

Annotating and Analyzing Biased Sentences in News Articles using Crowdsourcing

Sora Lim¹, Adam Jatowt¹, Michael Färber², Masatoshi Yoshikawa¹

¹ Kyoto University, ² Karlsruhe Institute of Technology

infosora@db.soc.i.kyoto-u.ac.jp, adam@dl.kuis.kyoto-u.ac.jp, michael.farber@kit.edu, yoshikawa@i.kyoto-u.ac.jp

Abstract

The spread of biased news and its consumption by the readers has become a considerable issue. Researchers from multiple domains including social science and media studies have made efforts to mitigate this media bias issue. Specifically, various techniques ranging from natural language processing to machine learning have been used to help determine news bias automatically. However, due to the lack of publicly available datasets in this field, especially ones containing labels concerning bias on a fine-grained level (e.g., on sentence level), it is still challenging to develop methods for effectively identifying bias embedded in news articles. In this paper, we propose a novel news bias dataset which facilitates the development and evaluation of approaches for detecting subtle bias in news articles and for understanding the characteristics of biased sentences. Our dataset consists of 966 sentences from 46 English-language news articles covering 4 different events and contains labels concerning bias on the sentence level. For scalability reasons, the labels were obtained based on crowd-sourcing. Our dataset can be used for analyzing news bias, as well as for developing and evaluating methods for news bias detection. It can also serve as resource for related researches including ones focusing on fake news detection.

Keywords: News Bias, Dataset, Media Bias, Crowd-sourcing

1. Introduction

News articles are not always written in a neutral manner, but may deviate from the norm by using dedicated words, a specific writing style, or a preferred author’s viewpoint (Hackett, 1984; Morstatter et al., 2018). Such characteristics of media is referred to as media bias and in the context of news articles as news bias. News bias has been a challenge for a long time in the world of media. Truthfulness, fairness, accuracy, and balanced viewpoints have been emphasized in the context of news reporting to avoid news bias, because news can have a large influence on the readers, creating viewpoints and attitudes of people towards social issues, and eventually changing political views and the society (Niven, 2002). Recognizing news bias is therefore an important goal in the world of media and it can support (or alert) not only readers consuming news but also help authors to write texts in neutral style. Nowadays, news articles are mainly published online and are read from various news channels. To monitor and prevent news bias in a timely and efficient manner, first computational approaches (Park et al., 2009; Hutto et al., 2015; Hamborg et al., 2018) have been developed which aim to identify news bias automatically. These methods are based on techniques from natural language processing (NLP) and machine learning. However, detecting news bias automatically is still a major challenge. This can be traced back to several factors.

First of all, news bias is often subtle. Because fairness, factuality, and veracity are considered as crucial for news reporting, the bias often appears with a slight difference of meaning between words and subtle word choice. For instance, it can make a difference when speaking of “climate change” or of “global warming” (Schuldt et al., 2011), or when speaking of “illegal immigrants” vs. “undocumented immigrants” (Lim et al., 2018). As these examples illustrate, identifying such differences, and, thus, media bias

overall, typically requires not only applying sentiment analysis on the news articles, but also obtaining deep understanding of reported news events and their context.

Secondly, there is considerable lack of datasets which contain proper labels for news bias. Authors of existing approaches to news bias detection often create their own (manageable) datasets (Morstatter et al., 2018). They mostly use RSS feeds or crawl news websites for the dataset construction. Naturally, creation of these datasets takes a lot of efforts. Moreover, news bias is understood in different ways and labels for bias are provided with different granularity. Particularly, there are very few datasets available that focus on the fine-grained differences on the sentence level. Note that bias in one or just a few sentences in an article might already cause biased opinions in readers and therefore focusing on sentences when evaluating bias is necessary. However, prior datasets mostly indicate the bias status on the document level or on a news source (outlet) level (see Section 2). In the latter case, all articles from the same source receive the same bias label due to the inheritance from their source.

