Target-Aware Contextual Political Bias Detection in News

Iffat Maab[1], Edison Marrese-Taylor[1,2], Yutaka Matsuo[1]
[1]The University of Tokyo
[2]National Institute of Advanced Industrial Science and Technology
{iffatmaab, emarrese, matsuo}@weblab.t.u-tokyo.ac.jp

Abstract

Media bias detection requires comprehensive integration of information derived from multiple news sources. Sentence-level political bias detection in news is no exception, and has proven to be a challenging task that requires an understanding of bias in consideration of the context. Inspired by the fact that humans exhibit varying degrees of writing styles, resulting in a diverse range of statements with different local and global contexts, previous work in media bias detection has proposed augmentation techniques to exploit this fact. Despite their success, we observe that these techniques introduce noise by over-generalizing bias context boundaries, which hinders performance. To alleviate this issue, we propose techniques to more carefully search for context using a bias-sensitive, target-aware approach for data augmentation. Comprehensive experiments on the well-known BASIL dataset show that when combined with pre-trained models such as BERT, our augmentation techniques lead to state-of-the-art results. Our approach outperforms previous methods significantly, obtaining an F1-score of 58.15 over state-of-the-art bias detection task.

1 Introduction

News media companies publish thousands of articles every day. While we generally regard these articles as containing factual, true information, studies have shown that various kinds of bias exist in news (Fan et al., 2019; Lim et al., 2020; Gentzkow et al., 2015; Prat and Strömberg, 2013). Further studies have studied the effects that these biases have on readers, particularly in voting. A study by Groseclose and Milyo (2005) suggests that indeed media has a sizable political impact on voting, where for example DellaVigna and Kaplan (2007) found significant effect of exposure to Fox News in increased turnout to the polls.

Clearly, biased media have the potential to sway readers in potentially detrimental paths. Therefore, it is crucial to unveil the true nature of media bias. Furthermore, as all journalism contains narratives (Unesco, 2018), given its role on transforming individual and public opinion, we consider it is worth measuring and understanding the political bias phenomenon. We think bias detection is important as a proxy or mechanism to assess the quality of information in news media. As stated by Unesco (2018), there is no problem with the existence of narratives in substandard journalism, rather poor professionalism.

Bias in news from different aspects has been studied in the past, where for example Chen et al. (2018) and Arapakis et al. (2016) created news quality corpus of 561 articles and study how various news constituents characterize the quality of editorial articles. While these works are highly relevant to the bias problem, they did not specifically or directly target at the issue.

Foundational work in political bias was performed by Fan et al. (2019), who released a human-annotated dataset named Bias Annotation Spans on the Informational Level (BASIL), containing 300 fine-grained bias annotations. Concretely, political bias is identified at the sentence-level, where spans are annotated and a target (the main entity) is identified, in addition to a few other labels. Significantly, BASIL stands as the first dataset to be annotated with different types of bias. **Informational** bias, which depends broadly on the context of the sentence (Guo and Zhu, 2022a) and arises from manipulation of information or selective presentation of content in a factual way, e.g., use of quotes, to evoke specific reader's emotions towards news entities (Fan et al., 2019; van den Berg and Markert, 2020), and **lexical** bias, which stems from the choice of specific words or linguistic phrases that influence the interpretation of a subject, and perpetuate the understanding of information (Re-

Proceedings of the 13th International Joint Conference on Natural Language Processing and the 3rd Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics (Volume 1: Long Papers), pages 782–792
November 1–4, 2023.

casens et al., 2013; Iyyer et al., 2014; Hube and Fetahu, 2019) are present in BASIL. To the best of our knowledge, BASIL is the first dataset that annotates informational bias together with specific targets.

With the release of BASIL, work on political bias detection has mostly focused on informational bias, with a strong emphasis on informational context within and across news media articles, as informational bias is highly content-dependent. In the seminal work, van den Berg and Markert (2020) feed the whole document/article as context for sentence-level bias classification. Though this approach worked relatively well in practice, using long documents in this context brings considerable noise, redundancy and can increase vocabulary size, which can ultimately decrease the performance of the classifier as evidenced by previous work (Akhter et al., 2020; Guo and Zhu, 2022b). Moreover, as shown by Chen et al. (2020), detecting bias at article level remains even more challenging and difficult task.

In light of this issue, several works have recently focused on introducing more specific contextual information to perform classification (Cohan et al., 2019b; van den Berg and Markert, 2020; Guo and Zhu, 2022b), for example by mixing contexts of informational and lexical bias at both the article-level (entire article encompassing target sentence) and event-level (triplet of articles discussing the same event).

While the aforementioned approaches have resulted in improved performance, we think their applicability is limited. On one hand, articles in BASIL have no overall bias label, instead each sentence is labeled as evidence of a certain kind of bias or as a neutral statement, suggesting that these should be treated separately when detecting different kinds of bias. Previous studies (Rao et al., 2018; Tripathy et al., 2017) have already shown that on document-level classification, paragraphs can belong to multiple categories, which Chen et al. (2020), also observed on BASIL, where paragraphs belong to either informational bias, lexical bias or no bias spans. Furthermore, as highlighted by Chen et al. (2020), by mixing contexts of informational and lexical bias, it becomes difficult for the model to distinguish and predict different type of bias, which may result in lower model performance.

In light of this issue, in this work, we provide a framework to generate more consistent and similar bias contexts to improve performance. As shown in Table 1, each instance of annotated bias span also identifies the "target", i.e., the main entity or topic of the sentence that is also annotated in BASIL. Using this information, our key insight is to create event-level contexts that are target-aware and also sensitive to the bias label.

