

Automatic Incongruent News Detection (From the Perspective of Body and Headline Centric Representation)

Sujit Kumar (186101107)

May 2024

1 Abstract

The prevalence of deceptive and incongruent news headlines has demonstrated their significant role in propagating fake news, which worsens the dissemination of both misinformation and disinformation. In the literature, incongruent news article detection has been studied from two aspects- *body-centric* and *headline-centric* encoding. However, earlier headline-centric and body-centric approaches in the literature fail in the following scenarios. (i) The hierarchical encoding in the earlier studies is limited to paragraph level only, and headline-guided attention highlights paragraphs that are contextually similar to headlines. However, considering the underlying incongruent news detection task, highlighting the paragraphs or sentences that are not contextually similar to the headline is essential. (ii) It fails to detect partially incongruent news articles. To address the first limitation of studies in literature, this thesis proposes a **Gated Recursive And Sequential Deep Hierarchical Encoding GraSHE** method for detecting incongruent news articles by extending the hierarchy structure of news body from body to word level and incorporating incongruent weights. The proposed model, (*GraSHE*) captures the long-term dependencies and syntactic structure by incorporating sequential information at the paragraph and body level (using BiLSTM), and syntactic structure at the sentence level using child-sum Tree LSTM. Further, unlike headline guided attention models, (*GraSHE*) also incorporates incongruity weight to capture non-dominant textual segments which are not congruent with other part of news body. To address the second limitation of studies in literature, this thesis proposed dual summarization and graph context matching based methods. This thesis proposes dual summarization-based methods *Multi-head Attention Dual Summarization MADS* and *dual-summarization based approach*, namely **DuSum**, which divides the news article body into two sets, positive and negative set. Sentences congruent to the headline are placed in the positive set, and sentences incongruent to the headline are placed in the negative set. Then, generate two different summaries of both the positive and

negative sets. Next, match the headline with a summary of positive and negative for incongruent news article detection. This thesis proposed graph context matching based methods, *Graph-based Context Matching* **GCM** and *Graph-based Dual Context Matching* (**GDCM**). Both **GCM** and **GDCM** methods first represent headlines and news bodies as a bigram network to capture contextual relations between words and document structure. Then, for every word in the headline, both methods extract context from the news body Bigram Network. Next, it estimates the similarity between the extracted context and the headline for incongruent news detection. **GCM** extract only positive context (context of headline from paragraphs or sentences where discussion regarding headline key is preset). Whereas, **GDCM** extracts both positive context (context of headline from paragraphs or sentences where discussion regarding headline key is preset) and negative context (context of headline from paragraphs or sentences where discussion regarding headline key is not preset). For all the proposed methods, we conduct extensive experiments on three publicly available benchmark datasets. Our experimental results suggest that the proposed model outperforms existing state-of-the-art models in literature and efficiently detects partially incongruent news.