In this research, we create a news bias dataset¹ which can be applied for detecting subtle differences in news articles, and, thus, for analyzing and creating approaches for detecting news bias on a fine-grained level. In particular, the dataset contains news bias annotations on the sentence level. Furthermore, unlike other resources, our dataset contains articles about different news topics. In particular, we selected 4 topics covering issues on the English news reported between September 2017 and May 2018. Note that the covered news topics are not only associated with politics, but also with other issues, such as “Republican lawmaker commits suicide amid sexual molestation allegations

¹The dataset can be downloaded for research use at <https://github.com/skymoonlight/biased-sents-annotation>.

(named as Johnson)". All these aspects make our dataset unique and promising for future research.

To understand how bias appears in the collected news, we also perform an initial analysis of our dataset. Among other things, we derive statistics concerning the agreement between crowd workers, and compare different user groups w.r.t. the perceived bias.

The rest of this paper is structured as follows: In Section 2, we outline existing datasets in the area of news bias detection and related approaches. Section 3 presents the way in which we constructed the proposed dataset. We also show the format of the dataset to facilitate an easy reuse and adoption. After describing our findings on the analysis of perceived bias in Section 4, we conclude in Section 5.

2. Related Work

News bias definitions. Several prior works have focused on media bias in general and news bias in particular. Generally, according to D’Alessio and Allen (D’Alessio and Allen, 2000), media bias can be divided into three different types: (1) *gatekeeping*, (2) *coverage* and (3) *statement bias*. *Gatekeeping bias* is a selection of stories out of the potential stories; *coverage bias* expresses how much space specific positions receive in media; *statement bias*, in contrast, denotes how an author’s own opinion is woven into a text. Similarly, Alsem *et al.* (Alsem *et al.*, 2008) divide news bias into *ideology* and *spin*. *Ideology* reflects news outlets’ desire to affect readers’ opinions in a particular direction. *Spin* reflects the outlet’s attempt to simply create a memorable story. Given these distinctions, we consider the bias type tackled in this paper as *statement bias* w.r.t. (D’Alessio and Allen, 2000) and as *spin bias* according to (Alsem *et al.*, 2008).

Hyperpartisan detection datasets. Although news articles are a popular resource for performing research in computational linguistics and natural language processing, the number of datasets dealing with news bias detection is very limited. Noteworthy is, first of all, our preliminary dataset (Lim *et al.*, 2018). For this dataset, only a single news event was considered and bias labels were provided on the word level. In contrast, the currently proposed dataset covers several events and considers bias on the sentence level. We can further mention the effort to promote the development of novel approaches to media bias detection within the frame of the SemEval 2019 Task 4 “Hyperpartisan News Detection” (Kiesel *et al.*, 2019). For this challenge, the organizers published a large dataset consisting of 754k news articles. However, bias has been defined as hyperpartisan, i.e., w.r.t. a political stance. News articles are then labeled as *right*, *left*, and *main stream*. Furthermore, apart from the relatively small sub-part which was annotated manually, most of the articles were simply labeled according to their news sources. Consequently, the source-level bias annotation has just been automatically used for assigning the document-level bias.

Horne *et al.* (Horne *et al.*, 2018) present a large dataset of 136k news articles from 92 news sources for studying the complex media landscape. Both fake sources, satire sources, and hyperpartisan political blogs are considered.

Similar to the other datasets, labels are provided on the document level.

The dataset of Creminini *et al.* (Creminini *et al.*, 2019) is a recently published dataset containing news articles concerning the Ukrainian Crisis of 2014–2015 from 43 countries. The bias of each article was classified as pro-Russian, pro-Western, or neutral. Although this dataset can be considered as the closest one to our proposal, there are still significant differences. For instance, the authors labeled data on the article level, while our dataset has labels on the level of sentences. Moreover, our dataset covers several news topics and can thus be used to compare biases across different topics and domains.

Stance classification datasets. The dataset of Ferreira and Vlachos (Ferreira and Vlachos, 2016) is an example of a dataset in the area of *stance classification*. Stance classification describes the task of determining the stance of the author of a text document: whether the author illuminates not only one party, but also the opposition. Consequently, stance classification is related to bias detection. However, again, while we consider here subtle differences in the writing, stance classification operates on the document level.

