Progress Toward "the Holy Grail": The Continued Quest to Automate Fact-Checking

Bill Adair
Duke University

Chengkai Li
University of Texas at Arlington

Jun Yang
Duke University

Cong Yu
Google Research

1 Introduction

Two years ago at this conference we issued what we said was a "call to arms" to advance automated fact-checking [1]. We said the "Holy Grail" was "a completely automated fact-checking platform that can detect a claim as it appears in real time, and instantly provide the voter with a rating about its accuracy." We acknowledged that goal "may remain far beyond our reach for many, many years to come," but we called on the journalism and computer science communities to redouble their efforts to make progress.

Since then there has been remarkable progress and the "Holy Grail" is no longer a distant dream. Although computer scientists and journalists still have significant hurdles to overcome, recent advances with the creation of a global database of structured fact-checks and fact-checking tools such as ClaimBuster[1] and iCheck[2] have laid a groundwork for additional advances in the next few years.

As we noted in our 2015 paper, fact-checking is a growing form of journalism. According to the Duke Reporters' Lab,[3] the number of fact-checkers has nearly doubled from 64 in 2015 to 126 today. Fact-checking is uniquely suited for automated journalism because the individual articles have value weeks after they have been published because of the tendency of government officials to repeat political claims. During live events such as debates and speeches, fact-checking organizations have typically relied on editors and reporters to manually match new statements with previously published fact-checks. But in a fully automated system, the statements could be detected and, if they had been previously fact-checked, a link or summary of the conclusion could pop up in real time for the reader.

A promising development in this effort is the creation of a schema to identify fact-checking articles. This project, which is led by members of our team from Google and Duke University, has created a global open standard known as ClaimReview so that organizations can identify the people and statements they are checking, as well as their conclusion about the accuracy of claims. Google and Bing are now using the schema for search results.

Publishers can generate the schema from their content management systems or use the "Share the Facts" widget developed by the Duke Reporters' Lab. The schema provides a consistent way for search engines to identify and index fact-checking articles. It also solves a problem that has bedeviled anyone who has tried to develop apps: the mishmash of ways that different publishers present their fact-checks. The database of fact-checks identified by the schema creates tremendous potential for automation projects because it could potentially include every fact-check article published around the world.

Another promising area for automation is to assist journalists with repetitive and time-consuming tasks such as identifying factual claims. Every day fact-checkers and their college interns have difficulty keeping up with the flood of new factual claims from legislative debates, TV talk shows and other news coverage. ClaimBuster, a tool developed by our team at the University of Texas at Arlington, addresses this need by automating the process of finding factual claims to check. ClaimBuster can do the work of many college interns by quickly analyzing voluminous transcripts and identifying claims that journalists are most interested in checking.

In the past two years, we have refined ClaimBuster and have begun to deploy it for daily use by journalists. In Australia, it is used for daily analysis of Hansard, the proceedings of the Australian parliament.[4] In the United States, we are using ClaimBuster to analyze the transcript of a cable news channel and identify the most "check-worthy" claims.

Once a check-worthy claim is identified, we look for ways to help journalists check it. There are many possibilities for automation, and one focus of our team at Duke University, Google, and the University of Texas at Arlington is checking claims based on data or statistics. These claims are often vague and may be factually correct, but they can still mislead by "cherry-picking" partial and biased vantage points of the data. We have developed a tool for "perturbation analysis," which puts the claim into a larger context by automatically exploring a large number of alternative vantage points of the data, in order to evaluate claim qualities such as fairness, robustness, and uniqueness in a principled manner.

As a proof of concept, we have developed a website called iCheck and released it to the public in September 2016. The website analyzes the voting records of the U.S. Congress from January 2009 to September 2016, and lets visitors compare how legislators vote with party majorities and the president, and more importantly, explore how the comparison stacks up under

[1] http://idir.uta.edu/claimbuster/
[2] http://icheckuclaim.org/
[3] https://reporterslab.org/fact-checking/
[4] http://idir.uta.edu/claimbuster/hansard

different contexts—over time, among groups of peers, and for "key votes" identified by lobbying/political organizations.

