Identifying Deceptive Content: A Study on Clickbait and Fake News Detection

Shashank Rangarajan, Chia-Yu Tung, Rishabh Ghosh, Michael Guastalla, Edwin Wang
Department of Computer Science, University of Southern California
{sr87317, ctung, ghoshris, guastall, kuanchun}@usc.edu

1 Overview

Social media has become a popular platform for sharing views and news due to the availability of digital devices, affordable internet, and free subscriptions. However, studies show that less credible individuals are more likely to propagate fake news and clickbait for personal gain. As a result, detecting fake news and clickbait on social media has become a crucial research topic, with various methods based on sources, transmission, styles, and knowledge. Natural Language Processing (NLP) is essential for detecting fake news and clickbait due to the large volume of social media content, which is primarily text-based. Automated tools based on NLP can analyze the underlying meaning and intent of the text, enabling the identification of patterns in deceptive content. Successful implementation can detect and flag malicious content before it harms user experience. This project aims to improve such existing state-of-the-art NLP methods and explore new modalities for more accurate detection and classification of fake news and clickbait.

1.1 Objectives

Objectives of our projects can be summarized as follows:

1. Implement existing state-of-the-art models for clickbait and fake news detection.

2. Compare the performance of the state-of-the-art models on various publicly available datasets and study their results.

3. Explore the relationship between fake news and clickbait to enable more fine-grained prediction in the models.

4. Investigate other modalities, such as image, audio, and video, to improve clickbait and fake news detection.

2 Related work

To achieve the goals of our project, we conducted an extensive literature review on each sub-goal. Our search focused on the Webis Clickbait Corpus dataset (Challenge, 2017), and we discovered that (Papadopoulou et al., 2017) proposed a two-level classification approach that combined 65 first-level classifiers in a second-level feature vector, and (Qiu et al., 2021) introduced Click-BERT, which effectively identified clickbaits using pre-training methods such as BERT and Longformer and recurrent neural networks such as BiLSTM and Bi-GRU with a parallel model structure.

Our research on clickbait detection and BERT models also led us to discover that BERT is a popular model used for fake news detection. In a paper titled "r/Fakeddit" (Nakamura et al., 2019), the authors present a multimodal dataset consisting of over 1 million samples from multiple categories of fake news and propose hybrid text+image models, highlighting the importance of the novel aspect of multimodality and fine-grained classification unique to Fakeddit. This inspired us to explore two avenues: fine-grained predictions within the clickbait and fake news domain, and leveraging multiple modalities in clickbait detection.

For the first avenue, we could not find models with a fine-grained prediction as to whether a clickbait can be classified as fake news or not, but we plan to use (Corpus, 2017) for our explorations. For the second avenue, we found that most work is done on the YouTube platform. (Xie et al., 2021a) emulates the user's perspective by relying only on the thumbnail image and title of the video, while (Zannettou et al., 2018a) uses metadata such as comments, tags, and video statistics. Additionally, (Gamage et al., 2021a) builds on (Zannettou et al., 2018a) by also incorporating the audio transcript of the video into its detection methods, consulting six different models.

3 Datasets

Our project aims to detect fake news and clickbait using publicly available datasets. However, while there are numerous datasets available for fake news detection, there is a limited availability of datasets for clickbait detection. To overcome this challenge, we plan to use multiple fake news datasets to fine-tune our model and evaluate its effectiveness in detecting both fake news and clickbait.

We plan to use the LIAR dataset (Wang, 2017) that has labeled statements with corresponding truthfulness labels extracted from news sources, but we recognize the dataset's political bias, which may affect the results. To mitigate this potential issue, we will incorporate the CREDBANK-data (Mitra, 2015), consisting of labeled statements from Twitter with a sample size of approximately 60 million. We also intend to use other datasets like the Fake News Corpus dataset (Corpus, 2017) and BuzzFeedNews dataset (BuzzFeedNews, 2017).

For clickbait detection, we will use the (Dataset, 2017), which comprises headlines from various news websites labeled as clickbait or not. Moreover, we will employ the clickbait-detector dataset (Mathur, 2017), consisting of clickbait headlines fetched from BuzzFeed, NewsWeek, The Times of India, and The Huffington Post. The genuine/non-clickbait headlines were retrieved from The Hindu, The Guardian, The Economist, TechCrunch, The Wall Street Journal, National Geographic, and The Indian Express.

These datasets only contain text and labels, which may be binary or multiple. To ensure consistency and remove any unnecessary noise, we will preprocess the data by removing stopwords, stemming, and converting the text to lowercase (except the headlines). Additionally, we may modify the labels to fit our problem requirements, such as converting them to a binary classification problem. This preprocessing will help in reducing the complexity of the dataset and improve the performance of the model.

We have identified a list of YouTube channels to be scraped from (Zannettou et al., 2018a) for clickbait detection, but we have yet to find a publicly available dataset that includes audio, video, image, together with text data for detecting fake news and clickbait. While we will attempt to explore this option, it's important to note that our ability to conduct a successful multimodal study may be hindered by the unavailability of such datasets.