Fake news detection datasets. Finally, various datasets have been published for fake news detection. Among the most recent and largest datasets in this regard is the one from Wang *et al.* (Wang, 2017). Note that this dataset covers 13,000 manually labeled short statements, but purely in the domain of politics. Also the fake news detection dataset of Perez-Rosas (Pérez-Rosas *et al.*, 2018) has been published recently. The authors cover seven news domains. Note, however, that fake news detection conceptually differs from bias detection, so that such datasets cannot be used for bias detection research.

3. Dataset

3.1. Constructing News Collection

News bias is always relative (i.e., in relation to explicit or implicit reference) and depends on the context of the news event. One way for handling the relative characteristic, is to compare the content of different news articles which are reporting the same news event, and then, contrast the words used in the articles through the different news outlets (Lim *et al.*, 2018). We also follow this strategy for creating our new dataset. We choose four different news events and collect the news articles reporting those events. The news events used by us are entitled as follows (using the titles of reference articles):

1. “Trump Clashes With Sports World Over Player Protests”²,
2. “Facebook critics want regulation, investigation after data misuse”³,
3. “Tillerson says U.S. ready to talk to North Korea; Japan wants pressure”⁴, and

²<https://www.voanews.com/usa/trump-clashes-sports-world-over-player-protests>

³<https://reut.rs/2p1NZaZ>

⁴<https://reut.rs/2BfEzFL>

| Event | News Source | Title of Target Article | Title of Reference Article |
|-------------|-----------------|---|---|
| NFL | ABC News | Trump: ‘Standing with locked arms is good, kneeling is not acceptable’ | (VOA) Trump Clashes With Sports World Over Player Protests |
| | Daily Beast | Trump’s Attack On Black Athletes May Bring a League to Its Feet | |
| Facebook | Daily Mail | Trump-linked Cambridge Analytica tapped 50M Facebook profiles | (Reuters) Facebook critics want regulation, investigation after data misuse |
| | New York Times | Data Firm Tied to Trump Campaign Talked Business With Russians | |
| North Korea | Newsweek | U.S. Will Talk To North Korea ‘Until The First Bomb Drops,’ Rex Tillerson Says | (Reuters) Tillerson says U.S. ready to talk to North Korea; Japan wants pressure |
| | ABC News | Tillerson tries to quell anxieties at State Dept. amid questions about his future | |
| Johnson | Washington Post | Dan Johnson suicide: Lawmaker accused of molesting teen killed himself. His widow calls it a ‘high-tech lynching.’ | (Reuters) Republican lawmaker commits suicide amid sexual molestation allegations |
| | NBC News | Dan Johnson, Kentucky lawmaker who killed himself, claimed he raised woman from the dead | |

Table 1: Sample news articles in our dataset. Bold texts show diverse aspects on the news events.

4. “Republican lawmaker commits suicide amid sexual molestation allegations”⁵.

These events are in connection with allegation of sexual molestation, conflict between racism and patriotism, a data firm intervened election, and international relationships, respectively. In the following, we refer to these events shortly as (1) NFL, (2) FACEBOOK, (3) NORTH KOREA, and (4) JOHNSON. Table 1 shows brief examples of each news event.

For collecting news articles on the same event from different news outlets, we choose *Google News* which provides clustered news articles on the main issues of the current time. We collect the articles linked from the Google news’ pages and extract the title and text content (after manual inspection in order to exclude image-only or unrelated articles). The resulting dataset consists of 371 articles for NFK, 103 articles for the event FACEBOOK, 39 articles for NORTH KOREA, and 44 news articles for the event JOHNSON. The titles of the articles belonging to each news event are very diverse, indicating that the news articles emphasize different aspects and therefore having potentially different degrees of bias. For example, for the event JOHNSON, one news article emphasizes the suicide event as “high-tech lynching”, while another one degrades the event by referring “claimed he raised woman from the dead”. Also the word choice (e.g., “attack on black athletes” vs. “clashes with sports words over players protests”) for the event NFL differs significantly. As our goal is to annotate the sentences, we split all the news articles into sentences by using the popular module from NLTK (Bird et al., 2009).