While the "Holy Grail" of fully automated fact-checking still poses significant challenges—some requiring more research and investments over a long term—we believe that some aspects of automated fact-checking are ready for prime time and can deliver substantial benefit to the journalists and the public. In the rest of the paper, we describe our progress, discuss lessons learned, and outline our vision of next steps.

2 ClaimBuster

Since December 2014, the team at the University of Texas at Arlington has been building ClaimBuster [2-5], a claim-spotting tool for assisting fact-checkers in discovering factual claims that are worth checking. ClaimBuster monitors the plethora of places where politicians and others make political claims such as interviews, speeches and debates. It gives each sentence a score that indicates how likely it is the sentence contains an important factual claim that should be checked. In this way, ClaimBuster provides a priority ranking on the sentences. The ranking helps fact-checkers avoid having to read massive transcripts and efficiently focus on the top-ranked claims.

ClaimBuster's claim spotter was tested in real-time during the live coverage of all primary election and general election debates for the 2016 U.S. presidential election. Closed captions of the debates on live TV broadcasts, captured by a decoding device, were fed to ClaimBuster, which immediately scored each sentence spoken by the candidates and posted top-scored claims to the project's website and Twitter account (@ClaimBusterTM). Post-hoc analysis of the claims checked by professional fact-checkers at CNN, PolitiFact.com and FactCheck.org reveals a highly positive correlation between ClaimBuster and journalism organizations in deciding which claims to check. ClaimBuster has also been continuously monitoring Twitter and retweeting the check-worthy factual claims it finds in people's tweets (see @ClaimBusterTM).

Our experience so far suggests a few directions for improving ClaimBuster's accuracy in spotting important factual claims. Currently, the tool scores individual sentences. This is a clear limitation as factual claims may span multiple sentences. Mitigating this limitation entails several natural language processing tasks, including coreference resolution and topic segmentation. Furthermore, structured representation of factual claims is imperative for deep understanding of the claims and thus more accurate spotting of important claims. Such structured representation should capture various aspects of a factual claim, including the domain and topic of the claim, the template of the fact being expressed, the involved entities, and their relationships. It is also crucial to capture the claim's important elements such as numbers, time points and intervals, comparisons, grouping, and aggregates.

ClaimBuster delivers the scores on claims through a variety of channels, including its website, Twitter account, API, and Slackbot. Particularly, the Slackbot allows users to supply their own text, directly as Slack input or through text files in a Dropbox folder, and to receive the claim spotter scores for the sentences in that piece of text. The Slackbot has been published in the public Slack App directory and can be installed from https://claimbotapi.herokuapp.com/. Furthermore, a public ClaimBuster API[5] enables developers to create their own fact-checking applications using ClaimBuster as an underlying service.

As part of the team's next step toward the "Holy Grail", we are extending ClaimBuster into an end-to-end fact-checking assistant for professional fact-checkers. A preliminary version of this extension already produces true-or-false verdicts for certain types of factual claims. Given a factual claim which is scored highly by the aforementioned claim spotting component, ClaimBuster may reach a verdict by a few methods. Particularly, one of the methods is to translate the factual claim into questions and their accompanying answers. It then sends the questions to question-answering systems and compares the returned results with the aforementioned answers. It produces a verdict based on the presence/absence of a discrepancy between these two sets of answers.

3 iCheck

We demonstrated iCheck at the 2016 Computation+Journalism Symposium; please see our paper [6] and website for additional details. Here, we focus on summarizing the challenges we identified during this project.

As explained earlier, we target number-based claims derived from data or statistics, and we have identified perturbation analysis as a way to formulate the human fact-checking process as a computational problem. Automated perturbation analysis [7] can quickly examine a huge number of different vantage points of data, quantitatively assess various aspects of claim quality, and intelligently suggest "counterarguments," thereby relieving human fact-checkers from the tedious, time-consuming, and error-prone aspects of manual fact-checking. While perturbation analysis has proven to be a remarkable fit for number-based claims, it is by no means a one-size-fits-all solution for all types of claims. For example, checking an assertion that some event occurred or somebody took a particular position on an issue would require different procedures that need to be automated differently. Given the diversity of domains and types of claims, it seems improbable for a single computational approach to be universally effective at automated fact-checking.

