What do journalists really want from NLP researchers?

How to help build trust in media and democracy by helping journalists understand big data
NANCY WATZMAN

Director of Strategic Initiatives, Dot Connector Studio

Current clients: Knight Foundation – Knight Commission on Trust, Media and Democracy; Democracy Fund

Past: Internet Archive, Sunlight Foundation, Center for Responsive Politics, Center for Public Integrity, and more...
Fact:
- A fact is a statement that can be proven as true.
- At the library you can check out books to read.
- Ice cream is delicious!
- Baseball players use a glove to catch the ball.
- Flowers need water, sun, and food to grow.
- A cat is a mammal.

Opinion:
- Ice cream is pretty to look at.
- Walking is so much fun!
- Cats are cute and cuddly.
- Baseball is the best sport to play.
- Talking on the telephone is boring.
- The library is the most wonderful place to go!
ClaimBuster.

0.39 Cats are cute and cuddly.
0.35 Markers are the only things you should use to color pictures.
0.32 In music class we sing songs.
0.27 Talking on the telephone is boring.
0.27 Baseball players use a glove to catch the ball.
0.24 Flowers are the most beautiful plant.
0.23 Fire is hot.
0.21 Singing is so much fun.
0.21 Telephones are used to talk to others.
0.20 Markers can be used to color and draw.
0.20 Ice cream is delicious.
0.19 Baseball is the best sport to play.
0.19 The library is the most wonderful place to go.
0.19 A cat is a mammal.
0.15 At the library you can check out books to read.
0.13 Flowers need water, sun and food to grow.
I. HOW DID WE GET HERE?
The good old days
Americans’ Trust in Government: 1958–2017

% who trust the govt in Washington always or most of the time

Source: Pew Research Center
Trust in All Four Institutions Declines

Percent trust in the four institutions of government, business, media and NGOs, 2016 vs. 2017

Source: 2017 Edelman Trust Barometer Q11-625. Below is a list of institutions. For each one, please indicate how much you trust that institution to do what is right using a nine-point scale, where one means that you “do not trust them at all” and nine means that you “trust them a great deal.” (Top 4 Box, Trust) General Population, 28-country global total.
Trust in Media Plunges to All-Time Lows

Percent trust in media, and change from 2016 to 2017

Source: 2017 Edelman Trust Barometer, Ben Boyd presentation, NYC, October 12, 2017
Chapter 2 draft findings: What Happened to Trust?

- Large-scale global economic, social and political “shocks” have shaken people’s faith in a variety of democratic institutions including government.
- Poor performance in effectively safeguarding the public interest and delivering expected services erodes confidence in government.
- Growing political polarization contributes to a diminished willingness to support a shared political enterprise that necessarily operates through compromise.
- An increase in economic inequality and a declining belief in fairness of government weaken trust in the institution.
- Declining economic mobility diminishes faith in government as a mechanism to provide opportunities for all.
Chapter 3 draft findings: The New Media Landscape

- Exponential advancements in digital technology, coupled with explosive growth in broadband Internet and ubiquitous mobile access, have dramatically shifted how news and information are produced and consumed.

- The drastic decline in advertising revenues for print newspapers over the past two decades has challenged the viability of business models for news.

- Social media platforms connect users across the world, but have raised concerns about privacy, manipulation and foreign interference.

- The “public square” has become a 24-hour, continuously connected mobile experience supported by social networks, broadcast, and cable television.

- Explanations for the decline in the public’s trust of media include the proliferation of news sources, the blurring between news and opinion, media disintermediation and the spread of misinformation and disinformation.
Chapter 3 draft findings: The New Media Landscape
(Cont’d)

- Navigating the new media environment and separating truth from non-truth will continue to become more challenging as emerging technologies, such as virtual reality, bots and deepfakes, become more sophisticated.

- A confluence of sociological and economic circumstances, along with technological advancements, have created a “perfect storm” that is destabilizing citizen trust in media and other democratic institutions more generally.
II. How can NLP researchers help?
Many philanthropic efforts underway. A few:

- **Knight Foundation** Trust, Media and Democracy Initiative: [https://kng.ht/2M4OFPr](https://kng.ht/2M4OFPr)


- **Disinformation projects and funders database**: [http://bit.ly/2vHg3sq](http://bit.ly/2vHg3sq)
More useful links

- Must reading: **Witness**, “Deep fakes and synthetic media: What should we fear? What can we do?”
- **MisInfoCon on Medium**
- **Digital Forensic Research Lab**, Atlantic Council
- **Alliance for Securing Democracy** (Russian influence)
- **International Fact Checking Network/Poynter.**
- **Hoaxy & Bot-o-Meter**, Indiana University
7 COMMON FORMS OF INFORMATION DISORDER

- **SATIRE OR PARODY**: No intention to cause harm but has potential to fool.
- **MISLEADING CONTENT**: Misleading use of information to frame an issue or individual.
- **IMPOSTER CONTENT**: When genuine sources are impersonated.
- **FABRICATED CONTENT**: New content is 100% false, designed to deceive and do harm.
- **FALSE CONNECTION**: When headlines, visuals or captions don’t support the content.
- **FALSE CONTEXT**: When genuine content is shared with false contextual information.
- **MANIPULATED CONTENT**: When genuine information or imagery is manipulated to deceive.
TYPES OF INFORMATION DISORDER

FALSENESS  INTENT TO HARM

Misinformation
Unintentional mistakes such as inaccurate photo captions, dates, statistics, translations, or when satire is taken seriously.

