FAKE NEWS

Zero Shot Fake News Detection

Group no 18

Asmita Mukherjee
Hanzalah Firdausi
Harsh Vardhan Bhadauriya
Shreyansh Jain
Sehban Fazili

Motivation

- The rise of social media has amplified the potential impact of misinformation.
- There has been significant investment, totaling over $300 million, from major media and tech organizations to combat fake news, along with support from various activists, NGOs, and interest groups.
- The prevalence of fake news and disinformation continues to grow, affecting millions of individuals globally, highlighting the necessity of implementing a solution.
- The persistent spread of inaccurate information calls for a technology-based approach to eradicate it from the internet.
- To address this issue, we introduce the FacQA tool which checks the trustworthiness of the news articles.

Problem Statement

- The objective of this project is to create a tool for checking whether news articles are trustworthy or not.
- The tool will have to perform an evaluation of the facts, events, and related information within a news article.

Literature Review

The field of fake news detection is expanding rapidly, and a plethora of research has been conducted on this topic. Notable works in this area include:

- "Fake News Detection on Social Media: A Data Mining Perspective" by Shu et al. (2017): This paper proposes a framework for detecting fake news on social media using data mining techniques.
- "Combating Fake News: A Survey on Identification and Mitigation Techniques" by Zubiaga et al. (2018): This survey paper provides an overview of various techniques that have been proposed for detecting and mitigating fake news.
- "Detection of Fake News in Social Media: A Machine Learning Approach" by Babalola et al. (2019): This paper presents a machine learning-based approach for detecting fake news on social media.
- "A Survey on Fake News: Detection, Analysis, and Mitigation Techniques" by Singh et al. (2020): This survey paper provides an in-depth analysis of various detection, analysis, and mitigation techniques for fake news.
- "Fake News Detection on Social Media using Geometric Deep Learning" by Pham et al. (2021): This paper proposes a novel approach for detecting fake news on social media using geometric deep learning.

Baseline Results

TF-IDF vectorization with ML classifier

Method	Train F1-score	Test F1-score
Logistic-regression	0.86	0.20
KNN	0.91	0.25
Decision Tree	1.00	0.26
SVM	0.89	0.19

BERT embeddings with ML classifier

Method	Train F1-score	Test F1-score
Logistic-regression	0.95	0.22
KNN	0.95	0.31
Decision Tree	1.00	0.27
SVM	0.94	0.19

Baseline Results

BERT Classifier

Method	Train F1-score	Test F1-score
Distilbert-classifier	1.00	0.19

Fine-Tuned BERT Classifier

Method	Train F1-score	Test F1-score
Fine-Tuned Distilbert-classifier	0.32	0.35

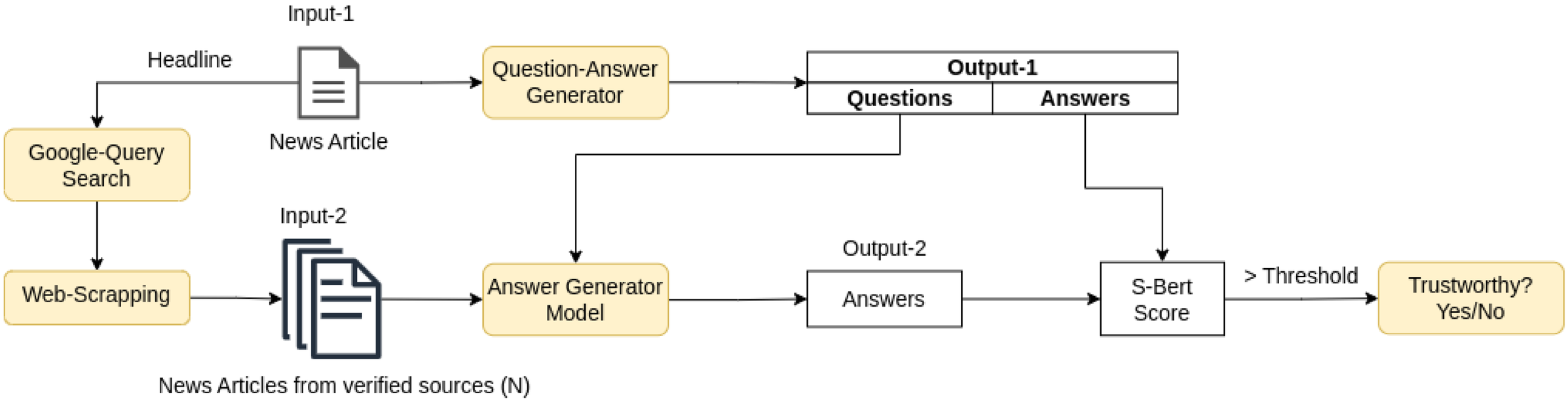
Novelty

- Since the method is zero-shot , it can classify news article almost instantaneously, while other machine learning and deep learning methods require additional training time.

- Zero-shot method performs really well on unseen classes.

- As seen before in our baselines , methods which require training work well when the test set is from similar domain, however since our method relies on question answering, it generalizes across different domains.

Model Architecture/Pipeline



Proposed Solution

- We propose a two step solution:
  - **Question-Answer generation**
  - **Answer generation task**
- The Question Answer Generator will be utilized to generate questions and answers from a given input news article.
- The verified news articles will be retrieved by conducting a reverse text search on credible and reputed news websites, focusing on a specific topic.
- The pre-trained answer generation model will then be used to generate answers to the questions generated by the QA Generator, using the trustworthy news article as context.
- The answers from the input news article will be compared with the answers from the credible sources using evaluation metrics.
- Based on the results of this comparison, a determination will be made as to whether the news article is trustworthy or not, using a set threshold.

Question-Answer Generator

The QA Generator is used to produce questions and answers from input news articles.

The Question and Answers are generated using an answer-aware question generation model, which has been fine-tuned using the t5-base model.

Answer Generation Model

A DistilBERT model, fine-tuned on the SQuAD dataset, is employed to generate answers to the questions using the verified news articles as context.

We give the questions generated from the QA generator on the input news article as an input and the verified news articles as context to the answer generation model to generate answers.

How Question-Answers are generated?

We use one such example using DistilBert Question-Answer generator model.

Input Text - While on the campaign trail, Donald Trump promised to revive the coal industry and after he took power, he signed an executive order rolling back a temporary ban on mining coal and a stream protection rule that was imposed by the Obama administration.

Question-Answer Pair: { 'question': "What did Donald Trump's executive order roll back after he took power?", 'answer': 'a temporary ban on mining coal and a stream protection rule that was imposed by the Obama administration'},

Evaluation Metrics

The metrics that can be used to evaluate the similarity between the answers of the input news article and the verified news article are as follows:

Bleu score

Rouge score

SentBERT

The threshold can be set by experimenting with the model.

Do you have **any questions?**