The process of readying iCheck to the public also taught us valuable lessons. Data extraction, cleaning, and linking took huge amounts of effort. Although we have been blessed with high-quality open-source APIs for the U.S. Congress (we relied heavily on GovTrack.us), a lot of work remained to get other related data for iCheck, such as lists of key votes from various lobbying/political organizations, and properly link them to the congressional voting records. These organizations publish their

[5] http://idir.uta.edu/factchecker/apidocs.html

information in different formats and refer to key votes in different ways. References are often incomplete or ambiguous—especially when many roll calls may be associated with the same bill—and linking is further complicated by occasional typos in the data source. While automated data extraction and cleaning techniques have come a long way, they still cannot achieve the accuracy desired for fact-checking. Some of the errors and ambiguities we encountered could only be resolved by input from human experts with intimate knowledge of the Congress. Improving accuracy for specific domains—without a lot of data, let alone expert-labeled data—remains a challenge.

Making iCheck user-friendly also required enormous effort. iCheck provides visualization, exploration, and recommendation features, but to make them useful and accurate, a very high degree of customization was necessary. Accuracy is difficult to achieve because the reality always manages to come up with exceptions to assumptions made by analysis and implementation. For example, legislators can switch party affiliations and voting rights of delegates change over time, complicating even simple accounting queries. Recommendation algorithms—for example, for suggesting related claims that are "surprising" or best "counter" the one being checked—also required lots of expert input and extensive tuning by our developers.

Looking back, we ask ourselves whether all the development effort was worthwhile for an application in a specific domain. iCheck was made public in September 2016. While there were some claims during the 2016 elections that perfectly fit iCheck, the bulk of the check-worthy claims in that season turned out to have nothing to do with congressional voting records. As we probably do not have the luxury of developing a system like iCheck from scratch for every single domain, the key question is whether and how we can develop a more general system or a set of tools that work across multiple domains. Since iCheck, our team has been exploring ways to build more general tools that can work with more types of claims and additional domains, while striking some balance between generality and user-friendliness. Recognizing data quality issues, we are also actively doing research on fact-checking in the presence of uncertain data, and developing techniques for prioritizing data cleaning efforts under resource constraints.

Overall, it has been a humbling experience for the computer scientists involved in the iCheck project to see the wide gamut of knowledge, skills, and efforts required of human fact-checkers and journalists. Plenty of interesting challenges remain in making automated fact-checking more general and more cost-effective.

4 The ClaimReview Schema and Share the Facts

The ClaimReview schema was developed by Jigsaw, a subsidiary of Google, and the Duke Reporters' Lab in an open process with schema.org. The markup is embedded in articles, providing a consistent way for fact-checkers to identify key elements such as the person or group being checked, the statement and the rating or conclusion.

Publishers can use their own content management systems to embed the markup or can use Share the Facts, a free service of the Reporters' Lab. In addition to the markup, Share the Facts also renders a "widget" that can be inserted in an article providing a visual summary.

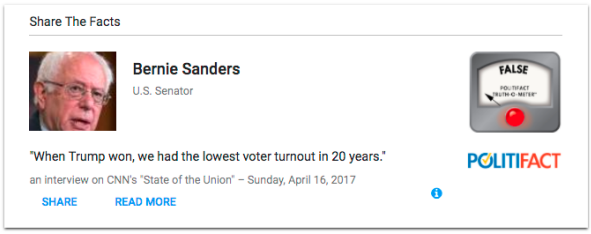

Figure 1: The Share the Facts widget provides a visual summary of a fact-checking article as well as embedding the ClaimReview schema.

The widget can be shared on social media and embedded in articles and blog posts like tweets.[6]

Use of the schema and widget is growing. Approximately 20 fact-checkers around the world are using the ClaimReview schema and an additional 11 are using the Share the Facts widget as of July 2017. Many others have said they plan to adopt one of the two methods in the next six months.

Over the past year, Google has announced a series of product features that leverage the ClaimReview schema to surface and highlight fact-checking articles in Google News and search results. In October 2016, the company began identifying articles that contain the markup with a "FACT CHECK" tag.[7]

Figure 2: In October 2016, Google News began identifying fact-check articles with a unique tag.

