Automated Fact-Checking

Joshua Chen

Universität Bonn
Seminar Natural Language Processing SS 2017

25 Jul 2017

Automated fact-checking

Automated fact-checking

Using machine learning, statistical and other techniques to automatically verify the truth of factual claims.

Automated fact-checking

Automated fact-checking

Using machine learning, statistical and other techniques to automatically verify the truth of factual claims.

Today:

- Knowledge graph based fact-checking of relational claims.
- Framework for checking general claims, finding counterarguments and reverse-engineering vague claims.

Knowledge graphs

Assume knowledge base in the form of **relational triples** `(subject, predicate, object)`.

Knowledge graphs

Assume knowledge base in the form of **relational triples**
`(subject, predicate, object)`.

e.g. (*"Obama"*, *"is a"*, *"Muslim"*),
(*"the Danube"*, *"flows through"*, *"Serbia"*)

Knowledge graphs

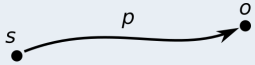
Assume knowledge base in the form of **relational triples** `(subject, predicate, object)`.

e.g. (*"Obama"*, *"is a"*, *"Muslim"*),
(*"the Danube"*, *"flows through"*, *"Serbia"*)

Knowledge graph $\mathcal{G}$

Vertices $\leftrightarrow$ subject and object entities

Edges $\leftrightarrow$ predicates between corresponding subject and object



$\mathcal{G}$ may be directed/undirected, labeled with extra info...

Link prediction

Key idea

Fact-checking can be viewed as a **link prediction** problem over the knowledge graph $\mathcal{G}$.

Link prediction

Key idea

Fact-checking can be viewed as a **link prediction** problem over the knowledge graph $\mathcal{G}$.

- $\mathcal{G} = (V, E)$ has a "true" edge set $\widetilde{E}$.

Link prediction

Key idea

Fact-checking can be viewed as a **link prediction** problem over the knowledge graph $\mathcal{G}$.

- $\mathcal{G} = (V, E)$ has a "true" edge set $\widetilde{E}$.
- $E \subseteq \widetilde{E}$ is potentially incomplete: some relational triples (s, p, o) may be true but missing from our knowledge base.

Link prediction

Key idea

Fact-checking can be viewed as a **link prediction** problem over the knowledge graph $\mathcal{G}$.

- $\mathcal{G} = (V, E)$ has a “true” edge set $\widetilde{E}$.
- $E \subseteq \widetilde{E}$ is potentially incomplete: some relational triples (s, p, o) may be true but missing from our knowledge base.
- Want to use structural properties of $\mathcal{G}$ to approximate $\widetilde{E}$: given a candidate edge e corresponding to (s, p, o), determine if $e \in \widetilde{E}$, i.e. if (s, p, o) is true.

Truth as proximity measures

Ciampaglia et al. [1]

Undirected knowledge graph $\mathcal{G}$, vertices labeled with their corresponding entity names.

Truth as proximity measures

Ciampaglia et al. [1]

Undirected knowledge graph $\mathcal{G}$, vertices labeled with their corresponding entity names.

Semantic proximity

For a path $P = v_1 v_2 \cdots v_n$ define

$$W(P) = \left(1 + \sum_{i=2}^{n-1} \log \delta(v_i)\right)^{-1}$$

where $\delta(v)$ is the degree of v in $\mathcal{G}$.

Semantic proximity captures the heuristic of *specificity*.

