SQUad at FIGNEWS 2024 Shared Task: Unmasking Bias in Social Media Through Data Analysis and Annotation

Asmahan Al-Mamari
Sultan Qaboos University, Oman
s133847@student.squ.edu.om

Fatma Al-Farsi
Sultan Qaboos University, Oman
s133941@student.squ.edu.om

Najma Al Zidjaly
Sultan Qaboos University, Oman
najmaz@squ.edu.om

Abstract

This paper is a part of the FIGNEWS 2024 Datathon Shared Task and it aims to investigate bias and double standards in media coverage of the Gaza-Israel 2023-2024 conflict through a comprehensive analysis of news articles. The methodology integrated both manual labeling as well as the application of a natural language processing (NLP) tool, which is the Facebook/BART-large-MNLI model. The annotation process involved categorizing the dataset based on identified biases, following a set of guidelines in which categories of bias were defined by the team. The findings revealed that most of the media texts provided for analysis included bias against Palestine, whether it was through the use of biased vocabulary or even tone. It was also found that texts written in Hebrew contained the most bias against Palestine. In addition, when comparing annotations done by AAI-1 and AAI-2, the results turned out to be very similar, which might be mainly due to the clear annotation guidelines set by the annotators themselves. Thus, we recommend the use of clear guidelines to facilitate the process of annotation by future researchers.

1 Introduction

The FIGNEWS 2024 Datathon Shared Task mainly aims to shed light on the bias and double standards found in the media coverage of the Gaza-Israel 2023-2024 war. The objective of this collaborative task is to build a shared corpus for comprehensive annotation by setting guidelines tailored to address conflicting views on such a sensitive topic. The task also helps build a generation of NLP researchers who will be able to handle raw data and set their own guidelines to analyze complex datasets.

The paper adheres to the required title format and aligns with the shared task's goals by addressing the challenges of bias in media coverage of the Gaza-Israel war. It also considers the influence of language and culture on the type of bias found in the articles. Through our participation, our team contributes to the development of a comprehensive corpus and guidelines to deal with media bias.

The FIGNEWS 2024 Datathon Shared Task requires the use of NLP on social media under two categories: bias and propaganda. The SQUad team focused on one subtask, which is detecting bias. The annotation process involved identifying bias against Palestine, Israel, both, others, or if the text is unbiased, unclear, or not applicable. Furthermore, manual and automated techniques were used to detect bias in a dataset in different languages (English, Arabic, French, Hebrew, and Hindi). Scholars such as Hamborg (2020) have previously attempted to explain the manual process of content analysis for identifying media bias. He mentioned that the analysis questions and hypotheses should be defined first. For our task, this step was already done by the FIGNEWS Task organizers. The researchers divided the actual annotation process into two steps: Inductive content analysis in which Annotators annotate text without any previous knowledge, other than the analysis question. In the same sense, the SQUad team started to analyze texts manually by doing inductive content analysis for a part of the given dataset. Deductive analysis in which Annotators then set some rules to guide the annotation process, based on the findings of the inductive content analysis. Elfardy and Diab (2016) addressed the challenge of annotating political and ideological perspectives in Egyptian social media. They developed an iterative process to create annotation guidelines by focusing on stance on political reform vs. stability, and views on religion's role. By refining their guidelines, they managed to boost the inter-annotator agreement from 76% to 92%. This showcases the importance of having guidelines to annotate data. Our team set the annotation guidelines to be followed in the annotation process.

In addition to manual content analysis approaches, there are also automated analysis methods. Regardless, they have taken a limited view, simply equating media bias with "diverse opinions" (Munson and Resnick, 2010) rather than analyzing it comprehensively. An example of advanced automated approaches is those that aim to detect bias through the process of word choice and labeling (WCL). Lim et al. (2018) developed a method to detect possible biased words in news articles. Their model inspected factors such as word sentiment, named entities, and parts of speech to flag biased language. The scholars managed to demonstrate that the method they developed was effective in identifying biased words. Target-specific sentiment classification (TSC) is another example of automated techniques used to analyze media bias. Yu and Jiang (2019) developed a TSC model using both textual and visual information from news articles. Their model adapts a BERT-based text encoder and a CNN-based image encoder to classify sentiment toward specific targets. The scholars found that their multimodal model outperformed text-only and image-only baselines. However, they note that TSC approaches still face challenges when applied to complex news texts. For this task, the SQUad team used the Facebook/BART-large-MNLI model through zero-shot classification to label the posts under the "Main" category. The process will be further explained in the following section.

2 Annotation Methodology and Examples

To ensure an accurate annotation process, clear guidelines were developed for this task.

