ASSESSING THE INCONSISTENCY IN ONLINE NEWS

Honour Chika Nwagwu, Guy Pascal Kibuh, Hyacinth Agozie Eneh and Stanley Abhadiomhen

Computer Science Department, University of Nigeria, Nsukka, Nigeria

ABSTRACT

The information on the web can be inconsistent across different web pages. News articles are examples of information on the web that are inconsistent and this paper proposes an approach that enables the visual analysis of inconsistencies in online news. It presents an approach which will enable the visual identification of inconsistencies associated to a news headline of interest. It uses a visual assessment approach that relies on two techniques, namely Fault Tolerance and Co-occurrence techniques. The Fault Tolerance technique is used in extracting related news headlines on the internet while the Co-occurrence technique is used for grouping and scaling related news headlines on the web. Also the bar-chart is used to plot charts that summaries the inconsistencies from which news readers can visually assess related news headlines of particular context.

KEYWORDS

Inconsistency, News, Fault Tolerance, Colour Coding, Charts, Co-Occurrence

1. INTRODUCTION

A new reader must have a holistic understanding of his news of interest to be able to filter the noise associated with the news. Evidently, existing approaches for searching and viewing online news articles require much effort, concentration, attention to detail and time to identify related news to the context of interest of the news reader. This is because the news reader must go through different web pages of different websites while reading and assessing the validity of information related to his news of interest to properly assess any associated inconsistency (diverse opinions of editors). News readers are likely not to have a holistic understanding of a news context published where they cannot assess the different dimensionalities to the news. For example, a real-life evaluation about “Trump’s Facebook ban” using Google news search on the 9th of June 2021 reveals the information depicted in Table 1. The Table shows different news headlines from different sources about Trump’s Facebook ban.

Certainly, different news articles about the same context often offer a variety of perspectives (Liu et al., 2019). Also, there are possibilities of bias when forming news title and these biases can be related to political affiliation, misinformation, and systematic errors among others. For example, Journalists can express their ideological view of news by misrepresenting the essence of the story (Gangula et al., 2019). Consequently, a news reader who reads about “Trump’s Facebook ban” for example, should be interested in reading about most of the different dimensionalities associated to Trump’s Facebook ban as evident in Table 1.

Unfortunately, the existing approaches for identifying related news articles require much effort, concentration, attention to detail and time to identify the different news relating to the context of interest. Existing approaches to dealing with the inconsistencies in news do not empower news readers or the data analysts with the ability to easily visualise, the multidimensionality associated to news of a particular context. For example, a traditional approach to reading news implies that the news reader has to search for news of interest, get a list of corresponding news headlines and repeatedly click on associated headlines to read the content of the news. This is very tedious as there are hundreds of related news articles with similar headlines on different web pages on the internet. Even news websites that display related news such as BBC and CNN, which provide the functionality for news reader to see related news headlines at the bottom of each news article, do not reveal the inconsistencies across the web unlike the approach proposed in this work. Also, these commercial news websites do not allow the news reader to determine the relatedness of the displayed headlines to the headline of interest which our approach allows.
### Table 1. Examples of Inconsistencies in news where “Trump’s Facebook ban” is a search string in Google News Search Engine

<table>
<thead>
<tr>
<th>S/No</th>
<th>Headline</th>
<th>News Media</th>
<th>Web Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>Trump is suspended from Facebook for 2 years and can’t return until ‘risk to public safety is receded’</td>
<td>The Washington Post</td>
<td><a href="https://www.washingtonpost.com/technology/2021/06/03/trump-facebook-oversight-board/">https://www.washingtonpost.com/technology/2021/06/03/trump-facebook-oversight-board/</a></td>
</tr>
<tr>
<td>5</td>
<td>Facebook to suspend Trump’s account for two years</td>
<td>The Guardian</td>
<td><a href="https://www.theguardian.com/us-news/2021/jun/04/facebook-donald-trump-oversight-board-instagram">https://www.theguardian.com/us-news/2021/jun/04/facebook-donald-trump-oversight-board-instagram</a></td>
</tr>
</tbody>
</table>