Truth as proximity measures

Truth score

Given a claim $c = (s, p, o)$, the **truth score** is

$$\tau(c) = \max\{W(P) \mid P \text{ is a path between } s \text{ and } o \text{ in } \mathcal{G}\}.$$

Truth as proximity measures

Truth score

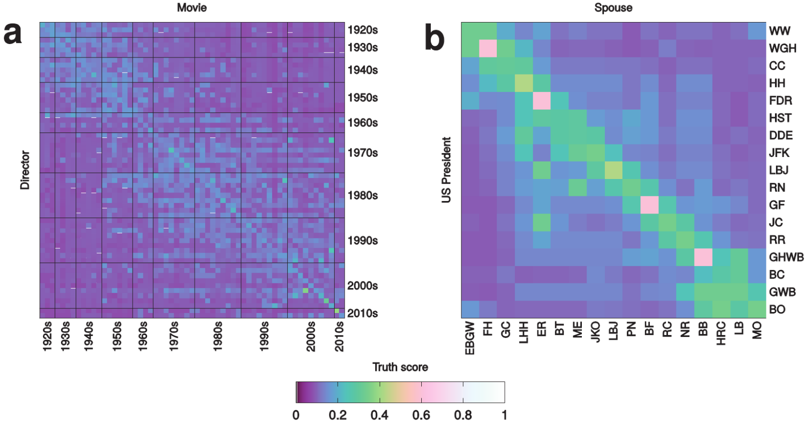
Given a claim $c = (s, p, o)$, the **truth score** is

$$\tau(c) = \max\{W(P) \mid P \text{ is a path between } s \text{ and } o \text{ in } \mathcal{G}\}.$$

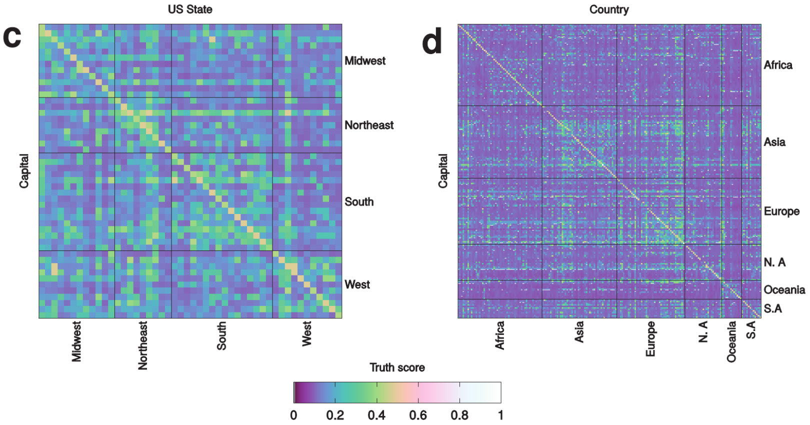
Algorithm

Given a claim (s, p, o) we compute its truth score. The higher the truth score the more confident we are that it is true.

Results


a
Movie
Director
1920s
1930s
1940s
1950s
1960s
1970s
1980s
1990s
2000s
2010s
b
Spouse
US President
WW
WGH
CC
HH
FDR
HST
DDE
JFK
LBJ
RN
GF
JC
RR
GHWB
BC
GWB
BO
EBGW FH GC LHH ER BT ME JKO LBJ PN BF RC NR BB HRC LB MO
Truth score
0 0.2 0.4 0.6 0.8 1

Results


US State
c
Midwest
Northeast
South
West
Capital
Midwest
Northeast
South
West
Country
d
Africa
Asia
Europe
N. A
Oceania
S.A
Capital
Africa
Asia
Europe
N. A
Oceania
S.A
Truth score
0
0.2
0.4
0.6
0.8
1

Other work

Discriminative path mining (Shi and Weninger [3]):

Other work

Discriminative path mining (Shi and Weninger [3]):

- $\mathcal{G}$ is directed. Vertices *and edges* are labeled with entity and predicate names.

Other work

Discriminative path mining (Shi and Weninger [3]):

- $\mathcal{G}$ is directed. Vertices *and edges* are labeled with entity and predicate names.
- Vertices are further labeled with their ontology (set of classes to which the entity belongs).

