

TATHYA: Towards Context Aware Identification of Check-worthy Claims in Political Discussions

**Ayush Patwari,
Dan Goldwasser,
Saurabh Bagchi
Purdue University**



Work funded in
part by Google
Faculty Award,
NSF SaTC

How this Work Came About?

- We were interested in the topic of man **and** machine to counter malicious communication in social networks
 - Example problem: Fake reviews on social sites
- Basic intuition: Humans and machines possess complementary capabilities in identifying fake or malicious communication
 - Example: Machines can rapidly process large numbers of reviews by evaluating features (e.g., word choice, sentence structure, history of contributor's account, etc.)
 - Humans are slower but can evaluate features that require genuine human understanding (e.g., bias, plausibility, factual contradictions, etc.)
- Work got funded by Google in late 2015
- But then we looked around and found the problem had recently been solved
- So ... we turned in mid 2016 to another problem that needed man and machine working together: Fact Checking

Why Fact Checking?

- Public figures such as politicians make claims about "facts" all the time
- They may be false, exaggerated or misleading, due to careless mistakes and even deliberate manipulation
- With technology, information is spread faster through all types of channels
- Can we use technology to check the veracity of factual claims important to the public?
- Fact checking is thriving: # organizations: 44 (2014) → 115 (Today)

Who is Doing Fact Checking Today?

Regular journalistic outfits



Non-traditional journalistic outfits

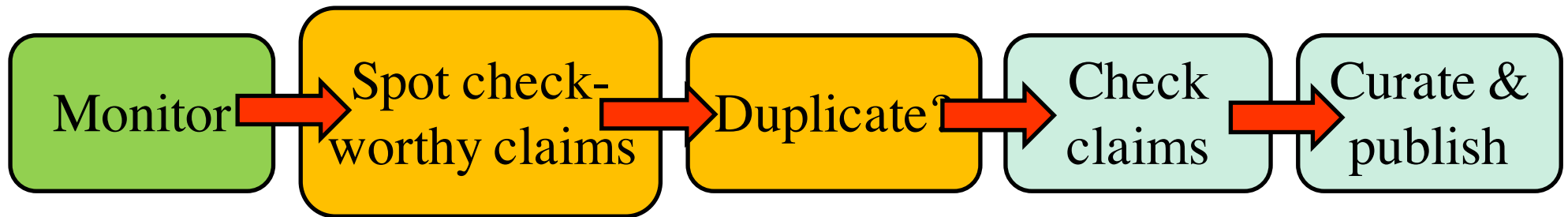


Why Automated Fact Checking?

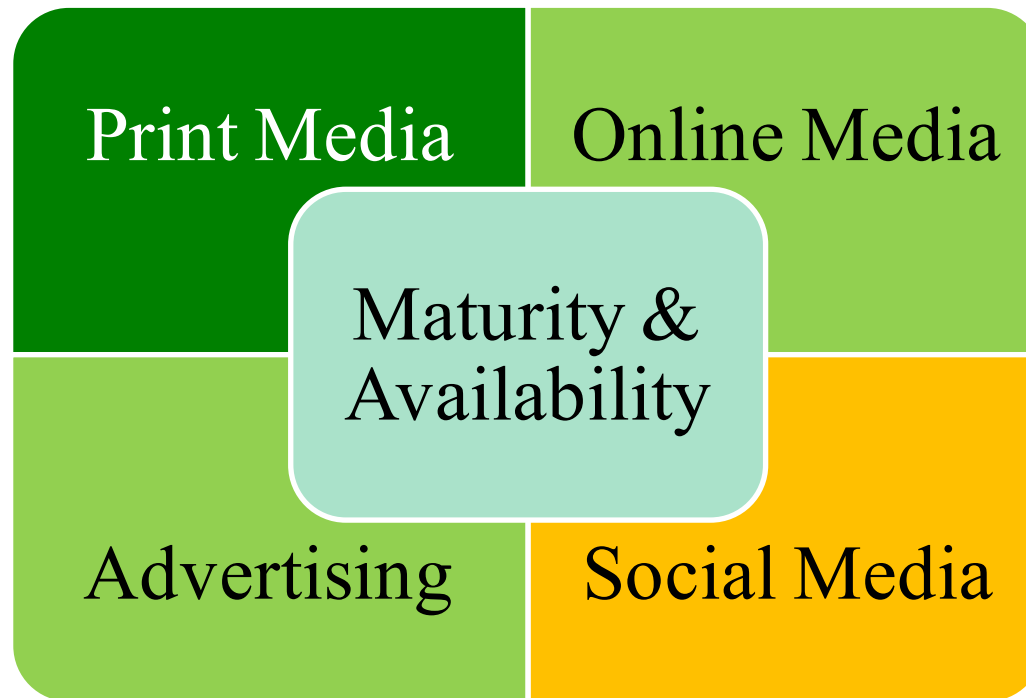
- Speed and volume of news cycle means purely manual fact checking is challenging
 - Time it takes you to read this slide, a software program will have “read” and parsed our 10 page paper on the topic
- Desire for real-time fact checking or within minutes before viral spread of fake news
 - Competitive advantage for fact checking organizations
- Proliferation of sources for spread of news
 - Print media, Online media, Campaign advertising, Social media