Disinformation
Fabricated or deliberately manipulated audio/visual content. Intentionally created conspiracy theories or rumours.

Malinformation
Deliberate publication of private information for personal or corporate rather than public interest, such as revenge porn. Deliberate change of context, date or time of genuine content.
Jonathan Stray, Columbia University

“Making Artificial Intelligence Work for Investigative Journalism”

“What do Journalists do with Documents? Field Notes for Natural Language Processing Researchers”

https://journalism.columbia.edu/faculty/jonathan-stray
From Jonathan Stray, “Making Artificial Intelligence Work for Investigative Journalism”

- Document classification
- Language analysis
- Monitoring for breaking news
- Lead generation
“The Echo Chamber,” Reuters, Biskupic, Roberts, and Shiffman 2014:

https://www.reuters.com/investigates/special-report/scotus/

_Reuters reporters showed how a small group of elite lawyers have argued most of the cases before the U.S. Supreme Court (Biskupic, Roberts, and Shiffman 2014). The reporters also broke down the number of accepted cases by type, for example whether filed by a business, individual, or government agency. They accomplished this mostly by hiring 20 freelancers to read 10,300 petitions over a period of three months, but were able to gain some additional information through LDA topic modeling (Stray 2016a)._
“DHHS downplayed food stamp issues," (Dukes 2013)

https://www.wral.com/records-dhhs-downplayed-food-stamp-glitches/13173174/

...a WRAL reporter used Overview (M. Brehmer et al. 2014) to automatically cluster 4,500 pages of state government emails obtained through a Freedom of Information Act request. One large cluster corresponded to messages posted to an inter-county government email list. The reporter manually read this cluster and found messages showing the government officials knew that a web browser compatibility problem was causing delays, ultimately affecting 70,000 people. (Dukes 2014)

https://wapo.st/2vTCStd

For the Washington Post story “Whistleblowers say USAID’s IG removed critical details from public reports,” reporters compared 12 draft reports with their final versions. Using sentiment analysis, they found that more than 400 negative references were removed before publication (Stray 2016a).
IIII. Case study: Duke Reporters’ Lab

Tech & Check
Duke Tech & Check – matching claims with published fact-checks

- Take sentences from
  - Cable news shows (CNN, Fox, etc.)
  - YouTube videos and livestreams
  - Microphone audio

- Find related factchecks from a Share the Facts database of previously fact-checked claims

- This reduces to the well-studied natural language processing (NLP) problems of identifying paraphrases and entailments

- Current algorithm identifies words with similar meaning between the spoken sentence and sentence from the database
Bob the Algorithm

• Written by Duke students based on research on natural language processing ("Bob" is the algorithm’s code name)
• Identifies words with similar meaning in the spoken sentence and sentence from the database
• Challenge: the common NLP problem of identifying claims that are paraphrases of claims in the database, or ones that logically follow from a claim in the database (entailment).
Paraphrase matching

We found the algorithm is good at matching semantics
Paraphrase matching

The algorithm performs well even with differing sentence structures

Last year we lost 500 billion in trade with China

Trade deficit with China was half a trillion in 2017
Paraphrase matching difficulties

The algorithm is most frequently fooled by sentences which share many words but differ in words critical to overall meaning.
Entailment

The algorithm is also frequently fooled when the database sentence contains the right information as well as additional information irrelevant to the spoken sentence (in NLP this is referred to as entailment).
Present state

• Current algorithm is ~93% accurate in determining if two sentences are paraphrases of each other

• Current algorithm is only ~64% accurate in determining if one sentence implies the other (entailment)
Next steps

• Shift focus from pure paraphrase matching to paraphrase and entailment matching

• Use a database of over 433k sentence pairs called the Multi-Genre Natural Language Inference Corpus to train our model

• Ideally this will increase our accuracy on difficult statements without decreasing accuracy on simple paraphrases
IV. Case study: **TV News Archive**
Trending Yesterday

- The Papers: BBCNEWS: August 14, 2018
  Topic: ITALY

- The 11th Hour With Brian Williams: MSNBC: August 14, 2018
  Topic: TRUMP

  Topic: TURKEY

- Hardball With Chris Matthews: MSNBC: August 14, 2018
  Topic: OMAROSA

- The Last Word With Lawrence O’Donnell: August 14, 2018
  Topic: DONALD TRUMP

Software > Stanford Named Entity Recognizer (NER)

The Stanford Natural Language Processing Group
Political ads broadcast 375,350 times over 26 markets

The Political TV Ad Archive collected and, using innovative open source technology, tracked airings of political ads in key markets the 2016 election cycle.

The collection also linked ads to fact-checks by national fact-checking organizations. In addition to tracking airings across key primary states, the collection includes ads that may air elsewhere or exclusively on social media.

2016 political ads: politicaladarchive.org
Fact Checked

A collection of TV news clips that have been fact checked by FactCheck.org, PolitiFact, The Washington Post’s Fact Checker, or other fact checking organizations. Downloadable table with fact checks organized by topic, date, and fact check sources here.

https://archive.org/details/factchecked
Third Eye: API for TV News Archive chyrons

https://archive.org/services/third-eye.php
Opened Captions

https://github.com/slifty/opened-captions

Dan Schultz, Bad Idea Factory
V. Conclusion

1. There’s a lot to do.
2. We need more collaborations between NLP experts and journalists.
3. Let’s get to work.

Slides available here: http://bit.ly/2Mvq0nn
Contact me:

- Nancy Watzman
- @nwatzman
- nancy@dotconnectorstudio.com