Bias Bubbles: Using Semi-Supervised Learning to Measure How Many Biased News Articles Are Around Us

Qin Ruan[1,3][0000−0001−5822−9260], Brian Mac Namee[1,2,3][0000−0003−2518−0274], and Ruihai Dong[1,2,3][0000−0002−2509−1370]

[1] School of Computer Science, University College Dublin, Dublin, Ireland
[2] Insight Centre for Data Analytics
[3] Science Foundation Ireland Centre for Research Training in Machine Learning
qin.ruan@ucdconnect.ie, {brian.macnamee, ruihai.dong}@ucd.ie

Abstract. The proliferation of web 2.0 technology allows us to easily create and share online content, but also leads to the rapid spread of misinformation and biased media, which has considerable negative effects on society. Deep learning-based classifiers are one common way of identifying media bias, but they suffer from a lack of large-scale labelled datasets. In this paper, we first explore the use of pseudo-labelling technology to mitigate this problem. Second, we exploit a masking method to identify biased sentences in news articles by iteratively masking each sentence from an article and observing the change in output of a bias detection model. These identified sentences not only contribute to evaluating the proposed model, but also enable end-users to understand where media bias arises in an article. Finally, we apply our well-trained bias detection model to a well-known news article dataset to show how widespread media bias is—the results show that it is rampant and has become a serious social problem that we cannot ignore.

Keywords: Media Bias · Pseudo-labelling · Semi-supervised Learning

1 Introduction

Online news websites are effective news transmission platforms, however, studies have shown that media bias is widespread in them, and is caused by inherent flaws in the news production process [2, 14]. The side effects of media bias—such as distorting readers' perception and negatively influencing social decision-making—have been widely recognized by social scientists [3]. In computer science solutions have been explored to identify media bias automatically—from traditional lexicon-based algorithms [8] to more recent deep learning-based models [4]. However, accurately detecting media bias in news articles and evaluating the degree of media bias that exists in our society remain significant challenges.

Some inherent characteristics of media bias are a major cause of these challenges. First, the forms of media bias are variant, such as using a tendentious or inflammatory vocabulary, adopting different writing styles, or reporting an event only in favour of one side. Second, bias is not a problem of honesty of reporting but journalists' own preferred opinions [9], and usually the bias is subtle rather than explicit because it is easier to affect unsuspecting readers that way [5]. These characteristics not only make media bias recognition more challenging than other text classification tasks, but also increase the difficulty and the cost of manually labelling news articles for media bias. Therefore, the scale of datasets released for media bias detection is usually quite small. For example, the Annotated Data dataset [16] contains only 46 news articles and 1,235 sentences from 4 news events. The lack of large-scale labelled data prevents researchers from adopting sophisticated models to improve classification accuracy.

In this paper we address these challenges in three ways. First, we explore the simultaneous use of unlabelled or machine labelled data (large-scale but with a lot of noise) and human labelled data (small-scale with better quality) by using two pseudo-labelling algorithms to augment training datasets containing the latter with the former. Second, we verify the generalization ability of bias detectors trained in the previous step by observing their performance on an unseen dataset. We also exploit a masking approach [22] to identify biased sentences by iteratively masking each sentence and making a comparison with human labels to further evaluate the proposed models. Finally, we leverage two well-trained bias detectors to analyze media bias in a large-scale news article dataset, MIND [20]. The results show that media bias is a widespread phenomenon and has become a serious social problem that we cannot ignore. The workflow followed in this paper is illustrated in Fig. 1.

2 Related Work

Recently computational approaches based on natural language processing and machine learning have been employed to detect biased news articles. A common perspective is to regard biased news detection as a text classification task. Therefore, feature mining and classifiers employed in text classification tasks can also be applied to biased news recognition. Kiesel et al. [13] noted how many entries into the Semeval-2019 task on *Hyperpartisan news detection* used standard text mining methods—including word n-gram, word embeddings, stylometric