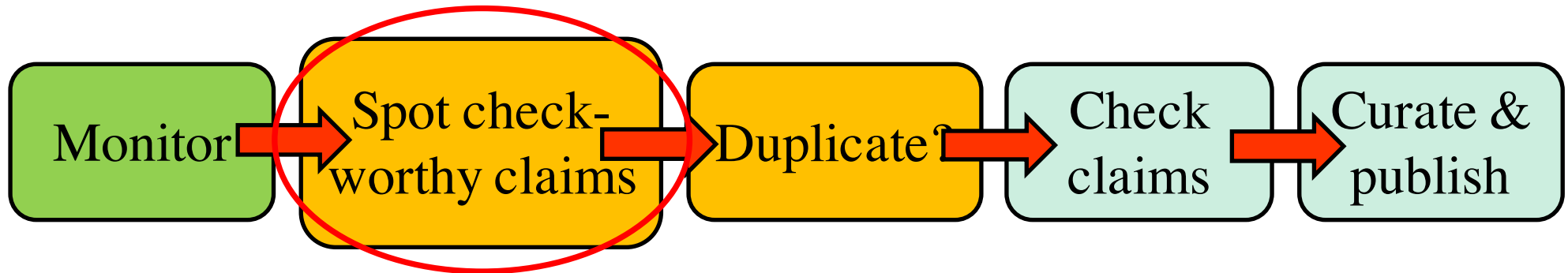
Nominal Pipeline for Automated Fact Checking



Monitor



Focus of This Talk



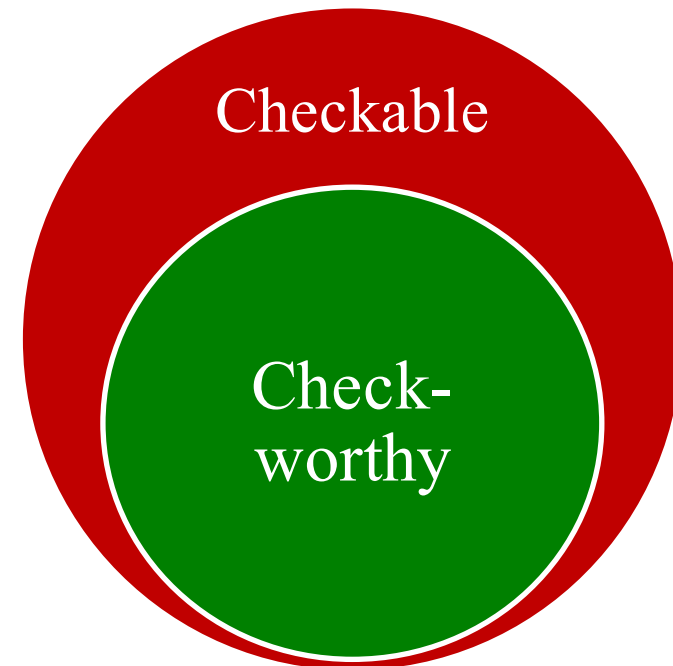
Hardness of Determining Check-worthiness

- **Checkable?**

- ✓ Factual statement, not opinion?
- ✓ Unambiguous?