The review process of the existing approaches showed a categorization of the existing approaches to dealing with the inconsistencies in news to include among others, those of the traditional approaches, the artificial intelligence approaches, and the sentiment approaches. These are discussed in section 2. A novel approach to enhancing the assessment of the inconsistencies among similar online news, which combines Fault Tolerance and Co-occurrence approaches is proposed in section 3. The results from implementing the approaches are discussed in section 4. Finally, a conclusion and future work are outlined in section 5.

## 2. EXISTING APPROACHES TO DEALING WITH THE INCONSISTENCIES IN NEWS

### 2.1 Traditional Web Search Approaches

An online news reader manually assesses news on the internet. This necessitates the news reader to search for news of interest, get the web links of corresponding news headlines and click on each web link to read the news contents. The news readers read through all of the articles to be able to mentally mine the contradictions, missingness or reason with any inconsistencies associated to the news. However, there are thousands of news articles on the internet and it is important that a news reader identifies the related news. The traditional approach takes so much time for an online news reader to find and read all the news articles relating to his article of interest.

Also, the traditional approach does not enable the news reader to easily identify false or misleading news. Examples of this approach as proposed by different authors to assessing news include; coding or theoretical frameworks such as discourse analysis and content analysis as outlined in (Gangula et al., 2019; Chen et al., 2018; Armstrong et al. 2018). The appropriateness of news headline to its content was manually analysed and presented in (Gangula et al., 2019). An approach for bias flipping news headlines is presented in (Chen et al., 2018). The authors in (Chen et al., 2018) manually annotated news headlines and automatically used Rouge Score in evaluating the headlines to identify opposite biased articles. The traditional approach also applies advanced deep learning model to flip the bias of news headlines. Also, several datasets that present bias or enable the analysis of news and news headline are curated from different sources and discussed in (Horne et al., 2018; Piotrkowicz et al., 2017).
The manual approaches to analysing the inconsistencies in news do not enable the evaluation of bias or fake news in real-time. They do not provide the online news analyst a platform to evaluate the news about his topic of interest. Also, the approach requires a lot of effort, concentration, attention to detail and time. This is because the news reader must go through different web pages of different websites while reading and assessing or curating the information of interest to properly assess inconsistencies in news.

2.2 Artificial Intelligence Approaches

The artificial intelligence approaches (MLA) include machine learning and natural language processing approaches. MLA involves algorithms that can learn and build mathematical models to make predictions and decisions. There are numerous MLA currently in use in assessing inconsistencies in news. A curation of crime-related information from multi-sources which are digitally published news articles were collected over a period of five years and analysed using deep convolution recurrent neural network model to extract different crime related entities and events is documented in (Dasgupta et al., 2018). The application of deep convolution recurrent neural network model enables the detection of crimes from the database. Even so, it does not enable the analysis of the inconsistencies associated to the news articles.

Also, it is explained in (Liu et al., 2019), how the frame detection approach is used to automatically detect frames of news headlines related to gun violence. This approach enables large-scale analysis of framing trends surrounding the gun violence issue in the United States. A hierarchical architecture that models a complex textual representation of news articles, and measures the incongruity between news headlines and body text approach is proposed in (Yoon et al., 2019). An approach to detect incongruity between a news headline and body text of a news article using a graph-based hierarchical dual encoder (GHDE) is the work of (Yoon et al., 2021). A deep hierarchical attention network that trained to extract hidden patterns in fake news using the concatenation of news headlines and their corresponding body text as input data-set is proposed in (Meel and Vishwakarma, 2021). Also, Natural Language Processing (NPL) network based detection approaches include the headline attention network as discussed in (Gangula et al., 2019) where it is used to automatically detect bias in newspaper articles.

