Project Proposal: Automatically Generating Clarifying Context for Misleading News Headlines

Anonymous ACL submission

Abstract

News headlines have been playing a crucial role in shaping public opinion for many years now. But many headlines, especially nowadays, are intentionally sensationalist or clickbait and often leave out important context from the full article, which can influence people's perception on the given topic. Several research groups have been working on identifying such misleading headlines, but without further explanation for the reader. This proposal presents a project idea that will be elaborated over the course of the semester: A novel NLP system that not only detects misleading headlines (reproducing existing baseline models) but also automatically generates clarifying context by extracting and summarizing relevant information from the article body. By bridging the gap between clickbait headlines and nuanced article content, the proposed project work aims to improve media literacy and reduce misinformation.

1 Introduction

News headlines are the first impression for readers and can therefor heavily influence public perception. Because of that, clickbait headlines can lead to widespread misinformation. Existing research focuses on detecting misleading headlines, which can be difficult as the question whether a headline is misleading can not always be answered with a clear "yes" or "no". Even though it seems sensible to use the whole grayscale of potential answers, I have found no work on automatically generating corrective context that clears up the true content of an article. This proposal introduces a project to develop a text generation model that, given a news article with a potentially misleading headline, produces a concise corrective statement that highlights the key details missing from the headline.

2 Research Question

Can I develop an NLP-based model that generates clarifying context for misleading news headlines by extracting and summarizing crucial information from the article body? This approach shifts the focus from binary detection to explanatory text generation, which might counteract misinformation more effectively.

3 Rationale and Motivation

Misleading headlines can distort public understanding and contribute to the spread of misinformation. Current systems usually flag such headlines without offering any additional information, leaving readers with an information gap. By generating clarifying context, this project will provide readers with a concise summary of the missing details, which could lead to a better informed public discourse.

4 Literature Review

Several studies have addressed aspects of this problem, though none have focused on generating corrective context:

- **Detecting Incongruity Between News Headline and Body Text via a Deep Hierarchical Encoder (Yoon et al., 2019)**: Introduces a dataset and a hierarchical encoder model to detect discrepancies between headlines and article texts. Does not implement any additional steps.

- **Detecting Misleading Headlines Through the Automatic Recognition of Contradiction in Spanish (Torres et al., 2023)**: Applies contradiction recognition in Spanish news articles to identify misleading headlines. The focus again is exclusively on classification.

- **Detecting News Headline Hallucinations with Explanations (Shen et al., 2023)**: Proposes a framework that explains headline hallucinations by highlighting discrepancies with the body text, but does not generate corrective context.

- **Fake News Detection with Headlines (Ramirez, 2023)**: Examines machine learning techniques for classifying news headlines as fake or genuine. Shows the limitations of detection-only approaches.

- **Clickbait Detection: A literature review of the methods used (Rakhmawati, 2019)**: Provides a comprehensive overview of clickbait detection methods and emphasizes the need for systems that go beyond flagging misleading content to actively supplement it with explanations.

5   Data Sources and Annotation Strategy

A comprehensive dataset is critical for training the clarifying context generation model. The following four datasets are proposed:

1. **Cornell Newsroom Dataset:** Contains over 1.3 million news articles with headlines and full texts from multiple publishers. Since it is not annotated for misleading headlines, I could try out heuristic rules or source-based proxy labels, or manually annotate headlines.

2. **NELA-GT-2018:** Contains approximately 0.7 million articles from 194 news stations, also including credibility ratings. These ratings can help decide which headlines might be misleading. This dataset is therefor offering a mix of reliable and misleading examples.

3. **News Aggregator Dataset (Kaggle):** Provides about 0.4 million news articles with headlines, full texts, and category labels. The same techniques as for the Newsroom Dataset could be used to label the headlines.

4. **Clickbait Challenge 2017 Dataset:** This dataset is older, but it contains annotated clickbait (misleading) versus non-clickbait headlines along with the associated articles, which could be a valuable baseline dataset for the task.

6   Methodology

The project will have several phases:

6.1   Data Preparation and Annotation

- Use the Clickbait Challenge 2017 dataset as a directly labeled baseline.

- Extract headline-article pairs from the Newsroom, NELA-GT-2018, and News Aggregator datasets.

- Use heuristic methods and source credibility ratings to flag potentially misleading headlines.

- Manually verify the results of this flagging method

- Manually annotate a subset for corrective context—identifying key details that are omitted in the headline.

- Preprocess the datasets into one single dataframe.

6.2   Model Development

- Fine-tune a pre-trained text-to-text generation model (e.g., T5 or BART) on the annotated dataset.

- Input: Concatenated headline and article text.

- Output: A short, corrective statement summarizing missing yet crucial details, or the statement that the headline seems valid.

- Optional: Experiment with reinforcement learning and other training techniques to optimize the quality of the generated context.

6.3   Evaluation

- Use automatic metrics like ROUGE, BLEU, and BERTScore to assess the similarity between generated corrective context and reference annotations.

- Conduct a human evaluation study where participants rate the clarity and usefulness of the generated context.

- Compare against baseline methods to demonstrate improvements.

2

6.4 Prototype Demonstration

The project will be developed in a Jupyter Notebook to demonstrate the end-to-end process:

- Input a headline-article pair.

- Display the generated corrective context.

- Outline future integration into a browser extension or web application.

7 Expected Contributions

The project is expected to contribute:

1. **A New Task Definition:** Formalizing the task of corrective context generation for misleading headlines.

2. **An Annotated Dataset:** A dataset of recent and labeled news datasets.

3. **A Generative Model:** An NLP system that produces clarifying context for headlines, validated through quantitative metrics and human evaluation.

4. **Practical Applications:** A demonstration that lays the groundwork for real-world deployment (e.g. a browser extension).

8 Conclusion

This proposal outlines a project to address a potentially critical gap in media literacy by automatically generating clarifying context for misleading news headlines. The project aims to shift the focus from sole detection to explanation, improving readers' understanding and fighting misinformation. It is also planned to create a new and widely usable dataset for this task. The proposed approach is innovative in that it emphasizes generation rather than only classification and could have practical impact for readers worldwide.

Limitations

Since this project deals with a new task, some challenges are the need for high-quality annotations and the difficulty of ensuring that the generated clarifying context is accurate and also useful. These limitations will be addressed through careful dataset curation and extensive evaluation.

References

Shen et al. 2023a. Detecting news headline hallucinations with explanations.

Torres et al. 2023b. Detecting misleading headlines through the automatic recognition of contradiction in spanish.

Yoon et al. 2019. Detecting incongruity between news headline and body text via a deep hierarchical encoder.

Rakhmawati. 2019. Clickbait detection: A literature review of the methods used.

Ramirez. 2023. Fake news detection with headlines.