FEVER shared Task

Tariq Alhindi
08/22/2018

Motivation

- 67% of consumers now look online for information before heading to a physical shop
- Yet, 61% of independent businesses, including restaurants, hairdressers, pharmacists and convenience shops, have inaccurate or missing opening hours listed on the web
- This is costing independent high street businesses £6.1 billion a year in lost revenue
- The UK Domain is urging businesses to check and take charge of their online information

MISINFORMATION ONLINE COSTS INDEPENDENT HIGH STREET BUSINESSES £6.1 BILLION A YEAR

2nd August 2017

https://www.nominet.uk/misinformation-online-costs-independent-high-street-businesses-6-1-billion-year/

Fake news is the promotion and propagation of news articles via social media. These articles are promoted in such a way that they appear to be spread by other users, as opposed to being paid-for advertising. The news stories distributed are designed to influence or manipulate users' opinions on a certain topic towards certain objectives.

Service	Price
Make a video appear in YouTube's main page for two minutes	RUB 35,000 ($621)
Make 20 videos trend on YouTube's main page for two minutes (up to six minutes, upon request)	RUB 450,000 ($7,992)
VIP services for social media	RUB 5,990–6,990 ($106–$124)
Press release distribution to news outlets	RUB 45,200 ($802)
10,000 site visitors	RUB 1,000 ($17)
100 dislikes on a YouTube video	RUB 100 ($1.7)
100 comments on a YouTube video	RUB 150 ($2.6)
1,000 group joins	RUB 650 (11)

https://ijnet.org/en/blog/real-news-about-fake-news-real-cost-spreading-misinformation

FEVER

The First Workshop on Fact Extraction and Verification

Join us at EMNLP on 1st November 2018

Overview

- FEVER: Fact Extraction and VERification of 185,445 claims
- Dataset
  - Claim Generation
  - Claim Labeling
- Systems
  - Baseline
    - Document Retrieval
    - Sentence Selection
    - Textual Entailment
  - Our System

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

[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 USA.

Verdict: Supported

Claim Generation

- Sample sentences from the introductory section of 50,000 popular pages (5,000 of Wikipedia's most accessed pages and their linked pages)

- **Task:** given a sample sentence, generate a set of claims containing a single piece of information focusing on the entity that its original Wikipedia page was about.
  - Entities: a dictionary of terms with wikipedia pages.
  - Create mutations of the claims.
  - Average claim length is 9.4 tokens

Claim Labeling

- In 31.75% of the claims more than one sentence was considered appropriate evidence
- Claims require composition of evidence from multiple sentences in 16.82% of cases.
- In 12.15% of the claims, this evidence was taken from multiple pages.
- IAA in evidence retrieval 95.42% precision and 72.36% recall.

Baseline System

- Document Retrieval: DrQA → returns k nearest document for a query using cosine similarity
- Sentence Selection: using TF-IDF similarity to the claim (above a certain threshold)
- RTE (with and without sentence selection)
  - MLP
  - DA
  - Note: RTE for NOTENOUGHINFO uses NEARST_P or RANDOM_S

Dataset size

Split	SUPPORTED	REFUTED	NEI
Training	80,035	29,775	35,639
Dev	3,333	3,333	3,333
Test	3,333	3,333	3,333
Reserved	6,666	6,666	6,666

Table 1: Dataset split sizes for SUPPORTED, RE-FUTED and NOTENOUGHINFO (NEI) classes

Document Ret.

k	Fully Supported (%)	Oracle Accuracy (%)
1	25.31	50.21
5	55.30	70.20
10	65.86	77.24
25	75.92	83.95
50	82.49	90.13
100	86.59	91.06

Table 2: Dev. set document retrieval evaluat

Results

Model	Accuracy (%)		
	NEARESTP	RANDOMS	SNLI
MLP	65.13	73.81	-
DA	80.82	88.00	38.54

Table 3: Oracle classification on claims in the development set using gold sentences as evidence

Model	Oracle Accuracy (%)	
	NEARESTP	RANDOMS
MLP	57.16	73.36
DA	63.68	69.05

Table 5: Oracle accuracy on claims in the dev. set using gold documents as evidence (c.f. Table 3).