2.1 Annotation Guidelines

Annotation guidelines are set to ensure consistent data labeling. They offer detailed instructions and criteria for annotators. These guidelines define bias, list specific indicators, specify labeling categories, provide examples, address potential ambiguity, and outline the overall annotation process. The team first started by writing a clear and concise definition of bias, specifying various types of biases to consider. In addition, The annotators agreed on specific indicators or cues to help them indicate any biases including explicit or implicit statements, tone, framing, and use of stereotypes. In the guidelines, a labeling scheme was established based on the categories that annotators should assign to each text. That included "Unbiased," "Biased against Palestine," "Biased against Israel," "Biased against both Palestine and Israel," "Biased against others," "Not Applicable," or "Unclear." Moreover, Annotated examples were included in the guidelines as reference points to illustrate how each category should be applied

2.2 Annotation Process

The annotators followed systematic procedures to label the data based on the guidelines they set. First, the annotators familiarized themselves with the guidelines and contacted the project team, Professor Zaghouani specifically, to address any uncertainties. When annotating manually, the annotators independently processed the texts, carefully considering bias indicators and assigned appropriate labels next to each article sample. Regular quality control checks were run during the annotation process. In addition, the team members used the Facebook/BART-large-MNLI model through zero-shot classification to label the posts under the "Main" category. This data processing model was chosen by the annotators for its strong performance in natural language processing (NLP) tasks. The data was divided into batches, with 100 samples per batch, to manage the large dataset efficiently. This process involved tokenization, input encoding, model inference, and label assignment. Furthermore, the labeled data was saved in Excel files after each batch processing step. Lastly, the pipeline architecture, using the "Facebook/BART-large-MNLI" model, was utilized to streamline the classification process from tokenization to inference.

2.3 Inter-Annotator Agreement (IAA) Analysis

The annotators conducted an Inter-Annotator Agreement (IAA) analysis to ensure reliability. The two annotators independently labeled the news article samples, following the provided guidelines. Furthermore, the IAA was set to measure the level of agreement between the annotators, however, a specific metric was not specified. Regardless, the guidelines played a crucial role in achieving consistent labeling.

3 Team Composition and Training

SQUad team has two Omani annotators specialized in English Education at Sultan Qaboos University. Both annotators are Arabic native speakers and learners of English as a second language. They share an academic foundation stemming from English education, literature, and linguistics. In addition, they specialize in English which aids their linguistic analysis abilities.

Annotators underwent some consultation meetings to discuss the task, and they were held with some professors at SQU specifically from the English Department of College of Arts and Social Sciences, including Dr. Najma Al Zidjaly who later became the supervisor of the team. Some meetings were held to increase the annotator's awareness of the complexities of the Israel-Palestine conflict to recognize the different forms of bias that could exist in media. According to some researchers, increasing the knowledge about news events can be a major factor that helps in bias annotation (Lim et al. 2020). In addition, NLP experts, such as professors and the task organizers, were contacted to provide guidance for the team members when needed. Also, relevant work related to the NLP filed was used for reference when needed.

The team members used online platforms such as WhatsApp, virtual and face-to-face meetings, and email updates which facilitated the communication process for them. The coordinated efforts of the team were overseen by a supervisor to address any issues or concerns raised by any of the team members. Furthermore, each member's role and responsibility were outlined to streamline decision-making processes. Also, regular check-ins were scheduled to identify any issues and monitor the team's performance.

4 Task Participation and Results

As our team contained two members only, we focused on the subtask related to detecting bias only. Setting clear guidelines helped the team members carry out a smooth annotation process. In addition, the annotators believed that including examples was important to facilitate the understanding of different forms of bias. For example, it was agreed on by both annotators that the use of words such as "terrorists" or "monsters" to describe Palestinian resistance groups is considered biased towards them. In addition, whenever the team faced any ambiguity, they maintained constant communication and referred to the guidelines they set earlier. Based on the final results decided by the FIGNEWS organizers, it was found that the NLP tool we used to annotate the data in the "Main" sheet scored a low centrality score on the Kappa scale (Zaoghouani et al., 2024). Regarding the IAA sheets, which were annotated manually, it was found the annotators had similar results under the different bias categories as shown in the following two tables. This is similar to Elfardy and Diab's (2016) case, where the annotation guidelines they set helped them boost their inter-annotator agreement up to 92%.

	Category						
Language	**Biased against both**	**Biased against Israel**	**Biased against Palestine**	**Biased against others**	**Unbiased**	**Unclear**	**Not applicable**
English	5%	16.25%	43.75%	6.25%	17.5%	6.25%	5%
Arabic	-	18.75%	37.5%	13.75%	21.25%	5%	3.75%
French	-	30%	20%	12.5%	22.5%	8.75%	6.25%
Hebrew	-	3.75%	55%	10%	11.25%	17.5%	2.5%
Hindi	8.75%	15%	36.25%	5%	17.5%	17.5%	-

Table 1: The first annotator's Inter-Annotator agreement results.