Source	Target	Index	Sentence	Bias
FOX	Obama Campaign	0	President Obama health care plan treats the treasured entitlement like a piggy bank, while the Romney-Ryan plan preserves it.	Inf
HPO		4	If any person in this entire debate has blood on their hands in regard to Medicare, it's Barack Obama.	Inf
NYT		4	Now when you need it, Obama has cut $716 billion from Medicare.	Inf
FOX	Romney Campaign	21	Obama campaign spokeswoman Lis Smith described the new Romney-Ryan ad on the subject as dishonest and hypocritical, considering Ryan's own proposals for Medicare.	Lex
HPO		27	Senator McCain and Governor Romney have subsequently opposed the savings that the president identified and demagogued the issue, ironically, since Governor Romney's running mate kept them in his budget.	Lex
NYT		7	Lis Smith, a spokeswoman for the Obama campaign, said, Mitt Romney's Medicare ad is dishonest and hypocritical.	Lex

Table 1: Bias sentences extracted from event 0 of BASIL with three news media sources, FOX (0fox; source:fox, event:0), HPO (0hpo; source:hpo, event:0), and NYT (0nyt; source:nyt, event:0), showing a single event can exhibit similar targets and bias types to manifest event-based target aware context.

For example, for the target "Obama Campaign", sentences from three different news sources are combined to form a single contextual example for informational bias classification, as highlighted in light gray. A similar procedure is applied for "Romney Campaign", where sentences are concatenated to form an example for lexical bias classification, highlighted in dark gray. Inspired by ideas from modeling context in informational bias detection (van den Berg and Markert, 2020; Chen et al., 2020; Guo and Zhu, 2022b), our approach is able to augment examples with richer contexts and less noise, and follows previous work in determining that the detection of lexical bias should hold equal importance as informational bias (Zhou and Bansal, 2020; Marinov and Efremov, 2019; Maab et al., 2023).

Following recent work (Maab et al., 2023), we tackle a variety of bias detection tasks including INF/OTH and INF/LEX using data from BASIL.

Through extensive experimentation, we demonstrate the effectiveness of our approach by obtaining state-of-the-art performance on all of our studied tasks. In addition, our holistic view on bias enables us to unveil inconsistent terminologies used for contextual information of BASIL, therefore we gather such contexts to improve clarity and uniformity, and to avoid previous work problems as indicated in our comparison with the state-of-the-art.

2 Related Work

Media bias has been scrutinized often with nuanced variations and under different contexts through diverse terminologies. Misinformation detection based on linguistic driven approaches are exposed by novel approaches (Pan et al., 2018; Pérez-Rosas et al., 2017). Powerful players in news media advance their interests by devoting plentiful resources to facilitate controlled communication in politics (Entman, 2007), therefore ideology prediction and trustworthiness of news media draws attention (Baly et al., 2019), while Hamborg et al. (2019) highlighted that distinctive contributions can be made by computer scientists to study bias.

Kulkarni et al. (2018) proposed an attention-based model to capture high-level contexts of news articles including title, link structure, and news information using both textual content and network structure to leverage cues from multiple views. Contextualized representations of sentences for better understanding of documents are studied using numerous pre-trained language models (Cohan et al., 2019b; Iyyer et al., 2014). Inspired from (Cohan et al., 2019b), van den Berg and Markert (2020) work on BASIL to propose several context inclusive models on article and event context, and use three BiLSTMs for encoding FOX, HPO, NYT news documents as triplets. Building upon existing study of (van den Berg and Markert, 2020), (Guo and Zhu, 2022b) use multi-level graph attention networks for bias detection by MultiCTX model that use contrastive learning from sentence embeddings to discriminate target sentences. Another recent study on BASIL (Lei et al., 2022) built distillation models on top of RoBERTa for informational bias classification and explore different types of local and global discourse structures. Similarly, article-level bias classifiers (Chen et al., 2020) use second order bias features of BASIL to manipulate context information using uncased BERT. Using BASIL, BERT by Devlin et al. (2018) remain as a major baseline model in majority of previous studies (van den Berg and Markert, 2020; Guo and Zhu, 2022b). Chen et al. (2020) find that fine-tuned BERT has a strong efficacy and use it to reimplement (Fan et al., 2019) results. In light of the findings, our proposed approach also utilize BERT (Devlin et al., 2018).

3 Proposed Approach

3.1 Bias-Aware Neighborhood Context

Previous work has shown that phrases surrounding a sentence annotated with bias can be used as local context to perform bias classification, and that this local context can contribute to the ability of models to identify and label types of bias. However, by ignoring the nature of these sentences, existing approaches that utilize neighborhood context (van den Berg and Markert, 2020; Guo and Zhu, 2022b) can run into problems by introducing ambiguous content, for example when adding sentences that are annotated with the opposite bias. As shown by (van den Berg and Markert, 2020), this can also lead to massive data leakage problems across train and test sets.

To account for the disparity in how different bias contexts are overlooked in previous work, in this paper, we propose to care for the bias label of neighboring sentences, advancing to generate Bias-Aware Neighborhood Contexts (BANC), and adding neighboring sentences to the model input as long as they have a related bias label. Table 2 shows an example of how this procedure works. Since, our approach is bias-sensitive, sentences with informational and lexical bias are treated separately. Therefore, for a given target sentence with index 1, the former (index 0) and next (index 2) sentences become neighbor sentences of lexical bias as they exhibit no bias span as highlighted in green. Correspondingly, to generate a BANC for informational bias classification, we combine sentences with indices 2, 3 and 4 as highlighted in blue. Teal (green + blue) color is shown by sentence index 2, since it is common between the two BANC text spans. According to the same principle, for cases where the first sentence of an article has bias, next sentence is checked and combined, whereas in the event where it is last sentence, former sentence gets checked and successively combined.

