Automated Evidence Collection for Fake News Detection

ICON 2021 : 18th International Conference on Natural Language Processing

Mrinal Rawat
rawatmrinal06@gmail.com

Diptesh Kanojia
d.kanojia@surrey.ac.uk

Content

- Motivation
- Introduction
- Dataset
- Approach
- Experiments
- Results & Discussion
- Conclusions & Future Work

Motivation

- Quality on social media is significantly affected due to the spread of fake news, misinformation and unverifiable facts.

- Panic due to Fake News in the epidemic situation like COVID-19.

Motivation

- Quality on social media is significantly affected due to the spread of fake news, misinformation and unverifiable facts.
- Panic due to Fake News in the epidemic situation like COVID-19.
- Websites such as Poynter, Factcheck, etc. currently rely on methods involving manual detection of fake news which is a cumbersome and challenging task.
- Automatic Fake News detection aims to mitigate the problem of misinformation with the help of evidence supported by multiple sources.

Motivation

- Quality on social media is significantly affected due to the spread of fake news, misinformation and unverifiable facts.
- Panic due to Fake News in the epidemic situation like COVID-19.
- Websites such as Poynter, Factcheck, etc. currently rely on methods involving manual detection of fake news which is a cumbersome and challenging task.
- Automatic Fake News detection aims to mitigate the problem of misinformation with the help of evidence supported by multiple sources.

Screenshot from PolitiFact (https://www.politifact.com/factchecks/2020/mar/25/facebookposts/water-boiled-orange-peels-and-cayenne-pepper-will-/)

Introduction

- Previous approaches rely on classical machine learning models (Vlachos, and Riedel, 2014) and deep learning based approaches (Malon, 2018; Vijjali et al.,2020) on LIAR and FEVER datasets.

- Previous work on CONSTRAINT-2021 dataset (Patwa et al, 2020) does not include evidence since the dataset only have claim and label pairs.

Introduction

- Previous approaches rely on classical machine learning models (Vlachos, and Riedel, 2014) and deep learning based approaches (Malon, 2018; Vijjali et al.,2020) on LIAR and FEVER datasets.

- Previous work on CONSTRAINT-2021 dataset (Patwa et al, 2020) does not include evidence since the dataset only have claim and label pairs.

- We address these issues and propose a novel approach that automatically collects the evidence from multiple sources.

- We incorporate a summarization component that helps outperforms the state-of-the-art approaches on CONSTRAINT-2021 shared task achieving a F1-score of 0.9925.

Dataset

We use the pre-released COVID-19 fake news dataset as a part of the CONSTRAINT- 2021 shared task (Patwa et al., 2020). Each post or tweet contains content in the English language and is classified as:

- **Real:** where tweets or articles which are factually correct and verified from authentic sources
- **Fake:** where tweets or posts related to COVID-19 which are factually incorrect and verified as false.

Dataset

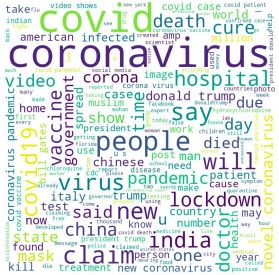
We use the pre-released COVID-19 fake news dataset as a part of the CONSTRAINT- 2021 shared task (Patwa et al., 2020). Each post or tweet contains content in the English language and is classified as:

- **Real:** where tweets or articles which are factually correct and verified from authentic sources
- **Fake:** where tweets or posts related to COVID-19 which are factually incorrect and verified as false.

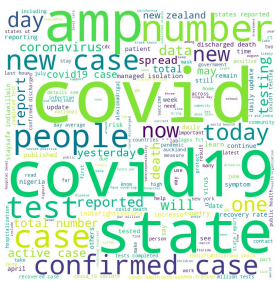
Split	Real	Fake	Total
Training	3360	3060	6420
Validation	1120	1020	2140
Test	1120	1020	2140
Total	5600	5100	10700

Label	Source	Text
Fake	Facebook	If you take Crocin thrice a day you are safe
Real	Twitter	ICMR has approved 1000 COVID19 testing labs all across India. #IndiaFightsCorona #ICMRFightsCovid19

