Automated Fact-Checking

Joshua Chen

Universität Bonn
Seminar Natural Language Processing SS 2017

20 Jun 2017

What is fact-checking?

Fact-checking is the task of verifying the truth of factual claims, in particular those propagated to a large public audience.

What is fact-checking?

Fact-checking is the task of verifying the truth of factual claims, in particular those propagated to a large public audience.

Examples of claims:

- News reports
- Statements made in political/other public debates
- Social media content — viral content, rumours
- Fake news — journalistic content designed to mislead

Examples of claims

Statements in public debates

"...adoptions went up 65 to 70 percent...when I was mayor."

– Rudy Giuliani, 2007 Republican presidential debate

Examples of claims

Statements in public debates

"...adoptions went up 65 to 70 percent...when I was mayor."

– Rudy Giuliani, 2007 Republican presidential debate

Rumours in Twitter tweets

Oh my god is this real? RT @AP: Breaking: Two Explosions in the White House and Barack Obama is injured
Is this real or hacked? RT @AP: Breaking: Two Explosions in the White House and Barack Obama is injured
Is this legit? RT @AP: Breaking: Two Explosions in the White House and Barack Obama is injured

Examples of claims

Statements in public debates

". . . adoptions went up 65 to 70 percent. . . when I was mayor."

– Rudy Giuliani, 2007 Republican presidential debate

Rumours in Twitter tweets

Oh my god is this real? RT @AP: Breaking: Two Explosions in the White House and Barack Obama is injured
Is this real or hacked? RT @AP: Breaking: Two Explosions in the White House and Barack Obama is injured
Is this legit? RT @AP: Breaking: Two Explosions in the White House and Barack Obama is injured

News headlines

Nigerian professor solves 150 year old maths problem – 2015 BBC interview with Nigerian professor claiming to have solved the Riemann hypothesis

Fact-checking today

> *"Falsehood flies, and the truth comes limping after."*
>
> – Jonathan Swift (1710)

Two challenges:

Fact-checking today

"Falsehood flies, and the truth comes limping after."

– Jonathan Swift (1710)

Two challenges:

1. Too much information to process — unbacked claims are easy to make but hard to confirm/refute in our "big data" age.

Fact-checking today

"Falsehood flies, and the truth comes limping after."

– Jonathan Swift (1710)

Two challenges:

1. Too much information to process — unbacked claims are easy to make but hard to confirm/refute in our "big data" age.
2. Social media enables anyone to create content with massive reach. Need methods to cope with this new channel of information to manage negative effects of misinformation (hysteria, dishonest mass influence of election outcomes, etc.).

Fact-checking today

Example:

Automated fact-checking

Automated aka **computational fact-checking** is using machine learning, statistical, and other techniques to automate (aspects of) fact-checking.

Various aspects/subtasks:

Automated fact-checking

Automated aka **computational fact-checking** is using machine learning, statistical, and other techniques to automate (aspects of) fact-checking.

Various aspects/subtasks:

- Detection and ranking of claims

Automated fact-checking

Automated aka **computational fact-checking** is using machine learning, statistical, and other techniques to automate (aspects of) fact-checking.

Various aspects/subtasks:

- Detection and ranking of claims
- Detection of rumours trending in social media

Automated fact-checking

Automated aka **computational fact-checking** is using machine learning, statistical, and other techniques to automate (aspects of) fact-checking.

Various aspects/subtasks:

- Detection and ranking of claims
- Detection of rumours trending in social media
- Automated checking of simple claims — "Einstein was born in Ulm", "The Danube flows through Serbia", "Obama is a Muslim".

Automated fact-checking

Automated aka **computational fact-checking** is using machine learning, statistical, and other techniques to automate (aspects of) fact-checking.

Various aspects/subtasks:

- Detection and ranking of claims
- Detection of rumours trending in social media
- Automated checking of simple claims — "Einstein was born in Ulm", "The Danube flows through Serbia", "Obama is a Muslim".
- Formulating more sophisticated mathematical models — checking more complicated claims (numerical/quantitative/dates etc.), finding counterarguments, reverse-engineering vague claims to make them more precise. . .