3.2. Annotation of News Bias

To overcome scalability issue concerning annotations, crowdsourcing has been recently widely used (Zubiaga et al., 2015). We also rely on this approach in our research to

obtain labels for the news article sentences w.r.t. their degree of bias. Crowdsourcing has also the additional benefit of allowing for understanding bias from the point of view of ordinary users who are likely to be typical readers of news articles.

We use Figure Eight⁶ as an underlying crowdsourcing platform, since this platform has already been widely used in many other researches, particularly for tasks which are rather hard for crowdworkers.

We note that it is rather difficult to provide bias-related labels such as binary judgements on each sentence of news articles, as the bias may depend in various ways on the news event and its context. Thus, to effectively design the bias labeling task, we pick one news article as a *reference article* for each analyzed news event. The reference article has been selected from well-known news agencies which supply news items to the news outlets like Reuters or AP. Typically, such news agencies are known to be neutral and containing the least bias. Having a reference article serves two functions: Firstly, reading first the reference news article about the described news event lets annotators become first familiar with the overall news event. If no such reference article were provided, crowd-workers might miss important context information to properly judge the bias of sentences later on. Secondly, assuming that the reference article is relatively bias-free, it might be easier for crowdworkers to recognize biased sentences in the target articles (e.g., due to noticing different word choices, e.g., “attack on black athletes” vs. “clashes over player protests”).

As the preliminary stage to construct a large news bias dataset, we run the task with a subset of articles which are equally sampled according to their news outlet’s political stance instead of labeling whole news articles in the dataset. For the sampling step, we make use of Media Bias/Fact

⁵<https://reut.rs/2AEnW78>

⁶<https://www.figure-eight.com>

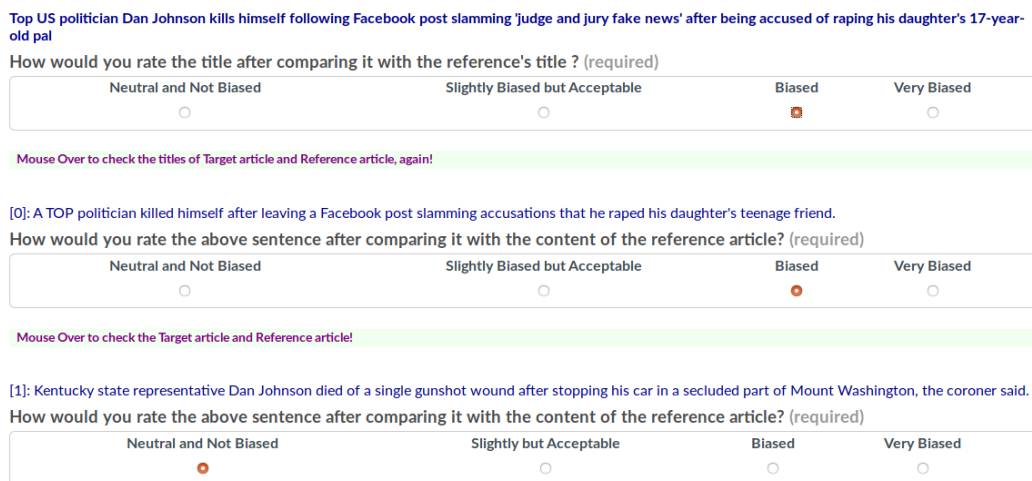


Figure 1: Screenshot of one part of the user interface for our crowd-sourcing task.

Check⁷ website which provides the information of political position of over 2900 media sources according to the ideology spectrum. Specifically, for each news event, we picked for each political position randomly four news articles.

In the annotation tasks, crowd-workers were asked to rate the degree of bias for each news sentence in each target article based on the four-scale category: neutral and not biased, slightly biased, biased, and very biased.

After the completion of the annotation for sentences in each target news article, the crowd-workers are also asked about the bias degree of the entire target article compared to the reference article, and whether they knew about the reported news event before. Our labeling interface is shown in Figure 1.