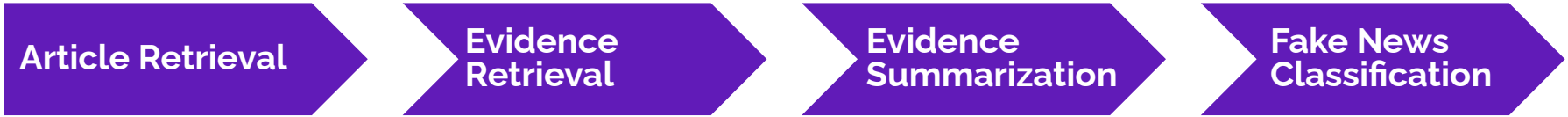
Fake



Real

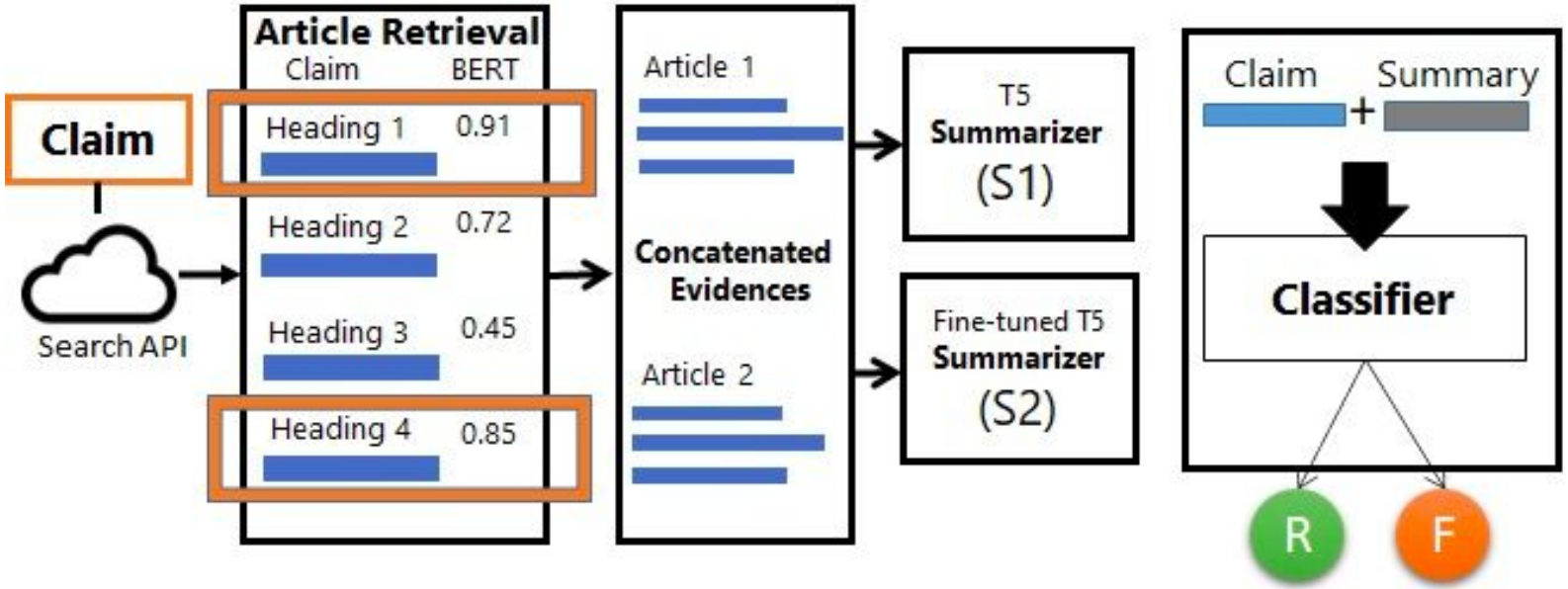


Approach


Article Retrieval
Evidence Retrieval
Evidence Summarization
Fake News Classification

Approach

Article Retrieval → Evidence Retrieval → Evidence Summarization → Fake News Classification



[CLS]	If	you	tested	...	[SEP]	SOURCES:	https://cdc.gov	[SEP]	Staying	away	...

Approach (Cont'd)

Evidence Collection

The original dataset of COVID-19, the evidence is not released along with the claim. Hence, we extract the evidence which assists the classifier.

- **Article Retrieval:** For each claim c, we perform a Google search and use BERT to get the similarity score of response text w.r.t. to the input claim. We select top 3 results which has the similarity score greater than 0.7.