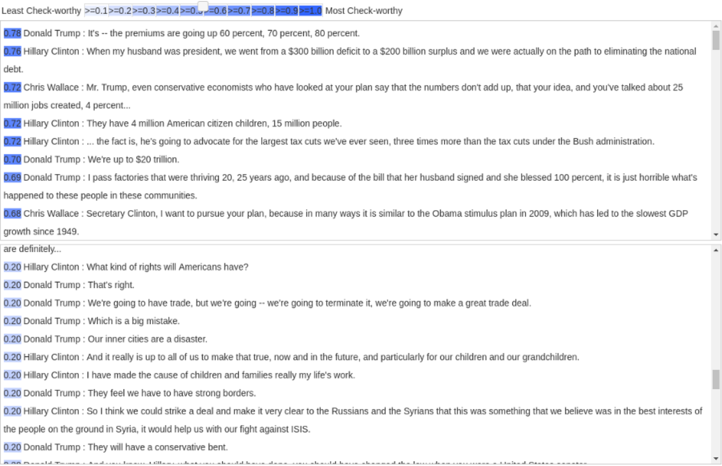
Claim detection and ranking

Hassan et al. [1, 2]

ClaimBuster (`http://idir-server2.uta.edu/claimbuster/`) scores sentences based on how "checkworthy" they are.

- Supervised learning — train a classifier to classify statements as "nonfactual statements", "unimportant factual statements" and "important factual statements".
- Use Platt scaling to obtain a probability score that a given sentence is an important factual statement.

Claim detection and ranking



Text from transcript of 2016 Third US presidential debate.

Automatic rumour detection

Zhao, Resnick and Mei [3]

Rumour: *"a controversial and fact-checkable statement"*.

Automatic rumour detection

Zhao, Resnick and Mei [3]

Rumour: *"a controversial and fact-checkable statement"*.

Automatically detecting if an individual tweet contains a rumour is difficult.

Automatic rumour detection

Zhao, Resnick and Mei [3]

Rumour: *"a controversial and fact-checkable statement"*.

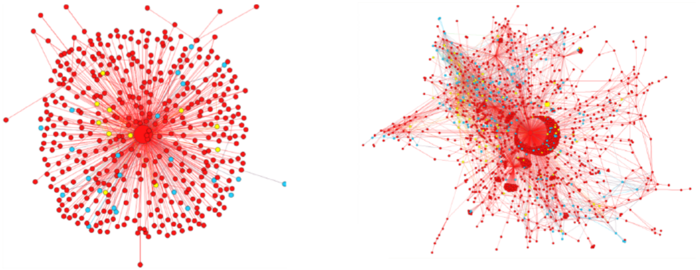
Automatically detecting if an individual tweet contains a rumour is difficult.

Instead, classify *clusters* of tweets containing a given rumour based on how many of them contain **enquiry signals**, i.e. tweets containing statements like "really?" or "what?"

Automatic rumour detection

1. **Identify signal tweets** containing predefined enquiry signal strings.
2. **Cluster signal tweets** into clusters of similar content.
3. **Determine the common text/topic** of the signal clusters.
4. **Capture all non-signal tweets** (i.e. those that contain the same topic but do not contain enquiry signals) and merge them into signal clusters to form candidate rumour clusters.
5. Using statistical features **rank candidate rumour clusters** by the likelihood that their topic is a rumour.

Automatic rumour detection



White House explosion rumour. Red, yellow and blue nodes: tweets spreading, correcting and questioning the rumour.

Left: 60 seconds after the source there were already sufficient enquiry nodes.

Right: After the rumour went viral.

Automated checking of simple claims

Ciampaglia et al. [4], Shi and Weninger [5]

Consider *relational claims*, i.e. claims linking two entities via some relation.

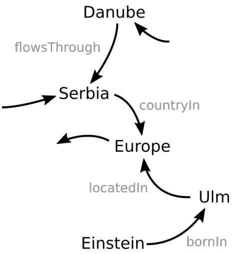
$$\text{Einstein} \xrightarrow{\text{was born in}} \text{Ulm}$$

$$\text{The Danube} \xrightarrow{\text{flows through}} \text{Serbia}$$

$$\text{Barack Obama} \xrightarrow[\text{(has religion)}]{\text{is a}} \text{Muslim}$$

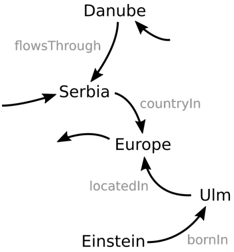
Automated checking of simple claims

- Form a **knowledge graph** from a collection of true relational triples. (Nodes are entities, edges are relation predicates.)