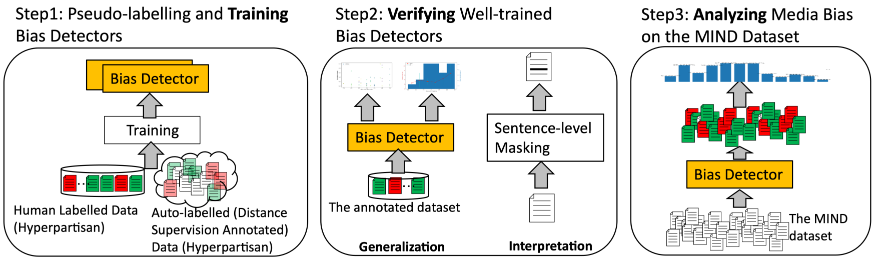

Step1: Pseudo-labelling and Training Bias Detectors
Bias Detector
Training
Human Labelled Data (Hyperpartisan)
Auto-labelled (Distance Supervision Annotated) Data (Hyperpartisan)
Step2: Verifying Well-trained Bias Detectors
Bias Detector
The annotated dataset
Sentence-level Masking
Generalization
Interpretation
Step3: Analyzing Media Bias on the MIND Dataset
Bias Detector
The MIND dataset

Fig. 1: The workflow of analyzing media bias on the MIND dataset.

features, sentiment and emotion features, and recognition of named entities in the news. However, in the Semeval-2019 task deep neural networks that adopt the most current trends (largely based on transformer models [19]) were the most common approaches used by competition teams [13].

Another type of approach deals with news bias, from discovering news bias texts to locating biased information. Some views find potential opinions by evaluating expressions of "bias target". For example, [17] designed a method called "stakeholder mining" because they treat important entities as stakeholders in the text. Similarly, [10] extract frame attributes and target words or phrases related to frame topics from political news articles.

Labelled training data is a prerequisite for applying machine learning to media bias detection. In the past two years, a number of important manually labelled biased news article datasets have been released, including the SemEval 2019 Task4 [13], Ukraine Crisis [6], NewsWCL50 [10], Annotated Data [16], and BASIL [5] datasets. These datasets cover news articles across different areas and feature different kinds of bias, e.g., the SemEval 2019 Task 4 dataset includes articles labelled from the perspective of political ideologies; and the Ukraine Crisis dataset identifies bias derived from different countries.

Based on the annotation granularity, these datasets can be mainly divided into three groups: article level, sentence level, and word group level. For example, the SemEval 2019 Task 4 dataset is the largest article-level news dataset, and contains 1,273 manually labelled news articles, each categorised as biased or unbiased. In the Ukraine crisis dataset, Färber et al. [6] extracted 90 news articles with a total of 2,057 sentences and labelled the data at both article and sentence level from multiple perspectives, such as subjectivity and the presence of hidden assumptions. The Annotated Data dataset is another sentence-level dataset including 46 news articles (made up of 1,235 sentences) [16] covering 4 news events. However, the major limitation of these manually labelled datasets is their small scale.

Researchers have studied automated labelling technologies to address the limited size of manually labelled datasets. Distant supervision is a popular technique for annotating datasets in the context of media bias detection. In this approach news articles are labelled not based on the detailed content of the articles themselves but rather based on the characteristics (e.g., political leaning) of their publisher. The SemEval 2019 Task4 dataset [13] also contains a large corpus for identifying hyper-partisanship, which has 754,000 news articles labelled via distant supervision. However, a recent study [1] shows that these types of datasets are very noisy and it is not yet clear how that can best be utilized in media bias detection tasks.

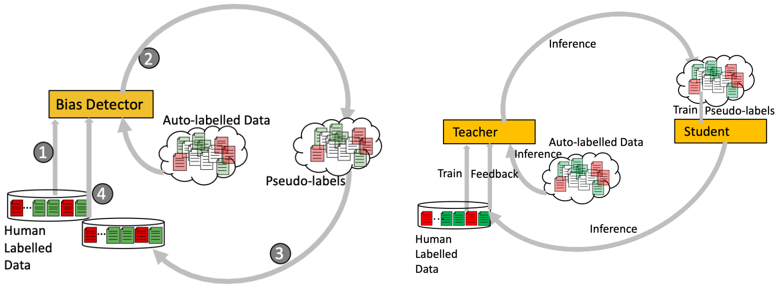
Self-training methods form a branch of semi-supervised learning, and leverage the probability output of a model to generate pseudo-labels for unlabelled data [23]. This approach can easily add more input data to help train a model. Due to its simplicity and effectiveness, self-training has been successfully used in various tasks. The entropy minimization (EntMin) method [7] encourages a model to make low-entropy predictions on unlabelled data through entropy regularization, and then employs qualified unlabelled data in standard supervised learning settings. Another simple and effective way to train neural networks in a semi-supervised way is pseudo-labelling [15]. A neural network model trained using labelled data through supervised learning directly predicts pseudo-labels for instances in the unlabelled dataset and these are then used along with the labelled data to retrain the model. Inspired by knowledge distillation, the noisy student approach [21] is a supervised method that transfers the knowledge of a teacher model to an equivalent or larger student model. The teacher model is first trained on labelled data to generate pseudo-labels for unlabelled examples. Then the equivalent or larger student model uses the knowledge of the teacher model to train on the labelled and pseudo-labelled data.

