

Misinformation Classification Using BERT models

Kris Riedman

Overview

- Introduction
- Literature review
- Experiment Design
- Results
- Conclusion

Introduction

- Misinformation classification of news articles
- Optimize for efficiency: can a smaller model be as accurate as the baseline
- Why is this important?

2020 Election and January 6th



Covid Misinformation



Literature Review

- PANACEA: An Automated Misinformation Detection System on COVID-19
 - Online tool that provides automated veracity assessment
 - Complicated results
- Neural Architectures for Fine-Grained Propaganda Detection in News
 - Sentence and fragment level analysis
 - May not scale to entire news articles

Literature Review Cont.

- Fine-Grained Analysis of Propaganda in News Articles
 - Identify spans of propaganda embedded in text
 - 18 different propaganda types
- Propppy: Organizing the News Based on Their Propagandistic Content
 - Maximum entropy text classifier for 4-way and 2-way classification
 - Does not use transformer architecture

Experiment Design

- Misinformation & Fake News text dataset
 - 35k true articles from sources such as Reuters, New York Times, and Washington Post
 - 44k misinformation articles from sources such as Redflag Newsdesk, Breitbart, Truth Broadcast Network but also data collected by the EUvsDisinfo project
 - 6k false articles from Russian national news

'True' Example

“The head of a conservative Republican faction in the U.S. Congress, who voted this month for a huge expansion of the national debt to pay for tax cuts, called himself a “fiscal conservative” on Sunday and urged budget restraint in 2018. In keeping with a sharp pivot underway among Republicans, U.S. Representative Mark Meadows, speaking on CBS’ “Face the Nation,” drew a hard line on federal spending, which lawmakers are bracing to do battle over in January.”

Misinformation Example

“Donald Trump just couldn’t wish all Americans a Happy New Year and leave it at that. Instead, he had to give a shout out to his enemies, haters and the very dishonest fake news media. The former reality show star had just one job to do and he couldn’t do it. As our Country rapidly grows stronger and smarter, I want to wish all of my friends, supporters, enemies, haters, and even the very dishonest Fake News Media, a Happy and Healthy New Year, President Angry Pants tweeted. 2018 will be a great year for America! As our Country rapidly grows stronger and smarter, I want to wish all of my friends, supporters, enemies, haters, and even the very dishonest Fake News Media, a Happy and Healthy New Year. 2018 will be a great year for America! Donald J. Trump”

Experiment Design Cont.

- Optimize three different models
 - Bert base cased (baseline), DistilRoberta, DistilBert
- 4 different experiments; 12 models in total
 - Max Length 200, Batch Size 4 & Batch Size 8
 - Max Length 500, Batch Size 4 & Batch Size 8
- Fine tune on 10k samples 5k from true and false dataset
- Evaluate on 15k samples 5k from each dataset (holdout)

Results

	200 Max length		500 Max length	
	Batch Size 4	Batch Size 8	Batch Size 4	Batch Size 8
Bert	0.53	0.17	0.58	0.99
DistilRoberta	0.97	0.17	0.84	1.00
DistilBert	0.48	0.29	0.94	0.97

Highlights

- DistilRoberta max length 200/batch size 4
 - Extended and expanded pre-training on DistilRoberta compared to other Bert models
- DistilBert and DistilRoberta max length 500/batch size 4
 - More efficient are able to take advantage of the smaller batch size
- DistilBert and the baseline model larger max length size improved performance
 - Similar architecture to the Bert base model
 - Enables the models to more accurately learn from each training example

Notable failures

- Too short
 - “The migration wave is part of preparations for the New World Order, says analyst David Icke.”
 - “Polish policy towards Russia is entirely based on the support of the United States.”
- Foreign language
 - “españa , almirante kuznetsov , ceuta , portaaviones Imagen del portaaviones Almirante Kuznetsov. Fuente:Reuters La ruta y la misión del portaaviones Almirante Kuznetsov, según el secretario de prensa del presidente de la Federación de Rusia, Dmitri Peskov, se encuentran en un sobre cerrado en el que se lee “Completamente secreto”.
- Artifacts from data collection
 - “***Next Swipe left/right*** Koala gets so excited it runs head-first into a tree Mason, a baby Koala at the Port Stephens Koala sanctuary in Australia, gets so excited he runs head-first into a tree.”

Notable failures Cont.

- ‘True’ article that contains misinformation as a quote
 - “The Turkish government’s spokesman on Wednesday said ***that the United States decision to recognize Jerusalem as the capital of Israel will plunge the region and the world into a fire with no end in sight . Declaring Jerusalem a capital is disregarding history and the truths in the region, it is a big injustice/cruelty, shortsightedness, foolishness/madness, it is plunging the region and the world into a fire with no end in sight***, Deputy Prime Minister Bekir Bozdag said on Twitter. I call on everyone to act logically, respect the agreements they signed and behave reasonably, avoid risking world peace for domestic politics or other reasons, he said. U.S. officials have said President Donald Trump is likely to give a speech on Wednesday unilaterally recognizing Jerusalem as Israel’s capital, a step that would break with decades of U.S. policy.”

Conclusion

- Smaller models can have similar performance to larger models
- Next steps:
 - Add on Lora to reduce model size further
 - Experiment with other hyperparameters
 - Run the experiment on a different dataset

Misinformation Classification Using BERT models

Questions?