- **Check-worthy?**

- ✓ Lot more subjective
- ✓ Relies on prior stand of speaker
- ✓ Relies on context of statement in current political discourse
- ✓ Competitive factors of fact-checking organization
- ✓ Bias of fact-checking organization



Checkable and Check-worthy: Examples

- SANDERS: *When this campaign began, I said that we got to end the starvation minimum wage of \$7.25, raise it to \$15.* ✓
Check-worthy

- SANDERS: *I think we have got to be clear, not equivocate, \$15 in minimum wage in 50 states in this country as soon as possible.* ✗

- Same debate **Not check-worthy**

Checkable and Check-worthy: Examples

- O' MALLEY: Senator Sanders  ed against the Brady Bill

Check-worthy

- CLINTON: I have been for the Brady bill, I have been against assault weapons



- Different debates

Not check-worthy

Current State-of-Art

- Supervised classification algorithm: ClaimSpotter (*)
- Labeled training set generated by non-expert human annotators from 30 debate episodes, 2004-12
 - Sentences spoken by 18 presidential candidates: total of 20,788 sentences
 - Annotators could use up to 5 previous sentences for context
- Support Vector Machine (SVM) classification into 2 classes: check-worthy and (not checkable + checkable but not check-worthy)
 - Best performance from using features: Bag-of-words (unigram), Part-of-speech, Entity type
 - Precision: 70%, Recall: 72%, F-measure: 70%

(*) Naeemul Hassan, Chengkai Li, and Mark Tremayne. "Detecting check-worthy factual claims in presidential debates." 24th ACM CIKM, pp. 1835-1838, 2015.

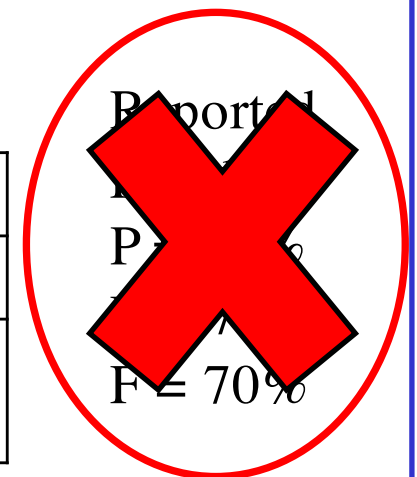
Naeemul Hassan, et al., "ClaimBuster: The First ever End-to-end Factchecking System," VLDB (demo), pp. 1-4, 2017.



Our Thought: First Let Us Replicate Results

- Dataset: 15 primary Presidential debates from 2015-16 (7 Republican, 8 Democratic); 3 Presidential debates
- Statement labeled check-worthy if *any* of 8 national fact-checking organizations checked it
- Dataset size:
 - 21,700 statements; 1,085 marked check-worthy
 - After pruning: 15,735 statements; 1,085 marked check-worthy (6.1%)

	Primary Debates			Presidential Debates		
	P	R	F	P	R	F
Claim-Spotter	19.4%	32.0%	24.1%	22.6%	14.8%	17.9%



Bring In Context

- What if we could bring context into the picture?
- What if we could bring topic modeling into the picture?
- Context: Rather than consider individual sentences, use chunk and use context around it as features
 - Definition: Named Entity Recognition (NER) labels sequences of words in a text which are the names of things, such as (Person, Organization, Location)
 - Named entity in current chunk – is it used in support or opposition in previous X chunks – how frequently is it used

Bring in Topics

- Topics: Claims in certain topics (e.g., gun control) are more likely to be checked by fact-checkers as compared to those on personal life
 - Definition: LDA represents “documents” as mixtures of topics and each topic spits out words with certain probabilities
 - LDA topic model with user-defined number of topics
 - 20 topics were found to be enough
- Change of topics: Could indicate evasion and repeated change could mean persistent desire to pin down speaker
 - Cosine similarity of topic distribution with $X/2$ previous and $X/2$ following chunks

