KDD2017 Submission 489

Submission information updates are disabled.
For all questions related to processing your submission you should contact the conference organizers. Click here to see information about this conference.
All **reviews sent to you** can be found at the bottom of this page.

Paper 489	
Title:	Toward Automated Fact-Checking: Detecting Check-worthy Factual Claims by ClaimBuster
Paper:	
Track:	Applied Data Science
Author keywords:	Fact-checking Computational Journalism Text Classification
EasyChair keyphrases:	fact checking (463), factual claim (460), worthy factual claim (237), claimbuster score (140), fact check (120), automated fact checking (110), check worthy (110), social medium (90), worthy factual (85), screening sentence (80), presidential debate (80), top quality participant (79), important factual claim (79), fact checking system (79), algorithmic fact checking (79), claim spotting (75), subjectivity classifier (70), knowledge graph (70), ground truth data (63), fact checking organization (63), major candidate (60), twitter account (50), topic distribution (50), presidential election (50), ground truth (50), fact checker (50), data collection (50), human fact checker (47), check worthiness score (47), non factual sentence (47)
Topics:	Computational social science, Natural language processing, Text mining
Abstract:	In this paper, we describe the current state-of-the-art of fact-checking research and describe the approach we have taken with ClaimBuster. We create a novel, human-labeled dataset of check-worthy factual claims using the sentences of the U.S. presidential election general debate transcripts and use natural language processing and supervised learning techniques to develop a factual claim identification model which is one of the core components of the presented fact-checking platform, ClaimBuster. We describe various components of the ClaimBuster system architecture and outline our development plan. We showcase how ClaimBuster is used to live cover the 2016 U.S. presidential election debates and monitor social media platforms and Hansard for identifying check-worthy factual claims. The performance of ClaimBuster is compared with the professional journalists and fact-checking organizations.
Time:	Feb 17, 05:21 GMT
The data used in the experiments is publicly available	No
The code used in the experiments is publicly available	No
The primary author is a full-time graduate student at the time of submission	No
I have adhered to the conference policies in the Call for Papers	I agree
Pick the category that best fits your submission (see CFP for details)	Deployed
Author conflicts:	none

Authors

first name	last name	email	country	organization	Web page	corresponding?
Naeemul	Hassan	nhassan@olemiss.edu	USA	University of Mississippi	http://cs.olemiss.edu/~nhassan	✔
Fatma	Arslan	fatma36dogan@hotmail.com	USA	University of Texas at Arlington		✔
Chengkai	Li	cli@uta.edu	USA	University of Texas at Arlington	http://ranger.uta.edu/~cli/	✔
Mark	Tremayne	tremayne@uta.edu	USA	University of Texas at Arlington		✔

Reviews

Review 3

Novelty:	**3**
Quality:	**3**
Presentation:	**2**
Overall evaluation:	**3**
Strengths:	The pressing need for automated fact checking is widely acknowledge - this is an important problem space The quality validation process described in section 3.2 is very thorough The paper is clear and well written
Weaknesses:	The paper would be improved by describing the approach that will be taken for actually evaluating the truth of check-worthy claims The scoring method could be improved by callibrating against the fact-checking organizations (CNN, Politifact) selected claims Much of section 4 describes analysis of the presidential candidates by external organizations as opposed to results of the study
Review:	This paper describes an interesting system that tackles a challenging problem widely-acknowledged to need more attention at the moment. To check the truth of claims stated in public forums, step one is to distinguish assertions that are worth checking and step 2 is to check them. This paper addresses the first step. The paper is fairly well written and provides a good description of the background and motiviation for the work. The classification setup with 3 categories seems well founded, and the crowd-sourced annotation step that uses screening sentences to validate labelers and redundant annotations with a stopping criteria should provide high quality evaluation data. Section 3.4 describes classification accuracy for checkability. The performance of ClaimBuster is in table 2 and the comparison with subjectivity classifiers is in table 4 and 5. This data would be easier to compare if the 3 tables were formated in a minipage all together. Also, please add more informative table captions. Table 4 and 5 are titled 'comparison with' but the data to be compared with is not shown in that table, only the output of the sentiment classifier. After section 2 sets up the problem as a 3-way classification problem, section 4.1 jumps to describing the individual claims in terms of a 'score', which I found jarring. I think what's being called the score is the probability of a claim being a CFS. Since the remainder of the paper relies on the reader knowing what the score is, it would help to call it out as a numbered equation. The distribution of scores in Figure 6 indicates that very few statements pass the .5 threshold filter that would be needed to survive to the next level of checking. I wondered if this means the .5 threshold is too rigid. Also, the average score for those items that CNN and Politifact thought worthy of checking is at .43. I expected to see some discussion of how these analyses would be used to tune the scoring but at this point the paper devolved into an analysis of the individual political parties. Also, I expected to see some discussion of using the CNN/Politifact analysis data as additional ground truth for tuning the scoring. The authors do point out that the scores of non-checked and checked sentences are significantly different. Some real estate is then dedicated to summarizing the political candidates truthfulness according to verdicts of external sources. This data while interesting in another venue is external to your results. I think the paper would be strengthened if you dedicated some space to discussing how the next phase of processing will be performed: to produce those verdicts algorithmically. Using shallow analysis is a good model to determine with enough power to determine sentence similarity for your checkability classifier but for actually evaluating the claims I anticipate that very different models will be needed. The paper leaves me wondering if there is any plan to tackle that deeper analysis to determine the truth of each claim in the ClaimBuster tool, unlike the other 'abandoned POC' products described in section 2.