4 Technical Challenge

It is challenging to demonstrate the effectiveness of our models on multiple data sources because of the inherent tendency to overfit individual datasets, and the need to gather sufficient data from various sources to establish a reliable ground truth for prediction. To overcome this challenge, we plan to fine-tune a version of our model on multiple data sources to study their performance and usability across various platforms.

While fake news and clickbait are related, current literature treats them as separate phenomena, making it challenging to establish a relationship between the two without risking misclassification and unreliable findings. To address this challenge, we plan to study the differences between latent representations of fake news and clickbait examples through an innovative strategy.

Multimodality poses the challenge of establishing a robust correlation between textual and visual information, which requires a sufficiently large dataset. However, most efforts currently involve crawling networking platforms such as YouTube, which may not be strong enough. Addressing this challenge may involve finding alternative datasets or improving the relationship between text and image data in the YouTube source.

Successfully addressing these challenges could result in better detection models that perform well across multiple platforms, provide insights into the nature of clickbait, and be robust enough to infer from either text or visual features alone.

5 Division of Labour (tentative)

Task	Member (last name)
Implement state-of-the-art clickbait detection model	Ghosh Tung
Implement state-of-the-art fake news detection model	Wang Rangarajan
Perform analysis of models on different datasets	Guastalla Tung
Explore models that make fine-grained predictions	Ghosh Wang
Explore other modalities to improve predictions	Rangarajan Guastalla

References

BuzzFeedNews. 2017. Buzzfeednews. https://github.com/BuzzFeedNews/

2016-10-facebook-fact-check/tree/master/data.

Webis Clickbait Challenge. 2017. clickbait challenge. https://webis.de/events/clickbait-challenge/shared-task.html#psubmission.

Fake News Corpus. 2017. Fake news corpus. https://github.com/several27/FakeNewsCorpus.

Webis Clickbait Challenge Dataset. 2017. clickbait challenge dataset. https://zenodo.org/record/6362726#.YsbdSTVBzrk.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Bhanuka Gamage, Adnan Labib, Aisha Joomun, Chern Hong Lim, and KokSheik Wong. 2021a. Baitradar: A multi-model clickbait detection algorithm using deep learning.

Bhanuka Gamage, Adnan Labib, Aisha Joomun, Chern Hong Lim, and KokSheik Wong. 2021b. Baitradar: A multi-model clickbait detection algorithm using deep learning. In *ICASSP 2021 - 2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2665–2669.

Saurabh Mathur. 2017. clickbait-detector. https://github.com/saurabhmathur96/clickbait-detector/tree/master/data.

Gilbert E Mitra, T. 2015. Credbank: A large-scale social media corpus with associated credibility annotations. https://github.com/compsocial/CREDBANK-data.

Kai Nakamura, Sharon Levy, and William Yang Wang. 2019. r/fakeddit: A new multimodal benchmark dataset for fine-grained fake news detection.

Olga Papadopoulou, Markos Zampoglou, Symeon Papadopoulos, and Ioannis Kompatsiaris. 2017. A two-level classification approach for detecting clickbait posts using text-based features.

Changyuan Qiu, Chenshu Zhu, and Haoxuan Shan. 2021. Click-bert. https://github.com/PeterQiu0516/Click-BERT.

Benjamin Riedel, Isabelle Augenstein, Georgios Spithourakis, and Sebastian Riedel. 2017. A simple but tough-to-beat baseline for the fake news challenge stance detection task.

William Yang Wang. 2017. Liar. https://www.cs.ucsb.edu/~william/data/liar_dataset.zip.

Tianyi Xie, Thai Le, and Dongwon Lee. 2021a. Checker: Detecting clickbait thumbnails with weak supervision and co-teaching.

Tianyi Xie, Thai Le, and Dongwon Lee. 2021b. Checker: Detecting clickbait thumbnails with weak supervision and co-teaching. In *Machine Learning and Knowledge Discovery in Databases*, Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), pages 415–430, Germany. Springer Science and Business Media Deutschland GmbH. Funding Information: Acknowledgement. The works of Thai Le and Dongwon Lee were in part supported by NSF awards 1742702, 1820609, 1909702, 1915801, 1934782, and 2114824. Publisher Copyright: © 2021, Springer Nature Switzerland AG.; European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases, ECML PKDD 2021 ; Conference date: 13-09-2021 Through 17-09-2021.

Savvas Zannettou, Sotirios Chatzis, Kostantinos Papadamou, and Michael Sirivianos. 2018a. The good, the bad and the bait: Detecting and characterizing clickbait on youtube.

Savvas Zannettou, Sotirios Chatzis, Kostantinos Papadamou, and Michael Sirivianos. 2018b. The good, the bad and the bait: Detecting and characterizing clickbait on youtube. In *2018 IEEE Security and Privacy Workshops (SPW)*, pages 63–69.