3 Pseudo-labelling Enhanced Bias Detectors

In this section, we present and evaluate our two pseudo-labelling frameworks: *Overlap-checking* and *Meta-learning*. We use the network from Jiang et al. [12] as our backbone model, as it is one of the best models submitted to the leaderboard of SemEval 2019 Task 4[4]. Baly et al. [1] showed that the top models on this leaderboard are trained purely on manual labelled articles. In this experiment, we demonstrate how to utilize by-publisher data (through distant supervision) in the training process via our pseudo-labelling frameworks and evaluate their performance on the SemEval 2019 Task4 hyperpartisan news detection dataset.

3.1 Pseudo-labelling Frameworks

This section describes our two pseudo-labelling frameworks: *Overlap-checking* and *Meta-learning*.



(a) Overlap Selection (b) Meta-Learning

Fig. 2: An overview of the overlap-checking and meta-learning frameworks

[4] https://pan.webis.de/semeval19/semeval19-web/#results

Overlap-checking. An overview of the proposed overlap-checking framework is presented in Fig. 2(a). The framework contains four steps: (1) the network is first trained on manually labelled data until it converges; (2) the training leverages the overlap-checking mechanism to select a batch of pseudo-labelled data; (3) new data is generated using labelled data and pseudo-labelled data; and (4) the model is re-trained on new data. For unlabelled samples, the pseudo-labelling method is used to label the data based on the probability distribution of the model prediction [15].

The overlap-checking method belongs to the branch of semi-supervised learning. Using this simple and efficient method, the system can easily add more data to help re-train the model. He & Sun [11] proved that using a batch of samples with the highest prediction probability of the model can help enhance the performance of the model. In the overlap-checking framework, the vanilla pseudo-labelling method selects the class with the highest predicted probability from the completely unlabelled dataset as the pseudo label of the sample. Assuming that there are L classes, denoted l as category instance, where a value of 1 indicates that category l is selected and 0 not selected, the formula is as follows:

$$y^{'} = \begin{cases} 1 & \text{if } l = \arg\max_{l \in L} f(x)^{'} \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

The system then combines both the pseudo-labelled annotation and the distant supervision annotation on the by-publisher dataset by considering their consistency. We denote the distant supervision dataset as A, the pseudo-labelled dataset as P, and the intersection set of A and P as candidate set $C = A \cap P$. Eventually, the top N pseudo-labelled samples are returned on the basis of descending order of the predicted probability value, where N represents the expected number of pseudo-samples.

Meta-learning. The meta-learning framework takes inspiration from the meta pseudo-labels approach [18]. The workflow of the meta-learning framework is shown in Fig. 2(b). The pseudo-labelling method maintains a network to be trained sequentially on a clean dataset and a pseudo dataset. Unlike the vanilla pseudo-labelling method, meta-pseudo-labelling trains the teacher network and the student network in parallel. The teacher network updates its own information from two aspects: a signal from the annotated data and feedback from the student network. Acquiring teacher network signals is by the standard process of training a supervised learning model based on annotated datasets. Getting feedback from the student network requires that the student network inherits the same network structure as the teacher network, but the update of the student network is based on noisy data.

3.2 Experiments and Evaluation

This section describes the evaluation experiment designed to assess the performance of the Overlap-checking and Meta-learning methods.

Table 1: Statistics of the Semeval-2019 Task 4 hyperpartisan news dataset.

Labelling type	#biased	#unbiased
By-article labelling	238	407
By-publisher (distant) labelling	15,008	14,992

Semeval-2019 Task 4 [13] focuses on detecting if a news article contains biased information. Released along with the competition is an article-level hyperpartisan news bias dataset. The dataset includes 1,273 manually labelled samples (of which 628 are kept private for evaluation) and 754,000 automatically labelled samples based on publisher attributes. We use the the hyperpartisan dataset as the training dataset for our models. We collect all published manually labelled samples and 30,000 samples selected randomly from the automatically labelled dataset. A summary of the training dataset is shown in Table 1.