- **Sentence (Evidence) Retrieval:** We extract the relevant article URLs U = (u1, u2, u3) and employ a similar method to find the top k sentences within each article.

Approach (Cont'd)

Evidence Collection

The original dataset of COVID-19, the evidence is not released along with the claim. Hence, we extract the evidence which assists the classifier.

- **Article Retrieval:** For each claim c, we perform a Google search and use BERT to get the similarity score of response text w.r.t. to the input claim. We select top 3 results which has the similarity score greater than 0.7.
- **Sentence (Evidence) Retrieval:** We extract the relevant article URLs U = (u1, u2, u3) and employ a similar method to find the top k sentences within each article.

Evidence Summarization

We summarize the evidence collected in previous step using following two ways: •

- **Google T5 (S1):** We use the Google T5-large (Raffel et al., 2019) model which is the current state-of-the-art for the summarization tasks.
- **Fine-tuning on FEVER Dataset (S2):** We first fine-tune the Google T5 on FEVER dataset which has ground truth claim and evidence, and then use it to summarize the evidence collected using our approach.

Approach (Cont'd)

Claim	Evidence	Summarization-1 (S1)	Summarization-2 (S2)
There is no evidence that children have died because of a COVID-19 vaccine. No vaccine currently in development has been approved for widespread public use. https://t.co/9ecvMR8SAf	Currently there is no coronavirus vaccine that has been approved for the American public. And there is no evidence that children have died because they received one of the COVID-19 vaccines being developed. PolitiFact found no evidence that anyone has died from complications related to a trial COVID-19 vaccination. There is no evidence that children have died because of a COVID-19 vaccine.	There is no evidence that children have died because they received a COVID-19 vaccine. No evidence that anyone has died from complications related to a trial COVID-19.	There is no evidence that children have died because they received one of the COVID-19 vaccines being developed. PolitiFact found no evidence that anyone has died from complications related to a trial COVID-19 vaccination.

Experiments

We evaluate the COVID-19 fake news detection using on following three methods:

- No evidence
- Summarized evidence without fine-tuning
- Summarized evidence with fine-tuning

The evaluation metrics used are as following:

Accuracy, Precision, Recall and F1-Score

Models Used:

- Machine Learning Models:

  Logistic Regression, Support Vector Machines, Random Forest Classifier

- Deep Learning Models:

  LSTM, BERT, RoBERTa, XLNet

Results & Discussion

	Previous Approaches		Our Approach w/ various Classification methods								
	Chen et al. (2021)	Li et al. (2021)	Logistic Regression			SVM			LSTM		
			-	S1	S2	-	S1	S2	-	S1	S2
Precision	0.9902	0.986	0.9531	0.9565	0.9701	0.9641	0.9671	0.9764	0.9589	0.9598	0.9612
Recall	0.9901	0.985	0.9531	0.9564	0.9700	0.9639	0.9668	0.9761	0.9584	0.9596	0.9612
F-Score	0.9901	0.985	0.9531	0.9565	0.9700	0.9639	0.9668	0.9761	0.9584	0.9596	0.9612

Results obtained after the fake news classification task where the values for previous approaches are from the latest shared task results and the results for each iteration of our approach are shown [P (Precision), R (Recall), and F (F-Score)]. (-) −→ No Evidence, S1 −→ Summarization-1 as Evidence, S2 −→ Summarization-2 as Evidence.

Results & Discussion (Cont'd)

	Previous Approaches		Our Approach w/ various Deep Learning Classification methods								
	Chen et al. (2021)	Li et al. (2021)	$BERT_{base}$			$RoBERTa_{base}$			$XLNet_{base}$		
			-	S1	S2	-	S1	S2	-	S1	S2
Precision	0.9902	0.986	0.9612	0.9916	0.9917	0.9918	0.9929	0.9922	0.9920	0.9934	0.9947
Recall	0.9901	0.985	0.9864	0.9888	0.9897	0.9897	0.9911	0.9916	0.9892	0.9911	0.9925
F-Score	0.9901	0.985	0.9858	0.9888	0.9893	0.9893	0.9908	**0.9908**	0.9892	0.9910	**0.9925**

Results obtained after the fake news classification task where the results for each iteration of our approach with various deep learning classification methods are shown [P (Precision), R (Recall), and F (F-Score)]. (-) −→ No Evidence, S1 −→ Summarization-1 as Evidence, S2 −→ Summarization-2 as Evidence.