In total 28 users contributed to our labeling task after excluding untrustworthy annotators. We filtered out those people who submitted all the same ratings only or irrelevant ratings and had unexpectedly quick answering time through their jobs by using a given function of the crowd-sourcing platform and manual examination. Overall, we collected 4,515 annotations for 966 sentences (5 annotations per each sentence) from 46 news articles through 215 unit labeling tasks. An overview of our dataset is given in Table 3.

4. Analysis of Perceived Bias

As we let five different crowd-workers rate the bias degree of 21 sentences (the article’s title and the first 20 article’s sentences) in a job unit, 996 sentences are tagged with multiple bias category, resulting in 29.97% neutral and not biased, 34.37% slightly biased but acceptable, 28.17% biased, and 7.49% very biased – as shown in Figure 2. When we converted bias degrees to binary judgments (such as the first two considered as unbiased and the latter two as biased), 35.55% sentences were assigned to biased sentence category. Table 2 shows examples of the users’ ratings.

After we collect the bias annotations, we examined the inter-rater reliability among the five crowd-workers’ answers. We calculated Fleiss’ kappa score (Fleiss, 1971), which is widely used measure to check the extent of the

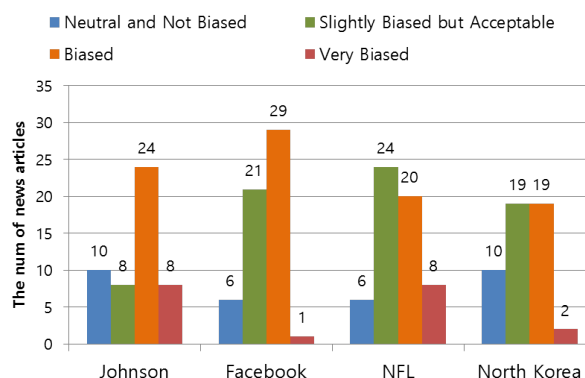


Figure 2: Bias label distribution on document-level

answer agreement among any number of raters giving categorical ratings to a fixed number of items (in our case, five raters giving 4-category ratings to 21 items). The mean scores calculated over all the target articles in each news events are -0.062, -0.078, -0.014 and -0.049 for the labels concerning the news events JOHNSON, FACEBOOK, NFL, and NORTH KOREA, respectively. As shown in Figure 3, the agreement tendency is low. We suspect this low agreement be the result of different sensitivity to the bias between users, as also reported during the creation of a similar dataset in (Lim et al., 2018). We next narrowed down the diversity caused by sensitiveness gap by aggregating the four-scaled bias category to a binary label. In this way, we obtained Fleiss’ kappa mean scores (-0.014, -0.114, 0.0004, and -0.0824) (see also Figure 4).

4.1. User Attributes

During the bias annotation, we also asked the raters whether they already had any knowledge about the reported news events. Based on the answer to this question, we categorized the crowd-workers into the user group “people who knew” and “people who didn’t know” the target news event. Here our hypothesis was that when people have some knowledge about the news event, they may already

⁷<https://mediabiasfactcheck.com>

| Sentence | Label |
|--|---------|
| In the post, written on Wednesday, Johnson paid tribute to his family, saying he had suffered post-traumatic stress disorder for 16 years - “ a sickness that will take my life ”. | biased |
| We’ve been promised by Google, Facebook, and other social sites that our personal information is protected and that when some of our information is provided to third parties, our identity will never be made known . | biased |
| Secretary of State Rex Tillerson is ready to talk about talking to North Korea. | neutral |
| This Wednesday, Kentucky state representative Dan Johnson was found dead by a bridge in Mount Washington with a “single gunshot wound” to the head, according to Bullitt County Coroner Dave Billings. | neutral |

Table 2: Example sentences from our dataset with the labels obtained via the performed crowdsourcing task. Bold texts are considered as a clue of the label judgements.