The solution of Semeval-2019 Task 4 winning team [12], is employed as the base detector in our approaches. This builds an Elmo-based sentence encoder to encode sentences to high-dimensional semantic vectors which are passed into differently initialized convolutional layers and batch normalization layers in parallel. The final output is a dense layer followed by a sigmoid function that concatenates output from the previous layers.

The training details follow the same configuration on the data processing and network side. Unlike the approach by Jiang et al. [12] that only uses manually labelled data, we improve performance by adding the data enhancement module to leverage the noisy by-publisher distant labels.

We conduct detailed comparisons of different data strategies combined with the bias detector, recording results in Table 2. The results in Table 2 are based on 10-fold cross-validation and bias detection performance is measured using accuracy, precision, recall and the F1 score. The Bias detector is precisely the same configuration as the original version [12].[5]

The overlap-checking and meta-learning methods are used with addition of distant supervised data to increase the size of the data by 1x, 2x, and 3x. Overlap-checking has the best performance in the case of using equal proportions, in which accuracy, precision and F1 score respectively exceed the baseline model. Providing more distant supervised data for the meta-learning method leads to better accuracy, recall and F1 score. However, with the addition of more data, the precision of the meta-learning model decreases.

4 Evaluating Generalization

In the previous section, we demonstrated the effectiveness of the proposed solutions on the Semeval-2019 hyperpartisan news dataset. We are interested in

[5] We re-implement the Jiang et al. approach [12] in PyTorch, and our accuracy score is 0.0116 higher than what they reported, which we assume is due to different initializations.

Table 2: The validation performance of different combinations on the Semeval-2019 Task 4 hyperpartisian news detection task.

Approach	Accuracy	Precision	Recall	F1
Baseline model	0.852±0.074	0.824±0.077	0.767±0.194	0.780±0.144
Overlap-checking (1:1)	0.867±0.052	0.863±0.071	0.760±0.132	0.803±0.091
Overlap-checking (1:2)	0.848±0.053	0.820±0.081	0.759±0.141	0.782±0.093
Overlap-checking (1:3)	0.854±0.063	0.826±0.083	0.764±0.122	0.791±0.096
Meta-learning (1:1)	0.856±0.066	**0.878±0.109**	0.725±0.175	0.777±0.134
Meta-learning (1:2)	0.864±0.066	0.853±0.081	0.760±0.160	0.797±0.120
Meta-learning (1:3)	**0.870±0.070**	0.840±0.102	**0.806±0.158**	**0.815±0.113**

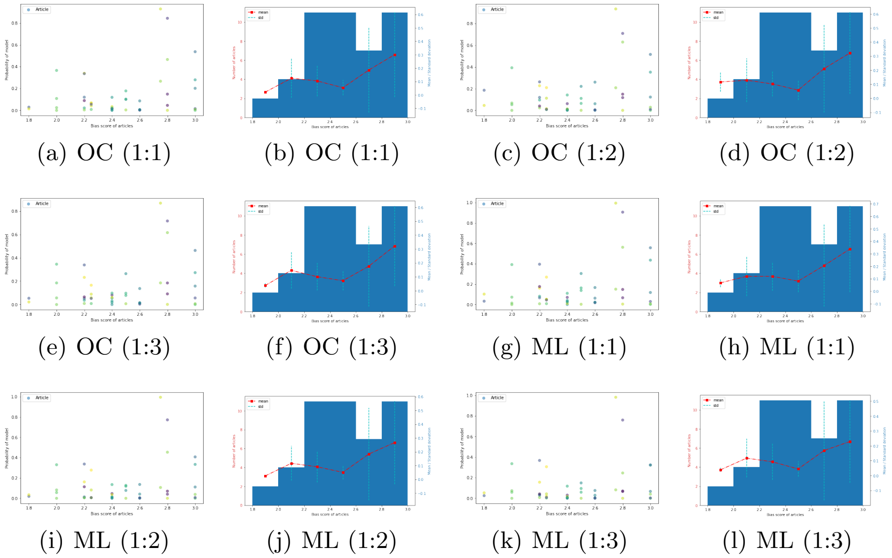
whether these trained detectors can generalise to other news article datasets to assess the degree of media bias in them, and also whether these models have the ability to recognize biased sentences within news articles. To address these two questions, we conduct experiments to evaluate the trained biased news detectors on a completely unseen dataset, the Annotated Data dataset [16], that contains article-level and sentence-level manual annotations.