Suggestions:	Go into more depth on analyzing the score values. Does the case study on 2016 presidential candidates provide useful material to adjust your score calculation? Describe how the truth of claims will be evaluated, even if this is still planned future work Share the crowdsourced checkability judgments so your results can be replicated

Review 1

Novelty:	**2**
Quality:	**4**
Presentation:	**2**
Overall evaluation:	**3**
Strengths:	well written, quite solid text-mining methodology, difficult problem
Weaknesses:	The author ambition to deploy their system on social networks, but the presidential debate dataset being quite specific, I am wondering if the transfer of the obtained model to other datasets really makes sense.
Review:	The design of a general automated fact-checking system is a really ambitious project. Indeed, some claims may require a lot of human investigation, some (like the existence of Russel's cosmic teapot) are clearly unfalsifiable. This paper is however focused on a more reachable objective: the design of a system to rank and classify check-worthy claims to help journalists in their fact-checking work. An important part of this work is the constitution of a solid training dataset from USA presidential debates labeled in term of NFS (for Non-Factual Sentences), UFS (for Unimportant Factual sentences) and CFS (for Check-worthy Factual Sentences). The labeling process seems rigorous with a screening system to control the labeling quality. The rest of the paper describes a rigorous, but classical, text-mining process where NLP features are extracted from the text and used as input to deploy classifiers (namely RF and SVM used as rankers for CFS vs NFS+UFS). The end of the paper gives some exploratory statistics obtained through the lens of the obtained CFS classifier (distributions, topics, comparison with CNN and Politifact). typo: section 3.5 : "an subjectivity classifier" -> "a subjectivity classifier"
Suggestions:	The authors should develop more thoroughly on how they plan to adapt this experiment to other sources of data (like Web and Social networks).

Review 2

Novelty:	**3**
Quality:	**3**
Presentation:	**2**
Overall evaluation:	**3**
Strengths:	1. This paper is well-written and generally easy to follow. 2. An interesting problem is studied. 3. It might bring significant impact if the authors could release the claim spotting dataset to the research community.
Weaknesses:	1. This paper shows no significant technical contributions. 2. More experiments and analyses could be conducted on the claim spotting task.
Review:	This paper addresses the problem of claim spotting, i.e., to identify factual claims that are worth checking. This task is formalized as a classification/ranking problem, which assigns one of the three labels to each sentence, i.e., non-factual sentence (NFS), unimportant factual sentence (UFS), and check-worthy factual sentence. A dataset with ground-truth collected from human annotators is created and used for this task. The check-worthy scores might also help in assisting professionals choose which facts to check and improving their work efficiency. This paper is well-written and generally easy to follow. The problem studied, i.e., claim spotting, might be of great interest and impact to certain communities. Both the technical and experimental parts are clearly presented and sound reasonable. It would be better if the authors could release the labeled dataset to the research community. I have two concerns about this paper. First, although claim spotting itself is an interesting problem, it is carried out by a typical classification procedure, with frequently used features. There seems no significant technical contribution in this paper.

	Second, as claim spotting is the main task of this paper, the authors could focus more on the task, and provide more experimental results and analyses. For example, maybe the authors could further verify the effectiveness of different features in the classification task (by testing the performance after removing a specific group of features). Overall, despite limited technical contributions, I don't see any fatal errors. Given that an interesting problem has been studied, I think this paper could be accepted if there is room.
Suggestions:	Please provide more experiments and analyses on the claim spotting task (Section 3).