Computational fact-checking: Problems, state of the art and perspectives

Ioana Manolescu

CEDAR team, Inria Saclay and Ecole polytechique
http://pages.saclay.inria.fr/ioana.manolescu, @ioanamanol
Extraction et Gestion de Connaissances, Metz, 2019
MOTIVATION
Bad memories: Romania, 1989
Bad memories: Romania, 1989

Ceaușescu re-elected at the 14th congress!
Bad memories: Romania, 1989

Ceaușescu re-elected at the 14th congress!

He had held power since 1965.
Bad memories: Romania, 1989
Bad memories: Romania, 1985
Things get better
... kind of

1000 dead (approx.)
No one convicted.
Democratic societies crucially need the press

• To debate and express dissent

• To analyze, confirm or refute public statements

• To expose and explain society functioning
Democratic societies crucially need the press

- To debate and express dissent
- To analyze, confirm or refute public statements
- To expose and explain society functioning
DATA JOURNALISM AND JOURNALISTIC FACT-CHECKING
Data journalism

Investigative journalism based on complex and/or large data

http://abonnes.lemonde.fr/les-decodeurs/portfolio/2017/04/18/les-fractures-francaises-1-5-le-logement-les-raisons-de-la-crise_5112859_4355770.html
Data journalism

Panama Papers (International Consortium of Investigative Journalism, ICIJ)
Fact-checking (since 1930 approx.)

Fact-checking: verification of facts mentioned in media content
- To protect media reputation and avoid legal action

“The day I became a fact-checker at The New Yorker, I received one set of red pencils [...] for underlining passages on page proofs of articles that might contain checkable facts. [...] confirmed with the help of reference books from the magazine’s library”

Fact-checking (2012 – ongoing)

Not everyone agrees, however, that Democrats are not flip-flopping on the issue.
Mark Kriskorian, executive director of the Center for Immigration Studies, a think tank that advocates for lower immigration, said that because the public doesn’t know exactly what border barriers the Trump administration wants to build, Mulvaney’s statement is not an “exact” comparison. But, he said, to dismiss it simply on that basis would be “tendentiously literal.”

“The fact is that, other than the ‘Mexico will pay for it’ stuff, Trump is simply channeling the 2006 Secure Fence Act, and Schumer et al. who voted for it out of political calculation are indeed hypocrites for opposing the attempt to finally bring that law to fruition,” Krikorian told us via email.

At the surface level, it is true in a broad sense that Democrats, including Schumer, Obama and Clinton have in the past supported border fencing. All three voted for the Secure Fence Act of 2006, and all three supported the 2013 Senate immigration overhaul that passed the Senate, and which called for tougher border security including some additional fencing. But to claim that those measures are the same as what Trump is proposing is a stretch.

Mick Mulvaney
Director, Office of Management and Budget

"We don’t understand why the Democrats are so wholeheartedly against [President Trump’s border wall]. They voted for it in 2006."

Fox News Sunday – Sunday, April 23, 2017

Did Democrats Once Support Border Wall?
By Robert Farley  Posted on April 26, 2017

White House Office of Management and Budget Director Mick Mulvaney made an apples-to-oranges comparison when he said he couldn’t understand why Democrats opposed supplemental funding for a border wall since many of them were for it back in 2006.

Mulvaney is referring to the Secure Fence Act of 2006, which called for construction of 700 miles of fencing and enhanced surveillance technologies, such as unmanned drones, ground-based sensors, satellites, radar coverage and cameras. Sen. Chuck Schumer and Sen. Sam Brownback among a bipartisan majority that voted in favor of the legislation, and it was signed into law by President George W. Bush.

In a very general sense, the Democrats named by Mulvaney supported a bill to build more fencing, not a wall, and the Secure Fence Act was not signed into law by President Obama or Hillary Clinton, who was among the bipartisan majority that voted in favor of the legislation.

Do Democrats actually support a border wall?

No. That false claim was spread by a network of false news websites.
It's not just checking

• Most aspects of modern reality are complex
• Explaining can be as important and useful as checking
  – Helps also analyze the future
Libération *Désintox*: cost of saving Cyprus

Bobards en stock sur les plans de sauvetage européens

« Les Français vont devoir donner 2 à 3 milliards d'euros pour des banques à Chypre. D'un côté on supprime les infirmières, on surtaxe les PME en France [...] mais quand il s'agit de donner de l'argent de l'UE, c'est-à-dire les Français, à des banques pourries à Chypre, on le donne ».