Index	Position	Sentence	Bias
0	Neighbor	Israel and Middle East policy have a tendency of surfacing in presidential politics in rather combustible ways.	-
1	Target	And a new advertisement that will run in areas of Florida with large Jewish populations attempts to stoke anxiety over American policies in the region, using a news clip of Prime Minister Benjamin Netanyahu of Israel warning of the risks of a nuclear Iran.	LEX
2	Neighbor	The fact is that every day that passes, Iran gets closer and closer to nuclear arms, Mr. Netanyahu is shown saying.	-
3	Target	For dramatic effect, a soundtrack fit for an episodic drama like Homeland plays as the prime minister continues.	INF
4	Neighbor	The world tells Israel, Wait. There's still time.	-
5	Neighbor	And I say wait for what?	-
6	Neighbor	Wait until when?	-

Table 2: An article of New York Times section extracted from BASIL showing bias-aware neighborhood context of informational bias in green and lexical in blue.

	Target	Sentences			Target-aware examples				Total
					Article-level			Event-level	
		FOX	HPO	NYT	FOX	HPO	NYT		
	Benjamin Netanyahu	1	-	-	1	-	-	-	1
18	Barack Obama	5	1	-	10	1	-	5 (fox × hpo)	16
	Secure America Now	-	2	2	-	1	1	4 (hpo × nyt)	6
Total					within Art. = 14			9	**23**
	Hillary Clinton	5	-	3	10	-	3	15 (fox × nyt)	28
22	Barack Obama	2	2	-	1	1	-	4 (fox × hpo)	6
	Nancy Pelosi	1	-	-	1	-	-	-	1
Total					within Art. = 16			19	**35**

Table 3: Detail of the number of contextualized instances obtained by applying our proposed ABTA and EBTA to a set of the original examples from BASIL, in this case taken from events (E) 18 and 22, for the case of informational bias.

3.2 Target-Aware Context

While our neighboring approach helps identify local context relevant for bias classification, we believe that global context, either at the article or event levels, can also be exploited. To that end, we note that BASIL contains annotations that also identify the "target" of a given sentence where either lexical or informational bias is present. This "target" label refers to the main entity or topic of the sentence that is annotated, with some of the most prominent targets in BASIL being entities or people that lie at the core of news reports, such as Donald Trump, Romney Campaign, Secure America Now, among others.

We further note that although the frequency of appearance of a given "target" varies substantially, as long as we keep the annotated label constant (e.g., lexical), the context remains the same. This motivated us to gather all surrounding linguistic cues pertaining to a specific target at both the article-level and event-level. Concretely, we create target-aware contextual information by making use of all possible combination spans having the same bias and target, and propose article-based target-aware (ABTA) and event-based target-aware (EBTA) contexts, which we explain below.

As show in Table 3, using ABTA context, for instance, the target "Barack Obama" which has 5 sentences annotated with informational bias in the FOX article and 1 in HPO, generates all possible combinations of two sentences within FOX giving us 10 contextualized examples, and 1 same example in HPO because this article has only one sentence, respectively. Note that possible combinations of sentences within articles are combined in groups of two only, which we do to emulate the natural distribution of occurrence of sentences with the same bias and same "target".

EBTA contexts shown in the "Event-level" column in Table 3, are computed for common targets across articles, for instance, the same target "Barack Obama" with informational bias appear across FOX and HPO with 5 and 1 sentences, therefore all unique possible combinations in groups of two generates 5 new contextualized examples across the two aforementioned articles. Finally, following the example in the table for "Barack Obama", the combined contexts of ABTA and EBTA give us a total of $10+1+5 = 16$ contextualized informational bias examples for a single target. Note that we repeat this procedure for generating target-aware lexical bias contexts.

Because of the way in which we combine sentences, it is evident that our approach is significant in providing contextualized examples for infrequent targets as well, therefore also contributing towards mitigating imbalanced bias distribution and skewed nature of "targets" as observed in BASIL articles (Chen et al., 2020).

Using our target-aware techniques, we are able to get more than triple the amount of training examples for training for lexical bias detection (462 sentences v/s 1,551 contextualized examples), and we observe a fourfold increase of examples for informational bias detection (1,221 sentences v/s 4,987 contextualized examples). Please see Table 4 for a detailed explanation on target-aware context generation for the most frequent "targets" in BASIL. Our separate use of lexical bias contexts guaran-

Target	Target Aware Context	
	Sentences	Possible Combinations
Donald Trump	340	2767 (Inf: 2386, Lex: 381)
Barack Obama	119	619 (Inf: 479, Lex: 140)
Barack Obama*	156	870 (Inf: 705, Lex: 165)
Hillary Clinton	62	327 (Inf: 292, Lex: 35)
Democratic Lawmakers	36	119 (Inf: 97, Lex: 22)
Joe Biden	32	325 (Inf: 241, Lex: 84)
Paul Ryan	25	122 (Inf: 97, Lex: 25)

Table 4: Most frequent bias targets in BASIL across events and their possible combinations using target-aware context. Barack Obama* includes three similar targets: Barack Obama, Obama's administration, Sasha and Malia Obama with 119, 21, and 16 bias sentences.

tees that the model is not relying solely on shallow lexical features of a complete article as in previous work (van den Berg and Markert, 2020; Guo and Zhu, 2022b), and instead looking for cues on the same categories or bias type (Rao et al., 2018; Tripathy et al., 2017) relevant for the task at hand.