4.1 The Annotated Data Dataset

The Annotated Data Dataset is a fine-grained news bias dataset [16]. Annotators have evaluated the degree of bias at the article and sentence level. Four to five annotators provide a bias score for each sample. The scores range from one (not biased) to four (very biased). To use the Annotated Data Dataset to conduct a generalization experiment, we assign a bias score to each sample (article or sentence) by aggregating the annotators' scores using the mean.

4.2 Generalization Experiment

Evaluating the performance of the model trained on the binary Semeval-2019 hyperpartisan news detection dataset directly on the Annotated Dataset is complicated because the labels in the Annotated Data Dataset indicate a degree of bias from one to four. The output of the trained bias detector, however, is a continuous probability of bias value between zero and one. We can, therefore, measure the generalization ability of the bias detection model by measuring the correlation between the model outputs and the aggregated bias scores. The first and third columns of Fig. 3 show scatter plots of aggregated human annotated bias degree (horizontal axis) and the probability of bias predicted by a model (vertical axis) for models trained using overlap-checking (OC) and meta-learning (ML) with different degrees of data augmentation (1x, 2x, and 3x). These plots show that humans score towards the biased direction for an article whose predicted results exceeds 0.5. Similarly, for articles that human annotators tend to



(a) OC (1:1) (b) OC (1:1) (c) OC (1:2) (d) OC (1:2)

(e) OC (1:3) (f) OC (1:3) (g) ML (1:1) (h) ML (1:1)

(i) ML (1:2) (j) ML (1:2) (k) ML (1:3) (l) ML (1:3)

Fig. 3: The leftmost and third column figures show the correlation between the bias score of articles calculated by annotators(x-axis) and the probability of models(y-axis). The second and rightmost column figures compare the trending of the probability of models with articles' bias score deviation.

be rate as unbiased, the prediction of the detectors is also between 0-0.5. There is, however, disagreement in the lower right corner of the plots.

The second and fourth columns of Fig. 3 show a trend analysis, where the degree of bias annotated by humans (horizontal axis) and the number of articles (left vertical axis) forms a histogram. The right vertical axis shows the mean and standard deviation of the probability of bias outputs in each bin. We see that as the biased score annotated by humans increases, the probability predicted by models also increases, showing a strong positive correlation between these two.

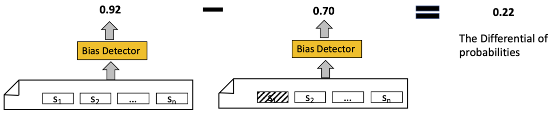
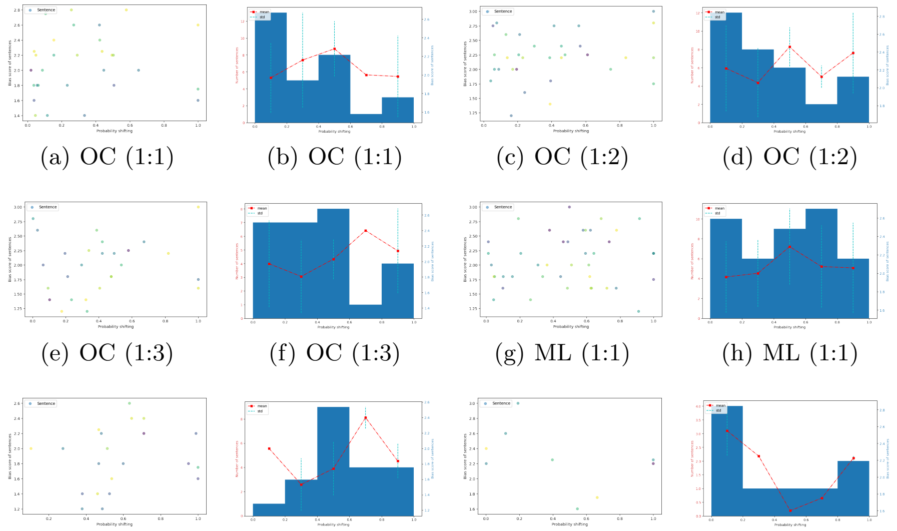


Fig. 4: The sentence masking approach for identifying influential setences.