2.3 Sentiment Analysis Approach

Sentiment analysis approaches are applied in analysing inconsistencies in news. They provide techniques through which the polarities of opinions are classified and analysed. A news can be positive, negative, republican, and democrat among others. Social emotion lexicons have been generated through calculation of the probabilities of the emotions as evident in (Lei et al., 2014). Similarly, a lexicon based sentiment analysis approach that investigates the emotion (positive, negative, neutral) in news articles is explained in (Taj and Meghji, 2019). The lexicon based sentiment analysis approach uses some predefined lists of words which are associated to specific sentiment to define the document based sentiment level of the investigated article through the use of the Rapid Miner tool. The authors applied the lexicon based sentiment analysis approach in investigating 2225 British Broadcasting Cooperation (BBC) news articles. It was observed that a majority of the investigated articles fell into the negative or positive categories with a minor percentage of articles having neutral sentiments. However, the sentiment based approaches does not present an efficient approach for evaluating the inconsistencies in massive news articles.

3. PROPOSED METHODOLOGY

The vast amount of news articles, disseminated about a news context, highlight the need for visual analytic systems where the information user can visually investigate news of interest. Such visual analytic systems will enable the web user to understand the trend in news relating to his interest and will enhance his experience when exploring the related news articles. Mining and visualization approaches provide human driven approaches to dealing with inconsistencies in news. It enables news readers to explore the interrelationships among data rather than reading mere numbers and text. There are many visualisation techniques for dealing with online information as presented in (Chen et al., 2018), and they include scrolling, pagination, distortion, and suppression. This proposed approach limits its techniques to the use of Fault Tolerance and Co-occurrence in mining and scaling news headlines.
This proposed approach is based on the principle that news headlines are more likely to convey the context of a story by providing the reader with insights about the context of the news. A news headline provides summarization of the major content of the news. Consequently, news readers are likely to have a holistic understanding about a news context which is published where he assesses the different dimensionalities (headlines) to the particular context of interest. Here, a web based platform that iterates through news websites, by using open source news aggregators to identify related news to a news headline of interest, is developed as ViewNews. News aggregators are portals which aggregate current news from different news sources with their associated web links. The proposed approach combines Fault Tolerance and Co-occurrence techniques in order to enable news readers to visually identify any inconsistency in a news headline.

A news headline is evaluated for instance(s) of inconsistencies by exploring associated headlines. An online reader selects a news headline of interest and the system tokenizes it, thereby enabling the news reader to select some of the keywords as mandatory tokens based on his interest. The percentage of the mandatory tokens (fault tolerance) is calculated by using the formula:

\[ FT = 100 - \left( \frac{\text{size}(m)}{\text{size}(t)} \right) \times 100 \]

Where \( FT \) = Fault Tolerance, size \( (m) \) = the number of tokens in the array of mandatory data, and size \( (t) \) = the number of tokens in the array of the news headline.

The identified headlines are depicted as inconsistencies in bars of a bar chart, where each bar is plotted against its Fault Tolerance value. Also, a View Links section is designed to enable a news analyst to view web links of all the depicted headlines enabling him to read the content of the news.

The co-occurrence technique enables the news analyst to select headlines with equal tolerance levels, depicted with the same colour in a bar chart and equal number of words. As a result, similar news headlines can be viewed by selecting particular co-occuring levels. This enables the news reader to view news that is biased towards a particular context. The use of Co-occurrence helps the news analyst to assess the level of popularity of a news headline. For example, when the level of co-occurrences is many, it means that the headlines are observed among many publication houses.

The use of ViewNews app is simple. The news reader or news analyst types the web link of ViewNews (https://view-news.herokuapp.com/#) into his browser to access the app. The news headline of interest is selected from the ‘select headline’ section. Also, mandatory keywords are selected from the select mandatory keyword(s) section. The mandatory keywords are words that make up the initially selected headline. It should be noted that the more the selection of words from the headline, the less tolerance level to be used by the system. A click on search reveals the inconsistencies associated to the news headline of interest. The co-occurrence is consequently selected to view the co-occurrences of keywords which are peculiar. More than one instance of the co-occurrences can be selected and viewed in a bar chart to enhance understanding of the investigated news headline. Examples of the use of Fault Tolerance and Co-occurrence techniques to visually analyse a news headline of interest for associated noise is as evident in section 4.0.