Since, prior studies focused solely on informational bias and overlooked other bias spans, we surmise that lexical bias detection is also significant as supported by Maab et al. (2023), and provide a more concise and sensitive bias narratives with neighborhood contexts together with target-aware contexts.

Finally, based on successful results reported by previous work (Mikolov et al., 2013; Maab et al., 2023), we additionally use a backtranslation approach to generate more data, which we apply to our contextualized samples using Spanish as a pivot language. By incorporating multiple viewpoints in our neighborhood and target-aware contexts, we facilitate our model in providing a broad and inherent semantics of biased targets to manifest variations in bias representations. Our extensive experiments will further demonstrate the impact of proposed context in different training settings.

4 Experimental Setup

To streamline the comparison with prior work (van den Berg and Markert, 2020; Guo and Zhu, 2022b), we use a 10-fold cross-validation setting where bias-aware neighborhood and event-based target aware contexts never appear at the same time in non-overlapping train-val-test split sets of 80-10-10, respectively. Average performance of our model using three seed runs is reported in all our experiments.

For the sentence-level bias detection, we perform two classification tasks: detection of informational bias (INF/OTH) and classification of bias type (INF/LEX). Inspired by Maab et al. (2023), for INF/OTH bias task we combine BANC, EBTA and ABTA with backtranslation of both examples labeled as containing either informational or lexical bias. For the INF/LEX, we do this only on lexical bias examples.

We refer to the original set of examples in BASIL, without augmentation as "regular". We do not perform any augmentation techniques for the testing examples . Furthermore, to examine the effectiveness of our proposed components in ablation studies, regular BASIL examples (Fan et al., 2019) are augmented with BANCand target-aware contexts in fractions of 10%, 20%, 30%, 40%, 50%, 100%, and 100% with BT (additional backtranslated examples).

Baselines Majority of approaches in previous studies concentrate on deep learning methods for identifying media bias. We compare our work with models that use different kinds of BASIL contexts for sentence-level bias detection to ensure consistent and impartial evaluation. We consider multiple contextual models that address the detection of informational bias, for example, SSC (Sequential Sentence Classification) Cohan et al. (2019a) and its variant WinSCC (windowed Sequential Sentence Classification) (van den Berg and Markert, 2020), RoBERTa, ArtCIM for target sentences within an article, and EvCIM for triplets of articles covering the same event (van den Berg and Markert, 2020; Guo and Zhu, 2022b). Guo and Zhu (2022b) further proposed MultiCTX model and reproduce the results using WinSCC and EvCIM for informational bias detection. We also compare against the fine-tuned RoBERTa model (Lei et al., 2022), as well as BERT (Maab et al., 2023; Chen et al., 2020; Devlin et al., 2018; Fan et al., 2019).

Implementation Details We use the PyTorch to implement our models, borrowing from HuggingFace (Face, 2021), our classifiers are based on BERT-base (Devlin et al., 2018), and all our models are trained with 5×10^{-5} as learning rate, 32 as batch size, and 15 as a maximum epoch count. We utilize a server with an NVIDIA V-100 GPU for our experiments.

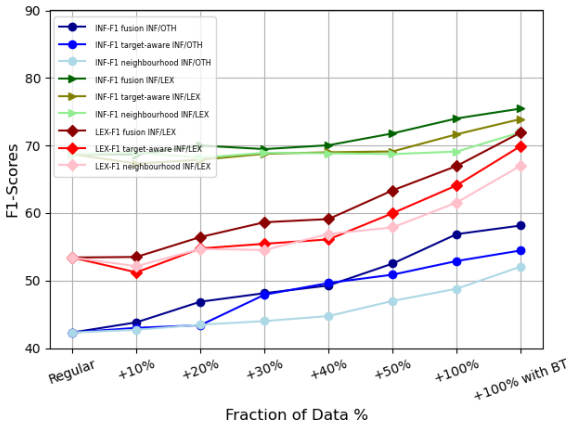

Figure 1: Plot showings various fractions of augmented contexts using BANC (neighbourhood), ABTA & EBTA (target-aware), and integration of multiple contexts (fusion = BERT + BANC + ABTA + EBTA) to examine the effect of INF-F1 score (blue) in INF/OTH task, and INF-F1 (green) & LEX-F1 (red) in INF/LEX task.

			INF/ OTH			INF/ LEX		
				F1 Score			F1 Score	
BANC	**ABTA+ EBTA**	**%**	**Acc.**	**INF**	**OTH**	**Acc.**	**INF**	**LEX**
-	-	Regular	80.14	42.32	87.11	76.69	68.71	53.42
✓	-	+ 10%	80.56	42.72	86.41	77.87	68.98	52.13
✓	-	+ 30%	82.35	44.01	89.32	79.87	68.91	54.54
✓	-	+ 50%	83.00	46.99	89.88	80.54	68.73	57.87
✓	-	+ 100%	83.72	48.90	90.77	79.97	69.10	61.56
✓	-	+ BT	85.31	52.07	89.97	82.21	71.97	67.02
-	✓	+ 10%	80.05	43.01	83.28	83.21	67.32	51.22
-	✓	+ 30%	81.44	47.91	88.22	80.30	68.77	55.45
-	✓	+ 50%	82.36	50.88	90.42	82.89	69.12	59.98
-	✓	+ 100%	84.36	52.91	89.07	83.30	71.66	64.10
-	✓	+ BT	86.05	54.46	91.53	86.67	73.92	69.92
✓	✓	+ 10%	81.13	43.82	84.17	78.56	68.67	53.50
✓	✓	+ 30%	84.72	48.14	88.06	81.23	69.47	58.64
✓	✓	+ 50%	85.51	52.52	89.06	82.60	71.79	63.36
✓	✓	+ 100%	84.90	56.88	90.17	83.36	74.01	66.97
✓	✓	+ BT	86.40	58.15	91.88	84.77	75.46	71.93

Table 5: Results of our ablation studies, in terms of accuracy and micro-F1 scores, when varying the amount (as percentage) of contextualized examples obtained with ABTA and EBTA that are added to the training data, where BT stands for the backtranslation augmentation approach from (Maab et al., 2023).