Nicolas Dupont-Aignan, le 25 mars sur France Bleu

« C'est nous qui allons encore verser de l'argent puisque le MES s'est engagé à hauteur de dix milliards d'euros, dont deux milliards au titre de la France et des contribuables français ».

Florian Philippot, le 25 mars sur France Info

« Comment peut-on imposer aux Français, aux classes populaires et moyennes, aux petites entreprises, de nouvelles taxes, des taxes sur les sodas, une hausse de la CSG de 550 millions d'euros, et d'un autre côté alimenter de 15 milliards d'euros supplémentaires en Grèce l'inexécution de la zone euro ? On a même pensé à une taxe sur les parcs d'attraction ! »

Marine Le Pen, en septembre 2011, au moment du deuxième plan d'aide à la Grèce
Saving Cyprus: how much does it cost?

The article reads:
- The money is not given but lent
- The European Mechanism of Stability will lend $9bn
- Out of which France contributes 20% (approximately $2bn)

However, things are complicated because:
- The initial contribution of France to the EMS ($16 bn) counts toward French public debt
- The remainder contribution ($124) is not
Do we need to understand?

"Populism is telling people that there are simple answers to complex problems"

Salvini’s populism on the march as EU-backing north Italy turns on Brussels

Orban’s populism prospers by challenging EU taboos

Hungary’s prime minister leads the nationalist charge by challenging liberal taboos
Fact checking vs. fake news detection

- Fact checking is based on some **background information source**
  - Truth commonly agreed upon
- Fake news detection may or may not use a source
- E.g., text classifier (true, fake) trained with major news agency / fake content (often virulent style)
Most common fact-checking scenarios

• "What is the **value** of **metric** \( X \) in **space** \( Y \) at **time** \( T \)?"
  – \( X= \) youth unemployment, \( Y= \) Germany, \( T=2018 \)
  – \( X= \) illegal immigrants, \( Y= \) Italy, \( T=[2015-2018] \)
  – \( X= \) budget for research, \( Y= \) France, \( T=2019 \)
  – \( X= \) average monthly wages, \( Y= \) China, \( T=2018 \)

• **Comparison patterns**
  – \( X_1 \) against \( X_2 \); \( Y_1 \) against \( Y_2 \); \( T_1 \) against \( T_2 \); temporal trend etc.
Most common fact-checking scenarios

Deuxième trimestre mis à part, la France n'a pas la "croissance la plus faible de la zone euro"
Most common fact-checking scenarios

• "What did X say about Y?"
• "Is X related [in sense S] to Y?" [at time T]?
A CONTENT MANAGEMENT PERSPECTIVE
Lines of past and current research

1. **Model fact-checking** through a data and information management perspective

2. Identify **existing tools and techniques** which could be directly applied
   – In a special journalistic context (see next)

3. Devise **new models, tools and techniques** for fact-checking and data journalism problems
Projects and collaborations

• **Google Award** (2015) with X. Tannier (LIMSI)
• **ANR ContentCheck** (2016-2019) with X. Tannier (Sorbonne Université), S. Cazalens, P. Lamarre, J.-M. Petit, M. Plantevit (U. Lyon), F. Goasdoué (U. Rennes 1), Les Décodeurs (Le Monde)

[http://contentcheck.inria.fr](http://contentcheck.inria.fr)

• **Inria Associated Team WebClaimExplain** with AIST Japan (Julien Leblay)
• **Collaborations** with H. Galhardas (Technical U. Lisbon), former PhD S. Zampetakis and others
Fact-checking as a content management problem

Claim to be checked (text or data)

Human actors (journalists, experts, crowd workers)

Analysis result + proof
« True / rather true / rather false / false

See sources: http://dataref.com... »

Human actors

Media context

Reference information source 1
Reference information source 2
Reference information source n
Fact-checking as a content management problem

Claim to be checked (text or data)

Human actors (journalists, experts, crowd workers)

Reconciliation, reputation

Verification tool (query, match, source search...)

Analysis result + proof
« True / rather true / rather false / false

See sources: http://dataref.com... »

Source selection

Reference information source 1
Reference information source 2
... Reference information source n

Reference source construction, refinement, integration
Fact-checking as a content management problem

Claim to be checked (text or data)

Verification tool (query, match, source search...)

Source search / source selection

Human actors (journalists, experts, crowd workers)

Analysis result + proof

« True / rather true / rather false / false

See sources: http://dataref.com... »

Reference information source n+1

Reference source construction, refinement, integration

Media content

Claim extraction

Social network analysis

Reconciliation, reputation

Reference information source 1

Reference information source 2

Reference information source n
Fact-checking as a content management problem


Which existing technologies can be used?