Language	Category						
	Biased against both	Biased against Israel	Biased against Palestine	Biased against others	Unbiased	Unclear	Not applicable
English	1.25%	5%	35%	10%	37.5%	8.75%	2.5%
Arabic	-	1.25%	38.75%	11.25%	43.75%	1.25%	3.75%
French	-	1.25%	21.25%	11.25%	48.75%	12.5%	5%
Hebrew	1.25%	1.25%	52.5%	2.5%	27.5%	3.75%	11.25%
Hindi	2.5%	3.75%	23.75%	17.5%	40%	12.5%	-

Table 2: The second annotator's Inter-Annotator agreement results.

5 Discussion

The team relied on a manual approach to annotate the IAA section of the data provided, however, as Hamborg (2020) has pointed out, manual analysis is time-consuming. Also, prior knowledge about the conflict was helpful in the bias annotation process as Lim, Jatowt, Farber, & Yoshikawa (2020) have mentioned. Based on tables 1 & 2, The percentages for each bias category under AAI-1 and AAI-2 turned out to be similar, which proves the importance of having guidelines. Both annotators found that the Hebrew data contained the most bias against Palestine, with over 50% in both cases. Regarding bias against Israel, there was a noticeable difference between the two annotators. For example, in the text "These are the IDF's Merkava tanks that caused devastation in Gaza," one annotator found the text to be biased against Israel while the other categorized it as unbiased. This might be due to misinterpretation of the guidelines at times. Lastly, it was noticed that the data in French contained the lowest percentage of bias against Palestine. Overall, all languages showed a level of bias.

It was noticed that the NLP model that was used in the "Main" section of the data provided could not pick on context cues the way humans do. As advanced as the current models are, they don't have human-level comprehension. This supports Munson and Resnick's (2010) beliefs that automated approaches take a limited view rather than analyzing the dataset comprehensively. Moreover, algorithms don't always capture the intended meanings as their interpretations depend on the context, which aligns with Yu and Jiang (2019) findings. Furthermore, the model we used does not detect sarcasm, irony, and other forms of figurative language well. Also, the data we used was divided into chunks which made the process easier. Analyzing unstructured text at a large scale might be more challenging. Lastly, NLP models, including the one we used, provided us with output but without explanations of their reasoning. This can sometimes make it difficult to understand the findings.

The team's guidelines can provide a framework for categorizing different forms of bias to facilitate further research on bias detection in NLP. In addition, this kind of task can provide the next generation with the skills needed to detect bias. Also, identifying the limitations of NLP models can help scholars focus on developing them.

6 Conclusion

In conclusion, the bias annotation task was done based on the guidelines set by the annotators, with reference to some related work. In this paper, the annotators discussed the methodology employed and the process followed to carry out this study. In addition, the findings demonstrate bias against Palestine was found to be the category with the highest frequency regardless of the language of the text. Overall, the use of examples is crucial to ensure consistency in the annotation process. Based on the results concluded by the FIGNEWS organizers, the guidelines set by our team scored the highest rate. As for the quantity of annotation, the team ranked 5th place out of the 16 teams, and our total data points were 4,400. However, our team did not score the highest overall ranking (11th place) (Zaoghouani et al., 2024) , which may be due to scoring low on the Kappa scale under some categories. Lastly, future research could examine the inclusion of bias annotation techniques in NLP models to develop more bias-aware natural language processing systems.

References

Elfardy, H., & Diab, M. (2016, August). Addressing annotation complexity: The case of annotating ideological perspective in Egyptian social media. In *Proceedings of the 10th Linguistic Annotation Workshop held in conjunction with ACL 2016 (LAW-X 2016)* (pp. 79-88).

Hamborg, F. (2020, July). Media bias, the social sciences, and NLP: Automating frame analyses to identify bias by word choice and labeling. In *Proceedings of the 58th annual meeting of the association for computational linguistics: student research workshop* (pp. 79-87).

Lim, S., Jatowt, A., & Yoshikawa, M. (2018). Towards bias inducing word detection by linguistic cue analysis in news. In *DEIM forum* (pp. C1-3).

Lim, S.,Jatowt, A., Färber, M., & Yoshikawa, M. (2020). Annotating and analyzing biased sentences in news articles using crowdsourcing. In *Proceedings of the Twelfth Language Resources and Evaluation Conference* (pp. 1478-1484).

Munson, S. A., & Resnick, P. (2010, April). Presenting diverse political opinions: how and how much. In *Proceedings of the SIGCHI conference on human factors in computing systems*(pp. 1457-1466).

Wajdi Zaghouani, Mustafa Jarrar, Nizar Habash, Houda Bouamor, Imed Zitouni, Mona Diab, Samhaa R. El-Beltagy and Muhammed AbuOdeh . 2024. The FIGNEWS Shared Task on News Media Narratives, In *Proceedings of the Second Arabic Natural Language Processing Conference (ArabicNLP 2024)*. Association for Computational Linguistics, Bangkok, Thailand.

Yu, J., & Jiang, J. (2019). Adapting Bert for target-oriented multimodal sentiment classification. In *Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence*(pp. 5408-5414).

Appendices