5 Results

5.1 Ablation Study

To show the effectiveness of our proposed techniques, we rely on both INF/OTH and INF/LEX tasks. For BANC, (ABTA and EBTA), we vary the percentage of augmented data that is added to the training, and compare against the "regular" setting. Table 5 and Figure 1 shows a summary of our obtained results. Overall, we observe that with the increase in size of context-augmented samples for both neighborhood and target-aware context, the model yields improvements in F1-scores and accuracy of both bias tasks. Furthermore, we see that by only using BANC, we can achieve substantial performance improvements, regardless of the fact that this technique neglects event information.

Owing to the fact that target-aware context contains comprehensive data augmentation contexts (ABTA & EBTA), an elevated performance in INF-F1 scores in INF/OTH and INF/LEX tasks is observed, as shown in Figure 1 with blue and olive lines, over BANC in light blue and light green lines, respectively. Higher percentage of context achieves higher performance, for instance, when 100% context of (ABTA & EBTA) is utilized, INF/OTH task shows INF-F1 score of 52.91 against 42.32 of regular, and INF/LEX task shows INF-F1 score of 71.66 against 68.71 of regular, respectively. In INF/LEX task, the rise of LEX-F1 scores highlighted in red are more prominent than INF-F1 highlighted in green after 50% context, because number of lexical bias contexts are partially comparable to informational bias contexts, whereas in INF/OTH task the informational bias contexts are still reasonably lower than OTH (non bias + lexical samples). Similarly, since backtranslation is only performed on lexical bias contexts in INF/LEX task, LEX-F1 scores are more amplified than INF F1-scores.

When we combine our neighborhood augmentation technique (BANC) and target-aware article-based and event-based contexts (ABTA and EBTA) as our final model, it is observed that the performance begin excelling against the regular even when 40% of the combined context-augmented examples are fed to the model. Our results further demonstrate the effectiveness of backtranslation-based augmentation technique on BASIL, following the findings of Maab et al. (2023), and showing that this technique can be combined with our proposed components to attain further performance improvements.

5.2 Comparison with Prior Work

Having established the efficacy of our proposed approach, we now proceed to compare our model with previous studies. Concretely, we can only compare our work against one studied INF/OTH bias task of BASIL using contextual information

Model	INF / OTH			
	Acc.	P	R	INF F1
Neighborhood Context				
SSC-5 (van den Berg and Markert, 2020)	-	41.90	36.16	38.19
SSC-10 (van den Berg and Markert, 2020)	-	43.84	34.88	38.22
WinSSC-5 (van den Berg and Markert, 2020)	-	42.28	36.94	38.67
WinSSC-10 (van den Berg and Markert, 2020)	-	43.20	35.12	37.44
RoBERTa (van den Berg and Markert, 2020)	-	43.12	41.29	42.16
MultiCTX (Guo and Zhu, 2022b)	-	47.18	44.01	45.53
BERT + BT (Maab et al., 2023)	83.86	**51.22**	46.32	50.70
BERT + BANC (ours)	83.72	49.07	45.32	48.90
BERT + BANC + BT (ours)	**85.31**	50.08	**48.12**	**52.07**
Article Context				
WinSSC (Guo and Zhu, 2022b)	-	41.47	34.37	37.58
ArtCIM (van den Berg and Markert, 2020)	-	38.81	47.78	42.80
Event Context				
EvCIM (van den Berg and Markert, 2020)	-	39.72	49.60	44.10
EvCIM (Guo and Zhu, 2022b)	-	47.07	44.64	45.81
BERT (Chen et al., 2020)	-	58.62	32.08	41.46
RoBERTa (Lei et al., 2022)	-	43.53	49.84	46.47
MultiCTX (Guo and Zhu, 2022b)	-	47.78	44.50	46.08
BERT + ABTA + EBTA (ours)	84.36	52.78	47.74	52.91
BERT + ABTA + EBTA BT (ours)	**86.05**	**54.10**	**49.82**	**54.46**
BERT + BANC + ABTA + EBTA (ours)	84.90	55.60	**53.93**	56.88
BERT + BANC + ABTA + EBTA + BT (ours)	**86.40**	**59.22**	53.12	**58.15**

Table 6: Comparison of our approach with previous work, separated by usage of context. We report average results of three runs with different random seeds. In the Table, Acc, P, and R stand for Accuracy, Precision and Recall respectively. BT denotes the augmentation approach from (Maab et al., 2023), who are also the only authors to report accuracy.

as indicated by prior work (Maab et al., 2023), therefore we solely present our work pertaining to this task with state-of-the-art.

Based on our comprehensive analysis on how prior studies use different contexts on BASIL, we align similar contexts of our proposed method to allow meaningful comparisons as shown in the Table 6, using three corresponding sections.