Model	Accuracy (%)	
	NoScoreEv	ScoreEv
MLP / NP	41.86	19.04
MLP / RS	40.63	19.42
DA / NP	52.09	**32.57**
DA / RS	50.37	23.53

Table 4: Full pipeline results on development set

Model	Accuracy (%)	
	NEARESTP	RANDOMS
MLP	38.85	40.45
DA	41.57	40.62

Table 6: Pipeline accuracy on the dev. set without the sentence selection module (c.f. Table 4).

Our System

Tariq Alhindi
08/22/2018

Document Retrieval

- Google Custom Search API
  top 2 results of "Wikipedia" + claim

- Named Entity Recognition (NER)
  Pretrained BiLSTM of (Peters et al)

- Dependency Tree

- Combined Method

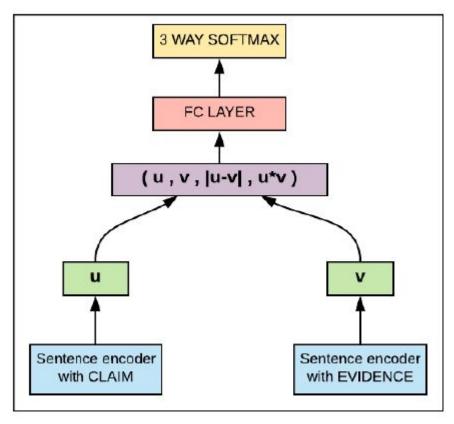| Method	Avg k	Coverage
Google API	2	79.5%
NER	2	77.1%
Dependency Parse	1	80.0%
Combined	3	**94.4%**
(Thorne et al., 2018)	5	55.3%

Matthew E. Peters, Waleed Ammar, Chandra Bhagavatula, and Russell Power. 2017. Semi-supervised sequence tagging with bidirectional language models. In ACL.

Sentence Selection

- Extract Top 5 Evidence from at most 3 documents
    - using TFIDF similarity
    - Evidence recall   78.4   (baseline system: 45.05)

- Top 5 evidence include a lot of wrong evidence! (Most gold has one or two evidence sentences)

- Only top 3 evidence sentences were used for entailment
    - using cosine similarity of ELMO-embeddings of claim and evidence

Textual Entailment

Alexis Conneau, Douwe Kiela, Holger Schwenk, Loïc Barrault, and Antoine Bordes. 2017.
Supervised learning of universal sentence representations from natural language inference data. EMNLP

Results

Data	Pipeline	FEVER
DEV	(Thorne et al., 2018)	31.27
	Ours	**50.83**

	S	N	R
S	**4635**	1345	686
N	2211	**3269**	1186
R	1348	1470	**3848**

Table 7: Confusion matrix of entailment predictions

Error Analysis

Fox 2000 Pictures released the film Soul Food	0.29
Soul Food is a 1997 American comedy-drama film produced by Kenneth "Babyface" Edmonds ,Tracey Edmonds and Robert Teitel and released by Fox 2000 Pictures	

Table 4: Cosine similarity between claim and supporting evidence

At seventeen or eighteen years of age, he joined Plato's Academy in Athens and remained there until the age of thirty-seven (c. 347 BC)
Shortly after Plato died , Aristotle left Athens and at the request of Philip II of Macedon ,tutored Alexander the Great beginning in 343 BC

Table 5: The top evidence is selected by Annotators and the bottom evidence by our pipeline

Error Analysis

Claim: Natural Born Killers was directed by Oliver Stone
Evidence: Natural Born Killers is a 1994 American satirical crime film directed by Oliver Stone and starring Woody Harrelson , Juliette Lewis , Robert Downey Jr. , Tom Sizemore , and Tommy Lee Jones .
Claim: Anne Rice was born in New Jersey
Evidence: Born in New Orleans, Rice spent much of her early life there before moving to Texas, and later to San Francisco

Table 6: Wrong gold label (NOT ENOUGH INFO)

Claim: Happiness in Slavery is a gospel song by Nine Inch Nails
Evidence: Happiness in Slavery,is a song by American industrial rock band Nine Inch Nails from their debut extended play (EP), Broken(1992)

Table 8: Example where our model predicts SUPPORTS for a claim labeled as NOT ENOUGH INFO

Thanks