• Database queries?
  – Journalists do not, generally, build databases.
  – "Not part of our job"
  – Persisting data is novel to some
  – SQL not user-friendly
  – Not always helped by journal information systems

• Remarkable exception: Ouest France
Which existing technologies can be used?

• **Database queries?**
  – Journalists do not, generally, build databases.
  – "Not part of our job"

• **Curse of the coverage**: they need to cover (almost) any topic

• **Curse of noteworthiness**: write about hot topics of today (or tomorrow)

• They work under **strong time pressure**
Which existing technologies can be used?

• **IR/NLP?**
  – Better options

• The outcome of fact-check has to be explainable
  • Not (only) "an ML algorithm said so"
  • But they don't like to train ML systems, either.

• Extremely **picky on data sources**
  – "Many Web sources claim it" is not an option
Automatic source selection: FactMinder


- Plug-in
- Brings relevant data from document and knowledge bases
- "Second screen"
Improving access to reference data sources


“Créations d’entreprises en France en 2015” → we return:
Keyword search across heterogeneous sources

A data model for facts, statement and beliefs

With F. Goasdoué and L. Duroyon (U. Rennes 1)

Modeling who said what when

- "On Jan 22, 2017, Le Canard Enchaîné wrote that François Fillon had stated on Jan 21 that Penelope worked for the NA from 2002 to 2008."
- What has one actor heard of?
- Who knew this at that time?
- Who reversed their position on a topic?
Summary of our work

• Automatically find most relevant reference data for a claim
  – RDF dump of full INSEE statistics
  – Table-aware full text search on the resulting graphs
  – Identifying statistic mentions in text (submitted)

• Identify connections between entities across heterogeneous content: ConnectionLens

• Modelling temporal facts, beliefs and statements
STATUS AND ROADMAP
Status

• Any **question answering framework** which can be plugged on trusted data sources is useful
  – Usually no time nor skills to integrate the data

• No IT infrastructure can be counted upon
  – More "a jumble of tools"

• ML promises a lot but delivery is hard(er)
  – Also: journalists ambivalence

• Strong and increasingly IT-literate community
  – E.g. [https://www.poynter.org/channels/fact-checking](https://www.poynter.org/channels/fact-checking)
Roadmap: computer science research

• DB, KR, IR, NLP
• In its most general statement, fact-checking supposes perfect NLP → study sub-problems!
• In fact-checking journalism, human writers chose topics, angle, style...
  – "A story wrapped around a query"
• Vision: build "perfect data machines" and give them to talented writers
A vision of journalistic dataspaces

- "Dataspaces": Franklin, Halevy and Maier, SIGMOD Record 2005
- Ingest **data of any nature**: structured (relational), semistructured (JSON, XML, [social] graphs), unstructured (text), KB...
- Storage, indexing
- Search across the data
  - Dong and Halevy [SIGMOD 2017]: kwd search, result can come from any data source
  - Bonaque, Cautis, Goasdoué, Manolescu [EDBT2016]: document search with social score component
  - ConnectionLens: find answers in any combination of data sources; "ad-hoc linked data"
Requirements for journalistic dataspaces

• Time
  – Of data acquisition
  – Of events described in the data

• Provenance
  – Authorship metadata
  – Annotation by users
  – Access control based on provenance and annotations

• Capacity to "derive" content (à la views)
• (Semi-automatic) semantic annotation and classification
• Social connections analysis
• Friendly interfaces
• Scalability
Roadmap: society

• Educate: the general audience, journalists, other social scientists
  – "Fake news creation" games, e.g. http://getbadnews.com
Roadmap: society

• Educate: the general audience, journalists, other social scientists
  – Education to media and the internet in schools
Is this worth it?

“Some people will never be convinced”

- “Facts have a liberal bias” (Paul Krugman) [https://www.nytimes.com/2017/12/08/opinion/facts-have-a-well-known-liberal-bias.html]

- "Scientists and humanity scholars believe in a constructed, logical discourse, and believe humans yield to reason. Businesspeople know this is not true, in general. Businesspeople have thus an advantage in winning political competitions." George Lakoff, former Berkeley professor [https://georgelakoff.com/2016/11/22/a-minority-president-why-the-polls-failed-and-what-the-majority-can-do/]

- Conspiracy theory adepts believe two obviously contradicting theories [Wood et al., 2012]
THANK YOU / QUESTIONS?

HTTP://CONTENTCHECK.INRIA.FR/