Bring in Text Normalization

- It is common to refer to entities using second and third person pronouns in a discussion
- This information is lost when analyzing a statement ‘out-of-context’
- We perform text normalization by propagating named entities along a discussion
- Example: *Yes look, I have made it clear based on Senator Sanders’ own record that he has voted with the NRA, with the gun lobby numerous times. ... He voted for immunity from gunmakers and sellers which the NRA said, “was the most important piece of gun legislation in 20 years.”*



Who is he?

Quiz

- The following are 20 statements made by our Presidential candidates during the Primary Debates in 2016
 - The statement to consider is in **blue**; **the rest of the text** provides context
- For each statement, determine if the statement is:
 - Check-worthy
 - Not check-worthy
- Ground truth will come from if the statement was checked by any of 9 reputed fact-checking organizations: Washington Post, Factcheck.org, Politifact, PBS, CNN, NYTimes, Fox News, USA Today
- If you determine that a statement is check-worthy, indicate on the following scale the result of the check:
 - True, Mostly true, Partly true-partly false, Mostly false, False



TATHYA: Putting It All Together

- SVM classifier that does binary classification: Check-worthy and (Not checkable OR Checkable but not check-worthy)
- Best result is obtained with the following feature sets:
 - Bag-of-words, POS, Entity recognition, POS tuples, Context

	Primary Debates			Presidential Debates		
	P	R	F	P	R	F
Claim-Spotter	19.4%	32.0%	24.1%	22.6%	14.8%	17.9%
TATHYA	19.3%	43.5%	26.3%	22.7%	19.4%	20.9%

Sources of Error

- False positives: For 1/3 of the case, the model was wrong – for remaining, we could not objectively decide why they had not been checked
- In discussion we found that these were driven by pragmatic factors
- Manual examination of sampled missed statements
 1. Sentence context not captured sufficiently: 21.6%
 2. Implication was checked: 8.1%
 3. Unit of checking was subjective: 10.8%
 4. Poorly structured sentences: 10.8%
 5. Model simply wrong: 46.8%

Error Examples

- SANDERS: *I am going to release all of the transcripts of the speeches that I gave on Wall Street behind closed doors, not for \$225,000, not for \$2,000, not for two cents.*
- TATHYA: *Check-worthy*
ClaimSpotter: *Check-worthy*
Reality: *Not check-worthy*
- CLINTON: *...and stood with the Minutemen vigilantes in their ridiculous, absurd efforts to, quote, "hunt down immigrants.*
- TATHYA: *Not Check-worthy*
ClaimSpotter: *Not Check-worthy*
Reality: *Check-worthy*

Take-Aways

- We tackle the problem of detecting whether political statements are check-worthy or not
- This is the first step in automating fact-checking
- We find that this problem is made difficult by: fact-checkers' subjectivity, understanding dynamics of discussion and incorporating world knowledge
- Our error analysis lays groundwork for further challenges in learning to automate fact-checking
- Our work: TATHYA
 - Focuses on exploiting the semantic context and debate dynamics.
 - Uses a set of different classes of features—bag-of-words, topic agreement, entity history and targeted part-of-speech tuples, reference resolution
 - Our best result for the classifier is an F1-score of 26.3%, only
 - Recall, arguably the more important metric for computational journalism, provides a better performance with a value of 43.5%

Resources

- Reporters' Lab (Duke University)
<https://reporterslab.org/fact-checking/>
- NPR, September 27, 2016
“Do Fact Checks Matter?”
- ClaimBuster (UT Arlington)
<http://idir-server2.uta.edu/claimbuster/>
- Flynn, D. J., Brendan Nyhan, and Jason Reifler. "The nature and origins of misperceptions: Understanding false and unsupported beliefs about politics." *Political Psychology* 38, no. S1 (2017): 127-150.