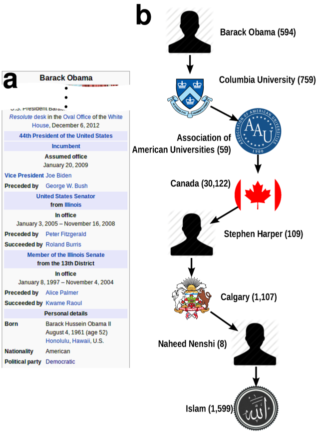
Automated checking of simple claims

- Form a **knowledge graph** from a collection of true relational triples. (Nodes are entities, edges are relation predicates.)



- Determining the truth of a new relational claim can be modeled by various graph problems, e.g. link prediction, path mining, path weight optimization. . .

Is Obama a Muslim?



(a) Knowledge graph defined from Wikipedia infobox data. Weight $W(P)$ of a path P between two nodes decreases with the degree of intermediate nodes.

(b) The path between "Barack Obama" and "Islam" with greatest weight passes through the "Canada" node which has large degree. The path thus has low weight, indicating low support for the statement "Obama is a Muslim".

(Ciampaglia et al. [4])

More sophisticated fact-checking frameworks

Wu et al. [6]

More complex claims:

> *"... adoptions went up 65 to 70 percent... [during 1996–2001 compared to 1990–1995]"*
>
> *"Jim Marshall [US Democratic politician] voted the same as Republican leaders 65 percent of the time"*

More sophisticated fact-checking frameworks

Wu et al. [6]

More complex claims:

> *"...adoptions went up 65 to 70 percent...[during 1996–2001 compared to 1990–1995]"*
>
> *"Jim Marshall [US Democratic politician] voted the same as Republican leaders 65 percent of the time"*

- Contain quantitative/date information.

More sophisticated fact-checking frameworks

Wu et al. [6]

More complex claims:

> *"...adoptions went up 65 to 70 percent...[during 1996–2001 compared to 1990–1995]"*
>
> *"Jim Marshall [US Democratic politician] voted the same as Republican leaders 65 percent of the time"*

- Contain quantitative/date information.
- Often don't present the whole story — can we find counterarguments?

More sophisticated fact-checking frameworks

Wu et al. [6]

More complex claims:

> *"...adoptions went up 65 to 70 percent...[during 1996–2001 compared to 1990–1995]"*
>
> *"Jim Marshall [US Democratic politician] voted the same as Republican leaders 65 percent of the time"*

- Contain quantitative/date information.
- Often don't present the whole story — can we find counterarguments?
- Often vague — can we reverse-engineer to make them more precise?

More sophisticated fact-checking frameworks

A very high-level overview:

- The numerical/date data in claims can be thought of as parameters that can be perturbed.
- Varying these parameters yield statements with different truth scores, in this way we obtain a function surface of the truth score against the parameters.
- Optimizing over the function lets us determine how "robust" a claim is, reverse-engineer imprecise claims, find counterarguments etc.

References

[1] Hassan, Adair, Hamilton, Li, Tremayne, Yang and Yu. *The Quest to Automate Fact-Checking*. 2015.

[2] Hassan, Li and Tremayne. *Detecting Check-worthy Factual Claims in Presidential Debates*. 2015.

[3] Zhao, Resnick and Mei. *Enquiring Minds: Early Detection of Rumours in Social Media from Enquiry Posts*. 2015.

[4] Ciampaglia, Shiralkar, Rocha, Bollen, Menczer and Flammini. *Computational Fact Checking from Knowledge Networks*. 2015.

[5] Shi and Weninger. *Discriminative Predicate Path Mining for Fact Checking in Knowledge Graphs*. 2016.

[6] Wu, Agarwal, Li, Yang and Yu.*Toward Computational Fact-Checking*. 2014.