To compare with previous work where only within article context is used, we concretely utilize our top performing models for comparison, i.e., BERT combined with 100% BANC (BERT + BANC), and with backtranslation (BERT + BANC + BT). Similarly, prior work using event contexts are compared with our BERT model trained on 100% target-aware (BERT + ABTA + EBTA), and with backtranslation (BERT + ABTA + EBTA + BT), respectively. Since MultiCTX by Guo and Zhu (2022b) uses multi-contrast learning of both article and event contexts, we compare and use our best BERT model with fusion of both proposed context techniques (BERT + BANC + ABTA + EBTA), and with backtranslation (BERT + BANC + ABTA + EBTA + BT), which in essence is our final model. Based on our results, and supporting findings of our ablation study, both BANC and target-aware (ABTA & EBTA) hold significance in our approach, however target-aware contexts contributes more than BANC parallel to previous findings (Guo and Zhu, 2022b). Our approach outperforms previous work significantly, obtaining an F1-score of 58.15 in INF label.

In summary, our results show that our proposed approach leads to state-of-the-art results, offering compelling empirical evidence suggesting that adding multiple contextual information is effective at recognizing sentence-level informational and lexical bias as a type of misinformation.

5.3 Role of "target" frequency

To confirm the effectiveness of using target-aware (ABTA & EBTA) contexts, we conduct a study on most frequent bias targets of BASIL, and consequently experiment with BERT (Devlin et al., 2018), which serves as a baseline model for recent studies (van den Berg and Markert, 2020; Guo and Zhu, 2022b; Chen et al., 2020). From Table 4, we see that the "target" "Donald Trump" appears as the most attracted and significant media entity with substantial coverage of informational and lexical bias sentences. Out of 6,538 total target-aware contexts that we create, we found that around 42% (2,767) of them come from this target. In light of this issue, we are interested in studying the effect of target frequency in the creation of richer context, and propose an ablation analysis to gain insight.

We begin by first introducing target-aware contexts of only "Donald Trump" in various fractions to the regular setting, again for the INF/LEX task. We compare the contribution of the most frequent target towards performance by testing models trained solely on this data, and compare to models trained on the entire target-aware contexts.

As shown in Figure 2, our results are consistent with the performance rise of LEX-F1 scores after 50% data using all targets, no significant performance change is observed until 50% of Donald Trump contexts, however there is a gradual rise in F1-scores of INF and LEX when 100% contextual data is introduced. Increase in LEX F1-score of 56.72 from 53.42 is seen with 100% Donald Trump context when compared with the regular model. Similarly, due to the fact that no BT is performed on INF bias contexts, the rise of LEX-F1 scores

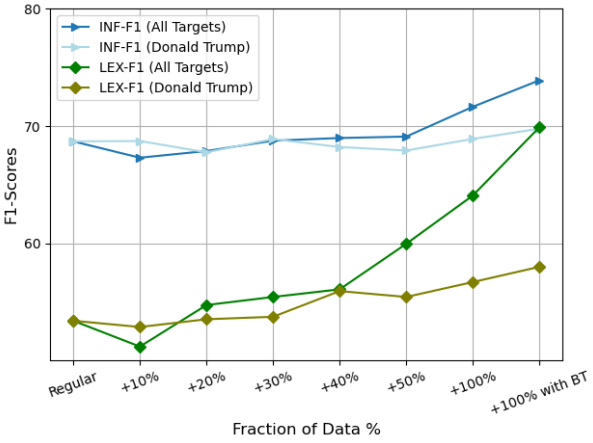

Figure 2: In INF/LEX task, plot showing comparison of performance on most frequent target "Donald Trump" v/s. All Targets-aware context (BERT + ABTA + EBTA) using INF/LEX bias task.

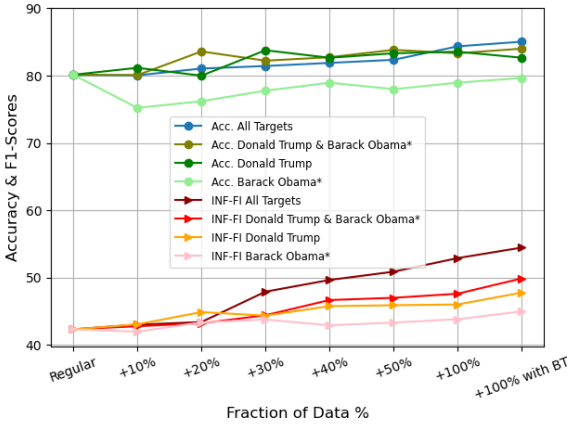

Figure 3: In INF/OTH task, plot showing comparison of F1-scrore and accuracy on target "Donald Trump", "Barack Obama*", and "Donald Trump + Barack Obama*" combined Vs. All Targets-aware context (BERT + ABTA + EBTA)

from 56.7 to 58.02 is more prominent in 100% with BT. In addition to the non-overlapping train-val-test, for this study we carefully choose testing examples so that the majority of targets have an equal %age in the test set to avoid the problems of overfitting the same target.

In addition to Donald Trump in INF/OTH task, we also introduce the second most frequent target "Barack Obama*", and the fusion of the two as shown in Figure 3. Consistent with our findings, our approach works well for even a single target like "Donald Trump" having approximately not far from half target-aware contextualized examples towards total. Following prior work (Maab et al., 2023), we also provide our model with back translated examples of Donald Trump, hence doubling up context examples from 2,767 to 5,534. Compared to the context free regular model, the best performance of INF F1-score of 46.01 from 42.32 is achieved, whereas through back translation INF F1-score further increased from 46.01 to 47.77 using Donald Trump. The combined context of "Donald Trump" and "Barack Obama*" is also examined for further confirmation of our proposed target-aware context approach which result in improved performance over a single target "Donald Trump". This study confirms the general nature of our approach in detecting different types of bias, since it is not uncommon in real world scenarios to run into similar and parallel target entities as reported by various media outlets (Arapakis et al., 2016; Lim et al., 2020).