4.3 Sentence Masking Experiment

A neural-network-based bias detector outputs a probability of bias, $p_{initial}$, when presented with an article as input. Inspired by [22], to measure the impact of a specific sentence, s_i, on the output of the model we can mask s_i out of the article, recompute the output of the model, p_{s_i}, and calculate p_{shift}, the difference between this and the original probability of bias: $p_{shift} = p_{initial} - p_{si}$. Fig. 4 illustrates this process.

The scatter plots in the first and third columns of Fig. 5 shows the relationship between p_{shift} values (horizontal axis) and human annotated bias values for sentences (vertical axis). The second and fourth columns of Fig. 5 illustrate the changes in bias scores of sentences under different probability intervals. These plots show that there is a certain positive correlation between the human-annotated bias scores of the sentences and their p_{shift} values.



(a) OC (1:1) (b) OC (1:1) (c) OC (1:2) (d) OC (1:2)

(e) OC (1:3) (f) OC (1:3) (g) ML (1:1) (h) ML (1:1)

(i) ML (1:2) (j) ML (1:2) (k) ML (1:3) (l) ML (1:3)

Fig. 5: The leftmost and third column figures show the correlation between the bias score of articles calculated by annotators(y-axis) and the probability shifting of models(x-axis). The second and rightmost column figures compare the trending of the article's bias score with the probability deviation.

5 Analyzing Media Bias in the MIND Dataset

This section analyzes Microsoft's large-scale English-based news recommendation dataset, MIND [20], which includes one million users and more than 160k news articles, and is widely employed in news-related academic research. Each article in the MIND dataset has a wealth of information associated with it: news id, title, summary, body URL and category. The number of articles used in our experiment is less than the total number in the MIND dataset as we removed articles that failed to get the body. The overlap-checking method and meta-learning method were used to infer the number of biased articles in the MIND dataset. The number of biased articles detected by each of the approaches is presented in Table 3. A more detailed analysis of the amount of bias in different news categories from the MIND dataset is shown in Fig. 6. From Table 3, we observe that the overlap-checking method identified 8.8804% of news as biased, while the meta-learning method is more conservative. Fig. 6 shows the number of biased articles detected by each approach—overlap-checking, and meta-learning—for different categories of news article in the MIND dataset. In the weather, music, and travel categories, the amount of biased news detected is relatively low, while in health, food and drink, as well as lifestyle, the amount of biased content detected is relatively high. In sports, there is quite a difference between the numbers of biased articles detected, which requires further in-depth analysis.

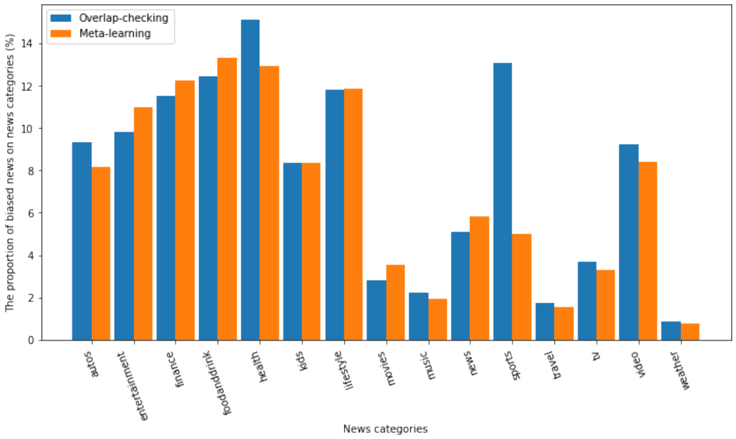


Fig. 6: The proportion of biased news calculated by base detector, overlap-checking framework and meta-learning framework, respectively.

6 Conclusion

In this work, we propose two pseudo-labelling based solutions—overlap-checking and meta-learning—to augment training sets with noisy automatically labelled data when training media bias detection models. The experimental results show that these data augmentation strategies have a positive effect on model performance. To validate the generalization capability of the proposed models, we

Table 3: The overall bias result on the bias detectors.

Method	#total	#biased	#unbiased	#proportion(%)
Overlap-checking (1:1)	122,134	10,846	111,288	8.8804
Meta-learning (1:3)	122,134	7,771	114,363	6.3627

re-evaluate them on a completely unseen dataset. The results show that the probability of bias values output by the models is highly consistent with human annotations. In addition, we exploit a masking method to identify essential sentences that affect the model's decision-making. The comparison results show a partial overlap trend between the biased sentences recognized by annotators and identified by models. Finally, we infer the amount of biased news in a large-scale news article dataset MIND. This shows that media bias is widespread–over 8% of all articles (and in some categories as much as 15%) are biased. In the future, we plan to look more in-depth at the MIND dataset to further understand the degree of bias that it contains.