| Name | Description | Content |
|---------------|--|--|
| event | news event of the articles | {Jonson, Facebook, NFL, NorthKorea} |
| date_event | published date of the news articles | {2017-12-15, 2018-03-18, 2017-09-24, 2017-12-13} |
| id_article | individual id of the news article | from 1 to 46 |
| loc_sentence | position of the sentence in the news articles | title, from 0 to 19 |
| source | news sources of the articles | {Washington Post, New York Times, ...} |
| source_bias | political stance of the news source of the article | {left, left-center, least, right-center, right} |
| ref | the reference article | Reuters, Voice of America (NFL) |
| url | urls of the news article | |
| article_bias | labeled bias degree of the target article | {neutral, slightly biased but acceptable, biased, very biased} |
| sentence_bias | labeled bias degree of each sentences | {neutral, slightly biased but acceptable, biased, very biased} |
| preknow | rater’s prior knowledge of the news event | {yes, no} |

Table 3: Information stored in our dataset

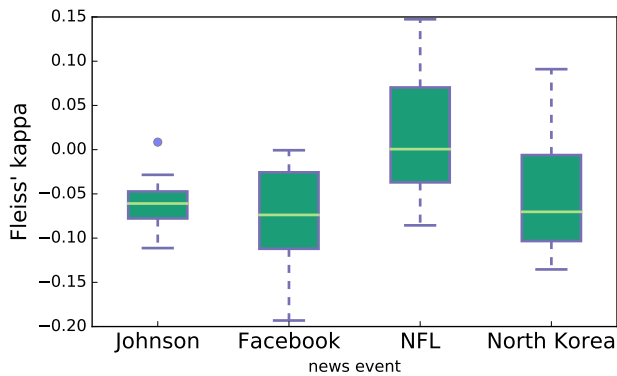


Figure 3: Inter-rater reliability on the Crowdsourcing result: four-scale case

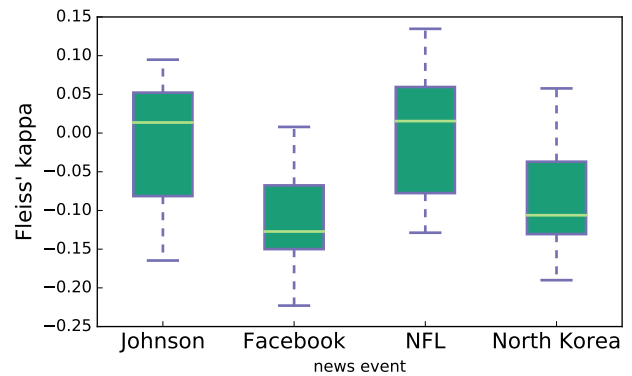


Figure 4: Inter-rater reliability on the Crowdsourcing result: binary case

have formed their own opinion based on the previously acquired information from the news media (which was uncontrolled for us as task providers). Consequently, the hypothesis was that this user group might differ more from the “people who didn’t know” group w.r.t. bias. Figure 5 shows the distribution of the user groups according to the prior knowledge about news events. To determine these user-group differences, we analyzed the inter-rater agreement scores according to the two user groups. The results show that the average agreement scores differ between these two user groups (see Figure 6). If we remove the special cases in which only one person among the 5 annotators has already prior knowledge or no prior knowledge, and then compare

the agreement scores among the two user groups, the “people who knew” group still has a higher agreement score. (Note that the event Johnson leads the “people who knew” group to empty after removing the above special cases. For the sake of comparison, the overall scores fill in for the blank in Figure 6.) However, as Figure 5 shows, overall there was only a considerably small ratio of people who knew the news event.