[6] http://www.sharethefacts.org/

[7] https://www.blog.google/topics/journalism-news/labeling-fact-check-articles-google-news/

In April 2017, Google began highlighting fact-check articles in search results. To distinguish them from other types of content, they were displayed with enriched textual snippets that concisely summarize the findings.[8]

Figure 3: Google uses the ClaimReview markup to display the fact-check as a rich text snippet in search results.

Most recently, in June 2017, Google started presenting a daily collection of fact-checking articles on the desktop homepage of Google News as part of the News Desktop redesign.[9] In July 2017, Bing, the Microsoft search engine, published information for publishers about how to use ClaimReview that said the markup will be used for "enhanced captions" in search results.[10]

An additional benefit of the markup is that structured summaries of the fact-checks can now be collected in a database that provides content for future applications. It is now possible to easily tap into the complete archive of articles by the world's fact-checkers.

5 Live Pop-Up Fact-Checking

We have made some early progress toward the goal of live pop-up fact-checking. When the television networks decided in the 2016 campaign that they were not going to do live fact-checking of presidential debates, the Duke Reporters' Lab developed FactPopUp, a Chrome browser extension.

FactPopUp is a manual tool to present short summaries of fact-checks on top of live video. It uses Twitter and Chrome's notification feature to display text and images. A fact-checker – for the tests, it was PolitiFact editor Aaron Sharockman – listens to the event and sends a tweet when one of the speakers makes a factual claim that has been previously fact-checked. That triggers a box that pops up on the browser.

We conducted beta tests of FactPopUp with mixed results. Our first test, during the final presidential debate of 2016, provided timely fact-checks after the candidates made factual claims. The web video of the debate was delayed about 15 seconds from the "live" event on television, which provided sufficient time for the PolitiFact editor to find the relevant fact-check that had been previously published and then tweet it so the pop-up appeared just a few moments after the candidate said the claim on the web video. During the debate, FactPopUp provided about 10-12 high-quality notifications about fact-checks on the candidates.

FactPopUp was less successful during the inaugural speech of President Donald Trump. There were only a few factual claims, and the live feed being used for the event happened to be running about 45 seconds behind live television. As a result, the PolitiFact editor triggered the pop-ups when he heard them on live TV, but on the web video, they appeared before Trump actually said the statement.

Currently, we are working on the next generation of pop-up fact-checking where the task of matching previously done fact-checks is automated, allowing the public to benefit from the database of fact-checks collected through the ClaimReview schema on a much bigger scale. A user could ask our app to monitor a web page or a video or audio stream for matching claims, or search the database via text or voice. Going beyond keyword searches, our back-end system could make use of any additional contextual signals provided by the app, such as the stream URL being monitored and time into the stream, to improve matching quality. For some streams, the back-end system can obtain additional information useful to matching, such as full-text transcripts and annotations by human experts.

Besides leveraging our growing database of fact-checks, our system also syncs with ClaimBuster. Search requests for specific claims indicate users find them check-worthy, and logs of such requests can be used by ClaimBuster as training data to improve its claim identification algorithms. Popularities of claims by request also serve an additional criterion with which human fact-checkers can decide what to check next.

Our apps will allow users to subscribe to claims with no existing fact-checks or streams with ongoing fact-checking activities; users are notified as soon as new fact-checks become available. The app can also continue to track published fact-checks, and send any updates and corrections to users.

6 The Challenges and Prospects for Fully Automated Fact-Checking

In the past two years we have made significant progress toward the "Holy Grail." ClaimBuster, iCheck and the ClaimReview schema have provided valuable lessons about the next steps toward fully automated fact-checking.