Furthermore, the incorporation of appropriate context in training samples serves significantly in enhancing the model's performance. To further illustrate this, Table 7 shows two examples of combined local (article-based) and global (event-based) contexts of informational and lexical bias for targets Adam Schiff and Liz Cheney, respectively. It can be seen that meaningful combination of local bias-sensitive contexts and a target-aware context approach in the examples are combined with target sentences which enables the model to detect various types of bias with increased precision, as shown by our results.

6 Conclusion

We study a challenging and significant task of detecting misinformation and shed light on bias prevalence in news media. Our work focuses on incorporating bias sensitive (BANC) and target-aware contexts (ABTA & EBTA) for sentence-level bias detection tasks. Our proposed approach exploits the distinct influence of informational and lexical bias in news media writing styles, emulating the principle of human learning. Our model encompass the process by which individuals acquire new knowledge in real-world settings, i.e., gathering the associated type of bias from common news media targets covering the same event coupled with past experiences, and subsequently utilizing such contexts to make predictions about unfamiliar aspects.

Our model concretely outperforms classification performance of strong baselines in all bias tasks

Source	Target	Local Index	Global Index	Sentence	Bias
FOX	Adam Schiff	3, 5	4	Your actions both past and present are incompatible with your duty as Chairman of this Committee, the letter stated. **We have no faith in your ability to discharge your duties in a manner consistent with your Constitutional responsibility and urge your immediate resignation as Chairman of this Committee.**The letter follows the conclusion of Special Counsel Robert Mueller's Russia probe, which turned up no evidence of collusion between Trump campaign members and Russia during the 2016 presidential election.	Inf
HPO		5	6	It doesn't appear that was any part of [special counsel Robert] Mueller's report. **In a letter dated Thursday, the GOP committee members accused Schiff of standing at the center of a well-orchestrated media campaign about a possible Trump-Russia connection.**	Inf
NYT		19, 21	20	They say Democrats will stop at nothing to ruin his presidency, and bristle at Democrats accusing them of turning a blind eye to the Russian threat. **And at the center of their wrath is Mr. Schiff, whose doughy-faced demeanor hardly evokes an attack dog.** The findings of the special counsel conclusively refute your past and present assertions and have exposed you as having abused your position to knowingly promote false information, having damaged the integrity of this committee, and undermined faith in U.S. government institutions, Representative K. Michael Conaway, Republican of Texas, said to Mr. Schiff.	Inf
FOX	Liz Cheney	-	-	-	-
HPO		9, 11	10	Previously, she had spoken out against the military's former "don't ask, don't tell" policy. **According to The Hill, Cheney sought to clarify her position after an alleged poll in Wyoming said she "supports abortion and aggressively promotes gay marriage.** Her opposition also puts her at odds with her father, who offered support for gay marriage in 2009.	Lex
NYT		7, 9	8	That position deferring to the will of the voters on a state-by-state basis may represent something of a compromise between total support or opposition. **But it did little to placate her sister.** It's not something to be decided by a show of hands, Mary Cheney wrote.	Lex

Table 7: Combined bias sentence examples with three news media sources extracted from event 7 (7fox, 7hpo, 7nyt) for informational bias and event 33 (33fox, 33hpo, 33nyt) for lexical bias showing the influence of local and target-aware global contexts that aids the model in effectively determining bias. In this example we see how sentences in bold, representing bias target sentences with global indexing (event-based), are harmoniously integrated with contextual information from neighboring sentences (local indexing, i.e., preceding and subsequent sentences within the article.)

and we provide statistical significance of our proposed components through extensive experiments. We find that the best performance is achieved when target-aware contexts are combined with BANC, and our methodological stand-point in using small-augmented data of frequent targets suggests that our model is better at recognising bias in mass media. In addition, we conclude its important to keep different bias separately for accurate prediction of bias and we intend to explore other bias features as part of future work. Consequently, future work could also extend contextual information to other misinformation tasks.

Limitations

Bias can vary based on human perspectives and existing NLP models have limitations to interpret the subjective nature of bias. Due to the lack of bias representations and annotated media coverage in other languages, our work is based only on English news articles. To the best of our knowledge, BASIL is the only dataset annotated with informational bias, and although our approach provides valuable insights and findings on detecting bias, we provide no evidence to suggest the significance of our findings regarding other contexts surrounding bias or misinformation detection tasks. Similarly, due to the disproportionate number of political ideologies in our dataset, we cannot say for sure if our model will perform equally well for other tasks, and we believe this requires further analysis.

Ethical Considerations

In this work, since we highlight some frequent bias targets in political news to propose the significance of our approach, we do not intend to promote media bias entities rather we advocate media literacy and ethical journalism practices. Further, the results we reported in our work highlight deeper understanding of bias contexts, and the need for bias mitigation at various levels of the mass media.

Acknowledgements

The authors wish to express gratitude to the funding organization as this work has been supported by the Mohammed bin Salman Center for Future Science and Technology for Saudi-Japan Vision 2030 at The University of Tokyo (MbSC2030).

References

Muhammad Pervez Akhter, Zheng Jiangbin, Irfan Raza Naqvi, Mohammed Abdelmajeed, Atif Mehmood, and Muhammad Tariq Sadiq. 2020. Document-level text classification using single-layer multisize filters convolutional neural network. *IEEE Access*, 8:42689–42707.

Ioannis Arapakis, Filipa Peleja, Barla Berkant, and Joao Magalhaes. 2016. Linguistic benchmarks of online news article quality. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1893–1902.

Ramy Baly, Georgi Karadzhov, Abdelrhman Saleh, James Glass, and Preslav Nakov. 2019. Multi-task ordinal regression for jointly predicting the trustworthiness and the leading political ideology of news media. *arXiv preprint arXiv:1904.00542*.