4.2. News Sources

Even though our focus is on the sentence-level bias in news articles, we should not overlook the political tendency of the news outlet which issues the news article. We can con-

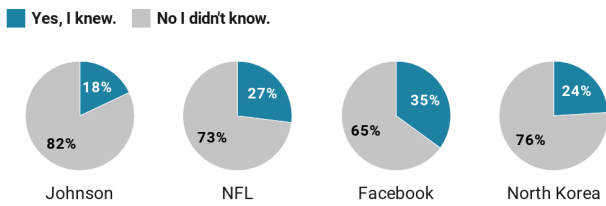


Figure 5: Crowd-workers' prior knowledge about the news events

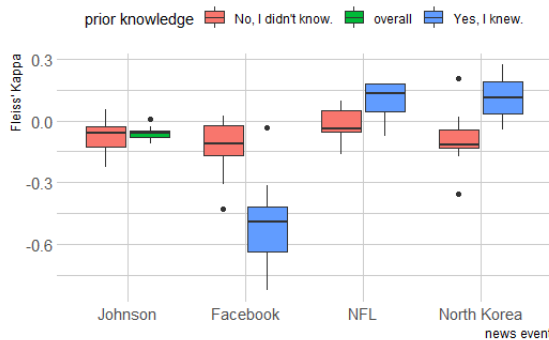


Figure 6: Inter-rater reliability according to the prior knowledge about the news events

sider that if people have established their political preference and support one specific party, they are less likely to recognize the bias in the news articles of the news sources which reflect their political views. However it seems to be inappropriate to ask directly people about their political preference. Therefore, we indirectly examine the political stance of the target news article's sources and its relationship with the inter-rater agreement. In this step, we obtained the political tendency of the news outlets by using the Media Bias/Fact Check website again and then we categorized the news articles according to the political positions of their sources. Political positions are categorized into left, left-center, least, right-center, or right. For intuitive understanding, we rearranged these categories to have (left, least, and right) as a simple version in which it regards left and left-center with left, and right and right-center with right. Figure 7 indicates that according to the news source of the news article, news reader from the different political preferences may differently perceive the bias. From this, if we restrict the annotation task to build on the news articles from the same political stance, so that we minimize the interference by inter-rater different preferences, we could further improve the quality of the achieved labels.

5. Conclusions

Detecting news bias is a challenging task for computer science as well as linguistics and media research areas due to the subtle nature and heterogeneous, diverse kinds of biases. In this paper, we presented a corpus of news articles where sentences of the news articles have been labeled in a crowd-sourcing task concerning their degree of bias. To

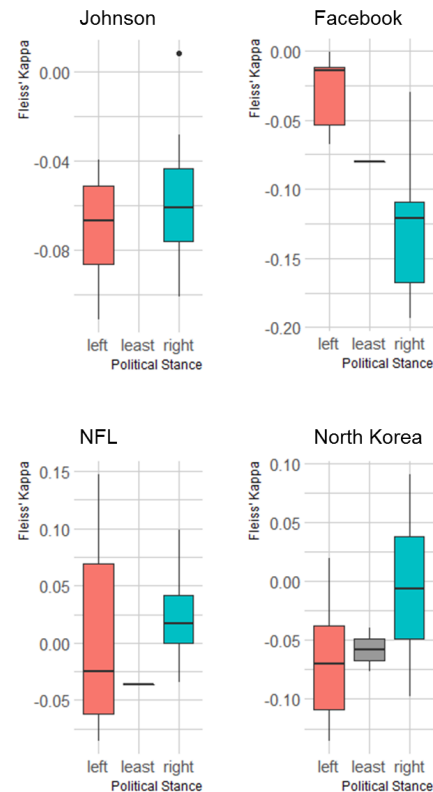


Figure 7: Inter-rater reliability according to the political stance of news sources

annotate subtle bias by using ordinary people on crowd-sourcing who can be considered as general news readership, we provided a standard of comparison of the news content as the reference article. We then analyzed these annotations in regards to perceived bias. Based on our analysis, we conclude that the prior knowledge about news event can be a major factor for bias annotation. In addition, we anticipate that limiting the annotation tasks with news articles having only a single political stance can improve the quality of bias annotation by minimizing the interference by inter-rater different preferences.

For future work, for fostering robust and practical bias detection mechanisms the greater topic variety needs to be considered for further extension such as not only for politics but also economy, global health topics and so on. Also, there is still room for further improvement for bias labels. In this research, any random workers put bias labels for sentences in a document as a job unit. In this regard, employing fixed raters through the whole job could be one way to increase label reliability as well as to decrease interference of personal political preference.