The ClaimReview schema provides one key element for the "Holy Grail"—a growing database of fact-checks organized as structured data and amenable to automated searching and matching. As we get buy-ins from major technology companies like Google, we are close to offering this part of the "Holy Grail" to the public, maximizing the impact of the hard work that human fact-checkers have created. In the near future, we see this direction as one where we can make the most practical gain with the current technology, and our Share the Facts widget and development of a better "pop-up" fact-checking app are important first steps. In the longer term, more study is needed on novel mechanisms for introducing the results of fact-checks to

[8] https://www.blog.google/products/search/fact-check-now-available-google-search-and-news-around-world/

[9] https://www.blog.google/topics/journalism-news/redesigning-google-news-everyone/

[10] https://www.bing.com/webmaster/help/markup-claim-review-7202cff4

individuals—especially those with strong prior beliefs who may be less receptive to the results. Regardless of the final delivery mechanism—and there may be many alternatives—the system infrastructure that we are building will serve as a solid foundation.

Besides disseminating the results of fact-checks, our work helping journalists produce more fact-checks remains a challenge, but we continue to make good progress and gain new insights. General, end-to-end automated systems are difficult. However, some steps of fact-checking are more amenable to general automation solutions than others. ClaimBuster has identified one sweet spot where AI can help learn what is check-worthy effectively. On the other hand, our experience with iCheck seems to indicate that checking a non-trivial statement automatically starting from just data still requires considerable work. With enough effort, we can probably achieve end-to-end automated fact-checking in specific domains for specific kinds of claims, but generalizing the success to other domains and claim types in a cost-effective manner remains challenging and would require long-term investment in collaborative research between journalists and computer scientists.

To continue our research, we have formed the Tech & Check Cooperative, a team that includes our researchers from Duke University, the University of Texas at Arlington and Google, as well as new partners from the Internet Archive and California Polytechnic State University. The Tech & Check Cooperative has received a grant from the Knight Foundation to continue this important research. In addition to developing apps for live fact-checking and expanding the use of ClaimBuster, the team will communicate with other researchers doing work in this area.

Acknowledgments

This work was partially supported by NSF grants IIS-1408846, IIS-1408928, IIS-1565699, a Knight Prototype Fund from the Knight Foundation, and a Google Faculty Award. Any opinions, findings, and conclusions or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the funding agencies.

References

[1] Naeemul Hassan, Bill Adair, James Hamilton, Chengkai Li, Mark Tremayne, Jun Yang and Cong Yu. The Quest to Automate Fact-Checking. In Proceedings of the 2015 Computation+Journalism Symposium, 5 pages, New York City, USA, October 2015.

[2] Naeemul Hassan, Chengkai Li, and Mark Tremayne. Detecting Check-worthy Factual Claims in Presidential Debates. In Proceedings of the 24th ACM International Conference on Information and Knowledge Management (CIKM), pages 1835-1838, Melbourne, Australia, October 2015.

[3] Naeemul Hassan, Mark Tremayne, Fatma Arslan, and Chengkai Li. Comparing Automated Factual Claim Detection Against Judgments of Journalism Organizations. In Proceedings of the 2016 Computation+Journalism Symposium, 5 pages, Stanford, California, USA, September 2016.

[4] Naeemul Hassan, Fatma Arslan, Chengkai Li, and Mark Tremayne. Toward Automated Fact-Checking: Detecting Check-worthy Factual Claims by ClaimBuster. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), Halifax, Nova Scotia, Canada, August 2017.

[5] Naeemul Hassan, Gensheng Zhang, Fatma Arslan, Josue Caraballo, Damian Jimenez, Siddhant Gawsane, Shohedul Hasan, Minumol Joseph, Aaditya Kulkarni, Anil Kumar Nayak, Vikas Sable, Chengkai Li, and Mark Tremayne. ClaimBuster:The First-ever End-to-end Fact-checking System. In Proceedings of the VLDB Endowment (PVLDB), 10(12), August 2017. Demonstration description.

[6] Brett Walenz, Junyang Gao, Emre Sonmez, Yubo Tian, Yuhao Wen, Charles Xu, Bill Adair, and Jun Yang. "Fact checking congressional voting claims." In Proceedings of the 2016 Computation+Journalism Symposium, Stanford, California, USA, September 2016.

[7] You Wu, Pankaj K. Agarwal, Chengkai Li, Jun Yang, and Cong Yu. "Computational fact checking through query perturbations." ACM Transactions on Database Systems, 42(1), 2017.