Automated Fact-Checking

Andreas Vlachos
University of Sheffield
What do fact-checkers do?

The United Kingdom has ten times Italy’s number of immigrants.

FALSE: We find no data to support this claim. The UK does not have "ten times Italy’s number of immigrants".

<table>
<thead>
<tr>
<th>Country/Immigration</th>
<th>Italy</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>4.92M</td>
<td>5.05M</td>
</tr>
<tr>
<td>2015</td>
<td>5.01M</td>
<td>5.42M</td>
</tr>
<tr>
<td>2016</td>
<td>5.03M</td>
<td>5.64M</td>
</tr>
</tbody>
</table>
Demand for fact-checking is growing

- “Falsehood flies, and the Truth comes limping after it” (*Swift, 1710*)
- Social media help *(mis-)*information spread (more) rapidly
- A manual fact-check takes avg. 1 day (*Hassan et al., 2015*)
Does fact-checking work?

Some studies found that:

- Backfire effect: even more convinced of their original opinions (Nyhan and Reifler, 2015: Does correcting myths about the flu vaccine work?)
- Voters do not change their preferences due to fact-checks (Nyhan et al., 2017: Taking Corrections Literally But Not Seriously?)

It works, but it takes time and consistent effort:

- Gregg et al., 2017: The SPOT effect: People spontaneously prefer their own theories.
Can we help with machine learning?
Automated fact-checking

The United Kingdom has ten times Italy’s number of immigrants.

FALSE: We find no data to support this claim. The UK does not have "ten times Italy’s number of immigrants".

<table>
<thead>
<tr>
<th>Country/Immigration</th>
<th>Italy</th>
<th>UK</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014</td>
<td>4.92M</td>
<td>5.05M</td>
</tr>
<tr>
<td>2015</td>
<td>5.01M</td>
<td>5.42M</td>
</tr>
<tr>
<td>2016</td>
<td>5.03M</td>
<td>5.64M</td>
</tr>
</tbody>
</table>
What do we want from automated fact-checking?

- Verdict justification, a.k.a. algorithmic transparency
  - Can’t convince otherwise
  - Need to check their correctness

- Generalization to different domains (economy, health, etc.)

- Learning with little data

*(Vlachos and Riedel, 2014)*
What claims should we fact-check?

- Syrian refugees are not properly vetted or tracked by the FBI once in the US
- Leaving the EU would put 3M jobs at risk

- Does it matter who said these?
- Is the linguistic style matter?
First step: simple numerical claims

Lesotho, a landlocked enclave of South Africa, has a population of 2 million and covers an area slightly smaller than the U.S. state of Maryland.

(Vlachos and Riedel, 2015)
Fact-checking simple numerical claims

Lesotho, a landlocked enclave of South Africa, has a population of 2 million and covers an area slightly smaller than the U.S. state of Maryland.

Identification

Entity: Lesotho
Property: population
Value claimed: 2,000,000

Verification

Absolute percentage error: 0.09

(Vlachos and Riedel, 2015)
Learning how populations are expressed in text

The population of France is 66M.

Iceland has 300K inhabitants.

32.4M tourists visited Russia.

<table>
<thead>
<tr>
<th></th>
<th>France</th>
<th>Russia</th>
<th>Iceland</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>population</strong></td>
<td>66,028,467</td>
<td>143,700,000</td>
<td>325,600</td>
</tr>
<tr>
<td>the population of X is _</td>
<td>66,000,000</td>
<td>140,000,000</td>
<td>325,000</td>
</tr>
<tr>
<td>X’s population is estimated at _</td>
<td>66,030,000</td>
<td>145,000,000</td>
<td></td>
</tr>
<tr>
<td>X’s inflation rate is _</td>
<td>0.9</td>
<td>4.0</td>
<td></td>
</tr>
<tr>
<td>X has _ inhabitants</td>
<td></td>
<td>140,000,000</td>
<td>300,000</td>
</tr>
<tr>
<td>_ tourists visited X</td>
<td>68,000,000</td>
<td>32,400,000</td>
<td></td>
</tr>
</tbody>
</table>
Did it work?

Data:

● 16 relations, e.g. population, inflation rate, GDP, etc.
● 100K pages collected from the Web with Bing

Results:

● Claim identification precision: 60% over 7,092 claims
● Claim verification: found many pages with out-of-date statistics
Sources for or against a claim

Claim: Doctors confirmed the first case of death by genetically modified food


Resolved Added Mar 9

It originated on a fake news website and is therefore false. Emergent is as of now the only site to offer a full debunking.

Sources

Sources Tracked: 3 Total Shares: 62,188

For

2

Against

1

Shares

60,596

1,592
Results

- 300 claims from debunking website www.emergent.info (Silverman, 2015)
- Automated stance classification with 73% accuracy (Ferreira and Vlachos, 2016)
- Advisor to the Fake News Challenge with 50 participants (Rao and Pomerleau, 2017)
Coming soon: Fact Extraction and VERification

Claim:
The Rodney King riots took place in the most populous county in the USA.

Evidence:
[wiki/Los Angeles Riots]: The 1992 Los Angeles riots, also known as the Rodney King riots, were a series of riots, lootings, arsons, and civil disturbances that occurred in Los Angeles County, California in April and May 1992.

[wiki/Los Angeles County]: Los Angeles County, officially the County of Los Angeles, is the most populous county in the United States.

100K claims verified manually on Wikipedia (with James Thorne and Amazon R&D)
Beyond research

- Build a website where users can fact-check automatically the claims they want
- Users will be invited to give feedback to evaluate and improve our models
- Help (human) fact-checkers scale to the challenges of modern media
- Increase factual accountability in society
Vision

I taught my 5th-graders how to spot fake news. Now they won’t stop fact-checking me.

It’s more important than ever to teach kids how to read news critically.

Updated by Scott Bedell | Mar 29, 2017, 11:30am EDT