6. Acknowledgments

This research was supported in part by MEXT grants (#17H01828; #18K19841; #18H03243).

7. References

Alsem, K. J., Brakman, S., Hoogduin, L., and Kuper, G. (2008). The impact of newspapers on consumer con-

- fidence: does spin bias exist? *Applied Economics*, 40(5):531–539.
- Bird, S., Loper, E., and Klein, E. (2009). Natural language processing with python o’reilly media inc.
- Cremisini, A., Aguilar, D., and Finlayson, M. A. (2019). A Challenging Dataset for Bias Detection: The Case of the Crisis in the Ukraine. In *Proceedings of the 12 International Conference on Social, Cultural, and Behavioral Modeling*, SBP-BRiMS’19, pages 173–183.
- D’Alessio, D. and Allen, M. (2000). Media bias in presidential elections: A meta-analysis. *Journal of communication*, 50(4):133–156.
- Ferreira, W. and Vlachos, A. (2016). Emergent: a novel data-set for stance classification. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, NAACL-HLT’16, pages 1163–1168.
- Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological bulletin*, 76(5):378.
- Hackett, R. A. (1984). Decline of a paradigm? bias and objectivity in news media studies. *Critical Studies in Media Communication*, 1(3):229–259.
- Hamborg, F., Donnay, K., and Gipp, B. (2018). Automated identification of media bias in news articles: an interdisciplinary literature review. *International Journal on Digital Libraries*, pages 1–25.
- Horne, B. D., Khedr, S., and Adali, S. (2018). Sampling the News Producers: A Large News and Feature Data Set for the Study of the Complex Media Landscape. In *Proceedings of the Twelfth International Conference on Web and Social Media*, ICWSM’18, pages 518–527.
- Hutto, C., Folds, D., and Appling, D. (2015). Computationally detecting and quantifying the degree of bias in sentence-level text of news stories. In *Proceedings of Second International Conference on Human and Social Analytics*.
- Kiesel, J., Mestre, M., Shukla, R., Vincent, E., Adineh, P., Corney, D., Stein, B., and Potthast, M. (2019). SemEval-2019 Task 4: Hyperpartisan News Detection. In *Proceedings of The 13th International Workshop on Semantic Evaluation (SemEval 2019)*. Association for Computational Linguistics.
- Lim, S., Jatowt, A., and Yoshikawa, M. (2018). Understanding characteristics of biased sentences in news articles. In *Proceedings of the CIKM 2018 Workshops co-located with 27th ACM International Conference on Information and Knowledge Management (CIKM 2018), Torino, Italy, October 22, 2018*.
- Morstatter, F., Wu, L., Yavanoglu, U., Corman, S. R., and Liu, H. (2018). Identifying framing bias in online news. *ACM Transactions on Social Computing*, 1(2):5.
- Niven, D. (2002). *Tilt?: The search for media bias*. Greenwood Publishing Group.
- Park, S., Kang, S., Chung, S., and Song, J. (2009). News-Cube: delivering multiple aspects of news to mitigate media bias. In *Proc. of SIGCHI on Human Factors in Computing Systems*, pages 443–452.
- Pérez-Rosas, V., Kleinberg, B., Lefevre, A., and Mihalcea, R. (2018). Automatic detection of fake news. In *Proceedings of the 27th International Conference on Computational Linguistics, COLING’18*, pages 3391–3401.
- Schuldt, J. P., Konrath, S. H., and Schwarz, N. (2011). “global warming” or “climate change”? whether the planet is warming depends on question wording. *Public Opinion Quarterly*, 75(1):115–124.
- Wang, W. Y. (2017). “Liar, Liar Pants on Fire”: A New Benchmark Dataset for Fake News Detection. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics, ACL’17*, pages 422–426.
- Zubiaga, A., Liakata, M., Procter, R., Bontcheva, K., and Tolmie, P. (2015). Crowdsourcing the annotation of rumourous conversations in social media. In *Proc. of WWW 2015*, pages 347–353.