4. RESULTS

ViewNews enables a news analyst to evaluate a news headline for instances of inconsistency by crawling the web through its news aggregators and displaying associated headlines. For example, the headline “Amazon To Open African Headquarters in South Africa” was assessed for issues of inconsistency across the web. Obviously, a look at the headline will imply that Amazon has agreed to open its headquarters in South Africa. However, when this headline is searched for, through the ViewNews platform at 75% tolerance where Amazon and Headquarters are selected as the mandatory keywords, more than 70 other headlines are identified (see Figure 1). Each bar in Figure 1 is associated to a headline that contains the selected keywords. It becomes obvious to the news analyst that the issue of a headquarters for Amazon is a controversial one. Consequently, the app user can click the ‘View Links’ button on the platform to view all the displayed news headlines and read his news of interest by clicking the associated web link.

\[ \text{FT} = 100 - \left( \frac{\text{size}(m)}{\text{size}(t)} \right) \times 100 \]

1 https://view-news.herokuapp.com/
Also, inconsistencies in the news headlines “Amazon To Open African Headquarters in South Africa” was investigated by selecting co-occurring levels 2 as evident in Figure 2. This enabled us to identify other headlines of related articles. For example, the headline as depicted in Figure 2 “Amazon may have just dropped a clue about the home of its new headquarters…” is written by Hayley Peterson at Dec 15, 2017, the second headline “Amazon may have just dropped a clue about the home of its new headquarters” is written by Leanna Garfield at Dec 18, 2017 while the headline that we are examining (100% tolerance) was written by SaharaReporters, New York at April 21, 2021. These headlines and their context are discussing about a different headquarters for Amazon in different years. The inconsistencies in the news were easy to identify through investigating the bar chart and clicking on the associated web links of the headlines while using ViewNews app.

![Figure 1. ViewNew exploration of “Amazon to Open African Headquarters in South Africa” news headline using the fault tolerance technique](image1.png)

![Figure 2. ViewNew exploration of “Amazon to Open African Headquarters in South Africa” news headline using co-occurrence technique for Co-occurrence 2](image2.png)
5. CONCLUSION AND FUTURE WORK

This paper describes the Fault Tolerance and Co-occurrence techniques. The Fault Tolerance technique is used for varying the associated news headlines according to their common tokens. It enables the extraction of related news headlines on the internet. The Co-occurrence technique is used for grouping and scaling related news headlines. Results from Fault Tolerance and Co-occurrence techniques are used to plot a bar chart. Interestingly, most news websites provide the functionalities of enabling their news reader to see related news headlines at the bottom of each news article. However, such related news headline is selected within the website and do not reveal the inconsistencies across the web. Unlike the use of Fault Tolerance in ViewNews, they do not allow the news reader to determine the relatedness of the displayed headlines to the headline of interest. The approach adopted in this work empowers news analysts and news readers to investigate news of interest for instances of inconsistencies across the web. It enhances the experience of the news analyst when exploring the web. It enables them to identify possible instances of fake news, popular news, and biased news among others. Nevertheless, this work falls short of efficiency issues in ViewNews algorithms and quantitative evaluation of ViewNews app. Consequently, further research is ongoing in visual analysis and assessment of online news. The Fault Tolerance and Co-occurrence algorithms of ViewNews, efficiency issues of implementing the algorithms and the application of ViewNews to analyze instances of incompleteness in news are being investigated. In addition, consideration is given towards possible collaboration with other researchers for the purposes of presenting novel approaches to dealing with the issues of inconsistency in online news.

REFERENCES