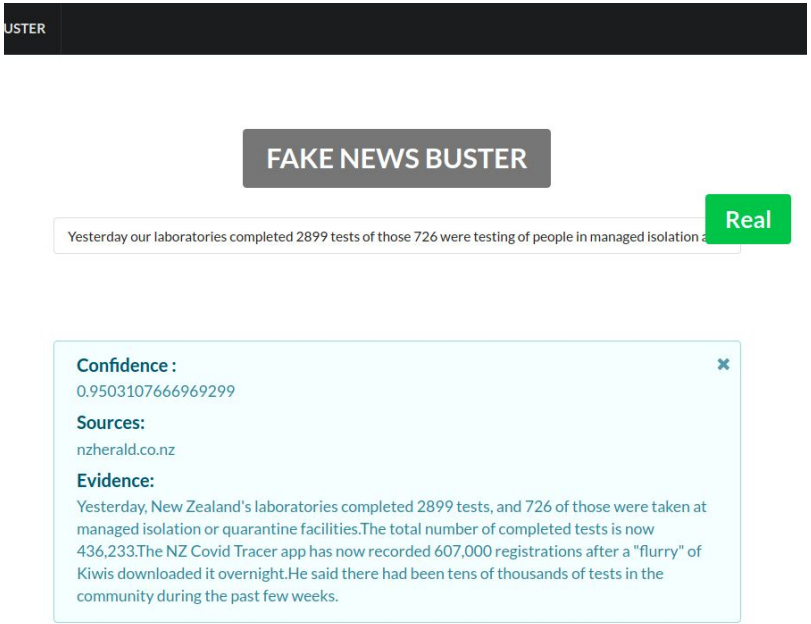
Conclusions & Future Work

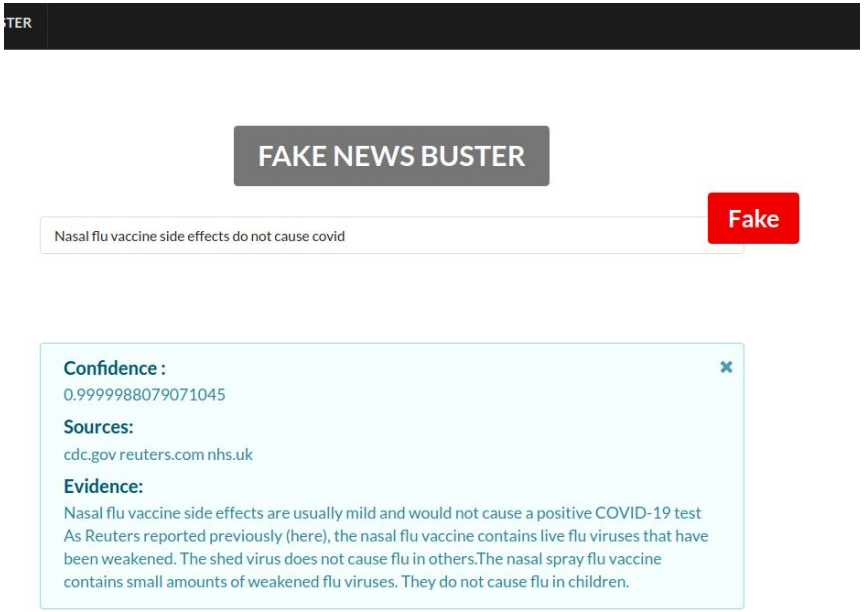
- We present an automated method to collect the evidence and summarize it for the fake news detection task.
- Our approach helps in augmenting the dataset released in the CONSTRAINT-2021 task.
- Our systematic framework achieves a F1- score of 0.9925.

Conclusions & Future Work

- We present an automated method to collect the evidence and summarize it for the fake news detection task.
- Our approach helps in augmenting the dataset released in the CONSTRAINT-2021 task.
- Our systematic framework achieves a F1- score of 0.9925.
- We show that a summarization module that can help collect evidence more effectively.
- In future, we want to experiment and reproduce the results on other fact verification tasks.

Conclusions & Future Work (Cont'd)





Source Code: https://github.com/rawat-mrinal06/fake_news/

THANK YOU