Acknowledgements. This publication has emanated from research conducted with the financial support of Science Foundation Ireland under Grant number 18/CRT/6183. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

[1] Ramy Baly et al. "We Can Detect Your Bias: Predicting the Political Ideology of News Articles". In: *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 2020, pp. 4982–4991.

[2] David P Baron. "Persistent media bias". In: *Journal of Public Economics* 90.1-2 (2006), pp. 1–36.

[3] Dan Bernhardt, Stefan Krasa, and Mattias Polborn. "Political polarization and the electoral effects of media bias". In: *Journal of Public Economics* 92.5-6 (2008), pp. 1092–1104.

[4] Yahui Chen. "Convolutional neural network for sentence classification". MA thesis. University of Waterloo, 2015.

[5] Lisa Fan et al. "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)*. 2019, pp. 6343–6349.

[6] Michael Färber et al. "A Multidimensional Dataset Based on Crowdsourcing for Analyzing and Detecting News Bias". In: *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*. 2020, pp. 3007–3014.

[7] Yves Grandvalet, Yoshua Bengio, et al. "Semi-supervised learning by entropy minimization." In: *CAP* 367 (2005), pp. 281–296.

[8] Stephan Greene and Philip Resnik. "More than words: Syntactic packaging and implicit sentiment". In: *Proceedings of human language technologies: The 2009 annual conference of the north american chapter of the association for computational linguistics*. 2009, pp. 503–511.

[9] Tim Groseclose and Jeffrey Milyo. "A measure of media bias". In: *The Quarterly Journal of Economics* 120.4 (2005), pp. 1191–1237.

[10] Felix Hamborg, Anastasia Zhukova, and Bela Gipp. “Automated identification of media bias by word choice and labeling in news articles”. In: *2019 ACM/IEEE Joint Conference on Digital Libraries (JCDL)*. IEEE. 2019, pp. 196–205.

[11] Hangfeng He and Xu Sun. “A unified model for cross-domain and semi-supervised named entity recognition in chinese social media”. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. Vol. 31. 1. 2017.

[12] Ye Jiang et al. “Team bertha von suttner at semeval-2019 task 4: Hyperpartisan news detection using elmo sentence representation convolutional network”. In: *Proceedings of the 13th International Workshop on Semantic Evaluation*. 2019, pp. 840–844.

[13] Johannes Kiesel et al. “Semeval-2019 task 4: Hyperpartisan news detection”. In: *Proceedings of the 13th International Workshop on Semantic Evaluation*. 2019, pp. 829–839.

[14] Juhi Kulshrestha et al. “Search bias quantification: investigating political bias in social media and web search”. In: *Information Retrieval Journal* 22.1 (2019), pp. 188–227.

[15] Dong-Hyun Lee et al. “Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks”. In: *Workshop on challenges in representation learning, ICML*. Vol. 3. 2. 2013, p. 896.

[16] Sora Lim et al. “Annotating and analyzing biased sentences in news articles using crowdsourcing”. In: *Proceedings of the 12th Language Resources and Evaluation Conference*. 2020, pp. 1478–1484.

[17] Tatsuya Ogawa, Qiang Ma, and Masatoshi Yoshikawa. “News bias analysis based on stakeholder mining”. In: *IEICE transactions on information and systems* 94.3 (2011), pp. 578–586.

[18] Hieu Pham et al. “Meta pseudo labels”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021, pp. 11557–11568.

[19] Ashish Vaswani et al. “Attention is all you need”. In: *Advances in neural information processing systems*. 2017, pp. 5998–6008.

[20] Fangzhao Wu et al. “Mind: A large-scale dataset for news recommendation”. In: *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 2020, pp. 3597–3606.

[21] Qizhe Xie et al. “Self-training with noisy student improves imagenet classification”. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2020, pp. 10687–10698.

[22] Linyi Yang et al. “Generating Plausible Counterfactual Explanations for Deep Transformers in Financial Text Classification”. In: *Proceedings of the 28th International Conference on Computational Linguistics*. 2020, pp. 6150–6160.

[23] Xiangli Yang et al. “A Survey on Deep Semi-supervised Learning”. In: *arXiv preprint arXiv:2103.00550* (2021).