The results showed that for datasets where the claims were well formulated, in terms of grammar and syntax, the question generation system performed relatively well. The questions come out ranked. However, from the results, it was evident that the top-ranked questions were not always the ones that Wolfram was returning the most results for. Most of the results returned were indeed in the top-k questions, with k varying from data-set to data-set. For a system like this to reliably return the user with a single concrete answer, the question generation process needs to return the best question as the top-ranked one a very large percentage of the time. This effect can be seen in Figure 2, where we see that the most results were returned for questions that ranked from 3-5. Comparatively, for the hand constructed questions no such pattern was present (see Figure 3), denoting that the ranking system was working to an extent just not as well as we would have liked.

From these results, we assessed the limitations of the current system and technologies. First, the question generation process needs to be improved or a reliable way to map natural language queries to database queries needs to be developed. There has been some research in this area, but there is still room for improvement [4]. It was also apparent that Wolfram, in many cases, lacked the knowledge to answer well-formulated questions. This could be alleviated by using a large network of knowledge bases. This, of course, would rely on successfully querying these databases which goes back to the first issue that I mentioned. Again some work has already been done in this area, but by no means is it complete or ready for deployment in a real-world system [3].

Ultimately, my results have shown that there is still much work to be done in question generation and claim to query mapping in general. Current technologies are not sufficient in constructing such a lofty system. I have also gained insight into some of the areas that need work the most. Going forward we plan to assess the viability of other techniques, such as neural networks, in claim to query mapping. The system discussed in the paper, as well as a single query version of the system, can be found at https://qgqa-cached.herokuapp.com/ and http://idir-server2.uta.edu/qgqa, respectively. Recently, we have also released a more complete end-to-end fact-checking system which is another step forwards in the goal of automated fact-checking and builds in some areas on the work discussed in this paper, that system can be accessed at http://idir-server2.uta.edu/factchecker/.

5. ACKNOWLEDGMENTS

The author would like to thank Chengkai Li, Gensheng Zhang, and Naeemul Hassan of the IDIR Lab at UT Arlington for their continued guidance and support. Their input and perspective have been paramount in the research process.

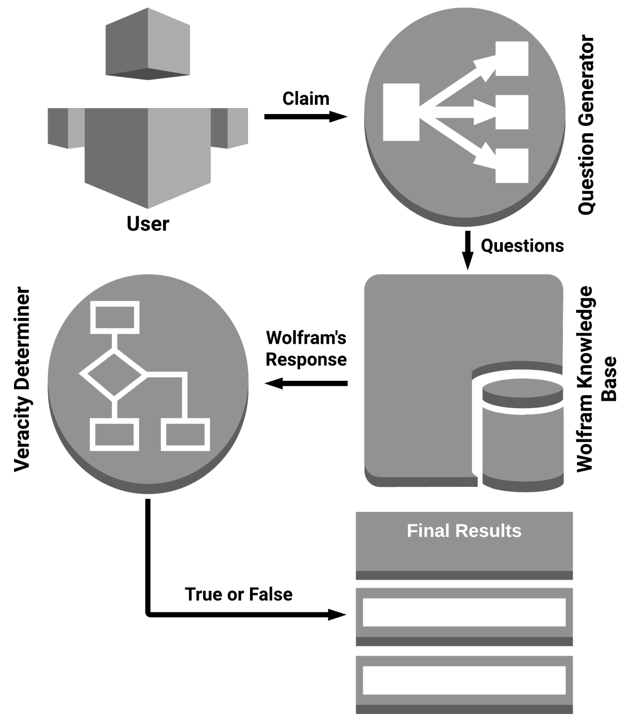

Figure 1: Figure showing how a claim moves through the system, to be classified as true or false depending on the answer received from Wolfram Alpha.

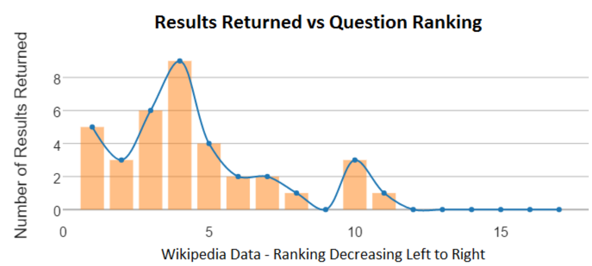

Figure 2: Chart depicting how many results were returned for a given question, all of which were ranked and generated by a question generation tool.

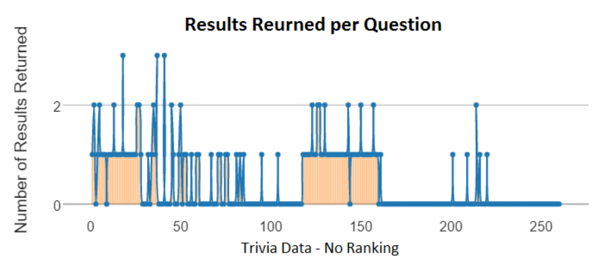

Figure 3: Chart depicting how many results were returned for a given question, none of which were automatically generated.

6. REFERENCES

[1] N. Hassan, B. Adair, J. T. Hamilton, C. Li, M. Tremayne, J. Yang, and C. Yu. The quest to automate fact-checking. In *Computational Journalism Symposium 2015*, pages 1–5, 2015.

[2] M. Heilman. *Automatic Factual Question Generation from Text*. PhD thesis, Carnegie Mellon University, 2011.

[3] D. Savenkov and E. Agichtein. When a Knowledge Base Is Not Enough: Question Answering over Knowledge Bases with External Text Data. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '16, pages 235–244, New York, NY, USA, 2016. ACM.

[4] K. Xu, S. Reddy, Y. Feng, S. Huang, and D. Zhao. Question Answering on Freebase via Relation Extraction and Textual Evidence. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, pages 2326–2336, 2016.