Wei-Fan Chen, Khalid Al-Khatib, Benno Stein, and Henning Wachsmuth. 2020. Detecting media bias in news articles using gaussian bias distributions. *arXiv preprint arXiv:2010.10649*.

Wei-Fan Chen, Henning Wachsmuth, Khalid Al Khatib, and Benno Stein. 2018. Learning to flip the bias of news headlines. In *Proceedings of the 11th International conference on natural language generation*, pages 79–88.

Arman Cohan, Iz Beltagy, Daniel King, Bhavana Dalvi, and Dan Weld. 2019a. Pretrained language models for sequential sentence classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. Association for Computational Linguistics.

Arman Cohan, Iz Beltagy, Daniel King, Bhavana Dalvi, and Daniel S Weld. 2019b. Pretrained language models for sequential sentence classification. *arXiv preprint arXiv:1909.04054*.

Stefano DellaVigna and Ethan Kaplan. 2007. The fox news effect: Media bias and voting. *The Quarterly Journal of Economics*, 122(3):1187–1234.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*.

Robert M Entman. 2007. Framing bias: Media in the distribution of power. *Journal of communication*, 57(1):163–173.

Hugging Face. 2021. The ai community building the future. *URL: https://huggingface. co*.

Lisa Fan, Marshall White, Eva Sharma, Ruisi Su, Prafulla Kumar Choubey, Ruihong Huang, and Lu Wang. 2019. In plain sight: Media bias through the lens of factual reporting. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 6343–6349, Hong Kong, China. Association for Computational Linguistics.

Matthew Gentzkow, Jesse M Shapiro, and Daniel F Stone. 2015. Media bias in the marketplace: Theory. In *Handbook of media economics*, volume 1, pages 623–645. Elsevier.

Tim Groseclose and Jeffrey Milyo. 2005. A measure of media bias. *The quarterly journal of economics*, 120(4):1191–1237.

Shijia Guo and Kenny Q. Zhu. 2022a. Modeling multi-level context for informational bias detection by contrastive learning and sentential graph network.

Shijia Guo and Kenny Q Zhu. 2022b. Modeling multi-level context for informational bias detection by contrastive learning and sentential graph network. *arXiv preprint arXiv:2201.10376*.

Felix Hamborg, Karsten Donnay, and Bela Gipp. 2019. Automated identification of media bias in news articles: an interdisciplinary literature review. *International Journal on Digital Libraries*, 20(4):391–415.

Christoph Hube and Besnik Fetahu. 2019. Neural based statement classification for biased language. In *Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining*. ACM.

Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik. 2014. Political ideology detection using recursive neural networks. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1113–1122, Baltimore, Maryland. Association for Computational Linguistics.

Vivek Kulkarni, Junting Ye, Steven Skiena, and William Yang Wang. 2018. Multi-view models for political ideology detection of news articles. *arXiv preprint arXiv:1809.03485*.

Yuanyuan Lei, Ruihong Huang, Lu Wang, and Nick Beauchamp. 2022. Sentence-level media bias analysis informed by discourse structures. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 10040–10050.

Sora Lim, Adam Jatowt, Michael Färber, and Masatoshi Yoshikawa. 2020. Annotating and analyzing biased sentences in news articles using crowdsourcing. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 1478–1484.

Iffat Maab, Edison Marrese-Taylor, and Yutaka Matsuo. 2023. An effective approach for informational and lexical bias detection. In *Proceedings of the Sixth Fact Extraction and VERification Workshop (FEVER)*, pages 66–77.

Martin Marinov and Alexander Efremov. 2019. Representing character sequences as sets : A simple and intuitive string encoding algorithm for nlp data cleaning. In *2019 IEEE International Conference on Advanced Scientific Computing (ICASC)*, pages 1–6.

Tomas Mikolov, Quoc V Le, and Ilya Sutskever. 2013. Exploiting similarities among languages for machine translation. *arXiv preprint arXiv:1309.4168*.

Jeff Z Pan, Siyana Pavlova, Chenxi Li, Ningxi Li, Yangmei Li, and Jinshuo Liu. 2018. Content based fake news detection using knowledge graphs. In *The Semantic Web–ISWC 2018: 17th International Semantic Web Conference, Monterey, CA, USA, October 8–12, 2018, Proceedings, Part I 17*, pages 669–683. Springer.

Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2017. Automatic detection of fake news. *arXiv preprint arXiv:1708.07104*.

Andrea Prat and David Strömberg. 2013. The political economy of mass media. *Advances in economics and econometrics*, 2:135.

Guozheng Rao, Weihang Huang, Zhiyong Feng, and Qiong Cong. 2018. Lstm with sentence representations for document-level sentiment classification. *Neurocomputing*, 308:49–57.

Marta Recasens, Cristian Danescu-Niculescu-Mizil, and Dan Jurafsky. 2013. Linguistic models for analyzing and detecting biased language. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1650–1659, Sofia, Bulgaria. Association for Computational Linguistics.

Abinash Tripathy, Abhishek Anand, and Santanu Kumar Rath. 2017. Document-level sentiment classification using hybrid machine learning approach. *Knowledge and Information Systems*, 53:805–831.

Unesco. 2018. Journalism,'fake news'& disinformation: handbook for journalism education and training. *UNESCO*.

Esther van den Berg and Katja Markert. 2020. Context in informational bias detection. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 6315–6326, Barcelona, Spain (Online). International Committee on Computational Linguistics.

Xiang Zhou and Mohit Bansal. 2020. Towards robustifying nli models against lexical dataset biases. *arXiv preprint arXiv:2005.04732*.