Other work

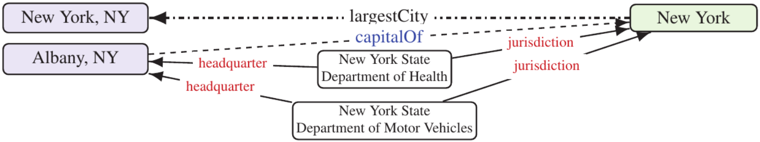
Discriminative path mining (Shi and Weninger [3]):

- $\mathcal{G}$ is directed. Vertices *and edges* are labeled with entity and predicate names.
- Vertices are further labeled with their ontology (set of classes to which the entity belongs).
- First seek to understand a claim e.g. (*"New York city"*, *"capital of"*, *"New York"*) by generalizing to the ontology e.g. (`U.S. city`, *"capital of"*, `U.S. state`)

Other work

Discriminative path mining (Shi and Weninger [3]):

- $\mathcal{G}$ is directed. Vertices *and edges* are labeled with entity and predicate names.
- Vertices are further labeled with their ontology (set of classes to which the entity belongs).
- First seek to understand a claim e.g. (*"New York city"*, *"capital of"*, *"New York"*) by generalizing to the ontology e.g. (`U.S. city`, *"capital of"*, `U.S. state`)



More general fact-checking

Wu et al. [4]

Example

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

Rudy Giuliani, 2007 Republican presidential debate

This claim is:

More general fact-checking

Wu et al. [4]

Example

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

Rudy Giuliani, 2007 Republican presidential debate

This claim is:

1. **vague**—the precise increase in adoptions is rounded, and the time frame is not stated.

More general fact-checking

Wu et al. [4]

Example

"Adoptions went up 65 to 70 percent... when I was mayor"
Rudy Giuliani, 2007 Republican presidential debate

This claim is:

1. **vague**—the precise increase in adoptions is rounded, and the time frame is not stated.
2. **misleading**—upon clarification, the exact time frame was cherry-picked to present the increase as greater than it actually was over the period Giuliani was mayor [2].

More general fact-checking

If claims are:

1. **vague**, then we want to reverse-engineer them to be more precise.
2. **misleading**, then we want to find counterarguments exposing the actual trends.

More general fact-checking

If claims are:

1. **vague**, then we want to reverse-engineer them to be more precise.
2. **misleading**, then we want to find counterarguments exposing the actual trends.

Key idea

A general factual claim contains **parameters** that we can vary in order to change its **result**.

e.g. *"Unemployment decreased by 20 percent between 2012 and 2016."*

Basic framework

Definitions

- A **parametrized query template** $q \colon \mathcal{P} \to \mathcal{R}$ maps the **parameter space** $\mathcal{P}$ of a claim to its **result space** $\mathcal{R}$.
- A claim is represented by a triple (q, p, r) where $p \in \mathcal{P}$ and $r \in \mathcal{R}$.
- A **relative parameter sensibility** function $S_P \colon \mathcal{P} \times \mathcal{P} \to \mathbb{R}$ gives the sensibility of one parameter setting relative to another.
- A **relative result strength** function $S_R \colon \mathcal{R} \times \mathcal{R} \to \mathbb{R}$ gives the strength of one result relative to another.

Relative sensibility and strength

Let (q, p_0, r_0) be a claim.
$S_R(r, r_0)$ is the strength of r relative to the reference result r_0.
$S_R(r, r_0) < 0$ means that r is **weaker** than r_0.

Relative sensibility and strength

Let (q, p_0, r_0) be a claim.
$S_R(r, r_0)$ is the strength of r relative to the reference result r_0.
$S_R(r, r_0) < 0$ means that r is **weaker** than r_0.

Example

"Unemployment decreased by 20 percent between 2012 and 2016."
If the actual unemployment rate was 10 percent we may say that the true result $r = 0.1$ is weaker than the claimed result $r_0 = 0.2$.

Relative sensibility and strength

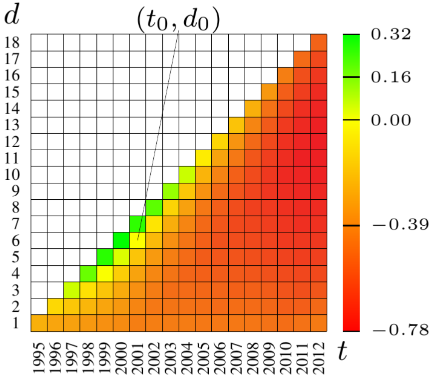
Let (q, p_0, r_0) be a claim.
$S_R(r, r_0)$ is the strength of r relative to the reference result r_0.
$S_R(r, r_0) < 0$ means that r is **weaker** than r_0.

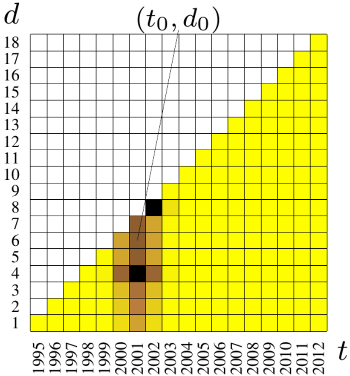
Example

"Unemployment decreased by 20 percent between 2012 and 2016."
If the actual unemployment rate was 10 percent we may say that the true result $r = 0.1$ is weaker than the claimed result $r_0 = 0.2$.

$S_P(p, p_0)$ is the sensibility of p relative to the reference parameter setting p_0. $S_P(p, p_0) > 0$ means that p is **more sensible** than p_0.

Relative sensibility and strength





Relative result strength (left) and parameter sensibility (right) surfaces relative to Giuliani's adoption claim.

Task formulations

Let a claim (q, p_0, r_0) be given.

Task formulations

Let a claim (q, p_0, r_0) be given.

Reverse engineering

Find parameter settings p maximizing sensibility relative to p_0, with result as close as possible to r_0 (i.e. minimizing $|S_R(q(p), r_0)|$).

Task formulations

Let a claim (q, p_0, r_0) be given.

Reverse engineering

Find parameter settings p maximizing sensibility relative to p_0, with result as close as possible to r_0 (i.e. minimizing $|S_R(q(p), r_0)|$).

Counterargument finding

Find parameter settings p maximizing sensibility relative to p_0, that weaken the original claim as much as possible (i.e. minimizing $S_R(q(p), r_0)$.

Task formulations

Let a claim (q, p_0, r_0) be given.

Reverse engineering

Find parameter settings p maximizing sensibility relative to p_0, with result as close as possible to r_0 (i.e. minimizing $|S_R(q(p), r_0)|$).

Counterargument finding

Find parameter settings p maximizing sensibility relative to p_0, that weaken the original claim as much as possible (i.e. minimizing $S_R(q(p), r_0)$).

Bicriteria optimization problems over the result strength and parameter sensibility surfaces, can be solved by enumerating $p \in \mathcal{P}$ in a suitable way.

References

[1] G. L. Ciampaglia, P. Shiralkar, L. M. Rocha, J. Bollen, F. Menczer, and A. Flammini. 2015. Computational Fact Checking from Knowledge Networks. *PLoS ONE* 10, 6 (June 2015), e0128193. DOI: `http://doi.org/10.1371/journal.pone.0128193`

[2] B. Jackson. 2007. Levitating Numbers. (May 2007). Retrieved July 17, 2017 from `http://www.factcheck.org/2007/05/levitating-numbers/`

[3] B. Shi and T. Weninger. 2016. Discriminative predicate path mining for fact checking in knowledge graphs. *Know.-Based Syst.* 104, C (July 2016), 123-133. DOI: `http://dx.doi.org/10.1016/j.knosys.2016.04.015`

[4] Y. Wu, P. K. Agarwal, C. Li, J. Yang, and C. Yu. 2014. Toward computational fact-checking. *Proc. VLDB Endow.* 7, 7 (March 2014), 589-600. DOI: `http://dx.doi.org/10.14778/2732286.2732295`