Information Engineering Express
International Institute of Applied Informatics
2024, Vol. 10, No. 1, IEE786

Characteristics of Datasets for Fake News Detection to Mitigate Domain Bias

Shingo Kato [*] , Linshuo Yang [*] , Daisuke Ikeda [*]

Abstract

ñFake newsò, news intentionally containing false information, has become quite com-mon and often causes social disruption. Many researches on automatic detection of fake have been extensively studied. The classification accuracy is improving, but a major challenge for practical application still remains: models can not work well for news in unknown fields, called ñdomainsò, due to bias caused by different words and phrases among domains. To improve the accuracy of cross-domain fake news detection, it is crucial to mitigate the domain bias since unknown news articles to be classified can be in unknown domains. As a preliminary experiment, we trained a classifier using news articles whose noun phrases were masked because they are con-sidered as a major source of the bias. However, contrary to expectations, masking did not improve accuracy. From the preliminary experiment, we obtained the hy-pothesis that pairs of fake and real news on the same topic can mitigate the domain bias. Using comparative experiments, we show that accuracy is higher when trained on paired news articles than when trained on unpaired ones.This result strongly suggests that a fake news dataset consisting of paired news could be effective for cross-domain detection.

Keywords: BERT, cross-domain, fake news detection.

1 Introduction

Thanks to the widespread use of the Internet, we can easily gather information. However, some false information is being spread intentionally on the Internet. One kind of them is ñfake newsò. There are various definitions for fake news, but all of them can be summarized as: news disseminating misinformation for some purpose. For example, Zhou et al. defined fake news as ñintentionally false news published by a news outletò [1].

Fake news has been a serious issue around the world, and has caused serious consequences not only within the Internet but also in the real world. For example, during the 2016 U.S. presidential election, so many fake news was spread that it

[*] Kyushu University, Fukuoka, Japan
[†] Kyushu University, Fukuoka, Japan
[‡] Kyushu University, Fukuoka, Japan

is even said those fake news changed the result of the election [2]. Therefore the need for fake news detection has been recognized. Determining the veracity of such news generally requires prior knowledge of the news and cost a lot to verify the information. Thus the need for automatic detection of fake news is increasing.

There are two major approaches to the fake news detection task: knowledge-based and feature-based. In the former case, a technique called fact-checking is often used. In the latter one, detection is based on capturing unique characteristics of fake news. Supervised learning is often used in this approach.

Feature-based detection has better accuracy as a result of large-scale pre-trained models such as BERT (Bidirectional Encoder Representations from Transformers) [3]. Although it has reached high accuracy in experiments, there is still a significant challenge remaining for its practical application, that is, fake news detection heavily depends on the genre (domain) of news in the training data and can not work well for news in unknown domains [4]. In other words, standard models for fake news detection are overfited to given domains. This is mainly due to *domain bias* caused by differences in vocabulary among domains. For instance, a detection model that judges news with "Donald Trump" words' frequent appearance as fake news is not effective for news in the sports domain. Therefore, *cross-domain fake news detection*—an approach that detect even unknown domains—is important.

Cross-domain detection can not be achieved while focusing on domain-dependent features. It is important for cross-domain detection to capture domain-independent features and reduce the impact of domain bias. In this paper we focus on stylistic characteristics used to deceive readers in fake news, such as extreme writing style and hearsay tone. These characteristics are considered to be common regardless of the domains. Furthermore, Benjamin et al. noted that there are differences in stylistic characteristics between fake news and real news [5]. The ultimate goal of our research is to achieve cross-domain fake news detection. Thus we conducted a preliminary experiment at the beginning. Since nouns are often considered as the main source of domain bias, we trained a classifier using news article data with nouns masked. However, the experiment results showed that this treatment does not improve the accuracy. The result suggests that the masking method is probably not an effective bias mitigation method for the dataset used in this study.

To reach the ultimate goal, it is necessary for us to reveal the reason. We noticed that the dataset in this study has the property that it always contains pairs of fake and real news on the exact same topic. Then we obtained the hypothesis that this property may have effect on domain bias. We focued on this property of dataset and examined how it may affect domain bias and conducted comparative experiments. As a result, models trained on a dataset consisting of fake-real paired news have better accuracy. In addition to our previous paper [6], we conducted an additional comparative experiment to detect the effect of paired and unpaired data on model. The result shows the misleading by the proper nouns is diminished when trained on paired data. This result strongly supports our hypothesis, and opens up a new way for cross-domain fake news detection.

2 Related Work

In this section, we briefly show three types of previous studies about cross-domain fake news detection. These studies commonly use stylistic features of fake news: the first type is a study using manually extracted features, the second one uses deep learning for cross-domain detection, and the third one proposed a method to mitigate bias between datasets. We introduce the third one because the method can be used to mitigate the domain bias.

2.1 Stylistic Characteristics of Fake News

Given a text of news, which characteristics do we intuitively find suspicious? The study by Benjamin et al. [5] statistically tested for differences between authentic and fake news by extracting three categories of features: stylistic features, complexity features, and psychological features. The results showed that characteristics differed significantly. Fake news have characteristics such as less jargon, more lexical redundancy, and more self-referential (e.g. "I", "We" are used more often). In addition, we find more differences between real and fake news in titles, fake news' titles have more capitalized words, fewer stop-words, and more named entities, in order to grab readers' attention by packing in as much information as possible. Classification using SVM with these features result in over 70% accuracy, well above the baseline of 50%. From this results, we can conclude that fake news can be detected by capturing features.

2.2 Cross-Domain Fake News Detection Using Deep Learning

Saikh et al. attempted to improve the accuracy of fake news detection using deep learning [7]. They also used a dataset called FakeNewsAMT, which contains six domains, to test the cross-domain detection accuracy. In the experiment, five domains were used for training, and data from the remaining one domain were classified as fake or non-fake using a neural network. The results show a relatively high accuracy of 73-91% for classification. In addition, factors that enabled highly accurate detection of cross-domain fake news—which is considered to be difficult—have not been verified. We consider that the method used to create FakeNewsAMT may be a contributing factor.

FakeNewsAMT is a dataset of news consisting of titles, contents, and labels. The dataset contains 40 news in each of six domains (business, education, politics, entertainment, sports, and technology) verified as factual, and then crowdsourced to create fake news based on each factual news, giving instructions to write the news in a journalistic style and avoid unrealistic content. Due to the instructions, the dataset can resemble actual fake news. We focus on the nature of FakeNewsAMT due to the method of its creation and examine its impact on cross-domain detection in Section 4.2.

2.3 Bias in Fake News Detection

The distribution of words used in each domain is different, which prevents generalization to unknown domains. That is, the domain bias in the training dataset makes cross-domain fake news detection difficult.

In addition, fake news is a type of news and is therefore influenced by trends and interests. Therefore, the data collected varies greatly depending on when the dataset was created. Murayama et al. named this "diachronic bias" [8], and noted bias among the fake news datasets.

Assuming that this bias is mainly caused by named entities such as person names, Murayama et al. attempted to improve the classification accuracy for datasets created at different times by masking these named entities. The results showed an improved accuracy, suggesting that masking named entities can mitigate the bias between datasets and let the model more generalizable to unknown datasets.

Since there are biases among the six domains included in the FakeNewsAMT dataset used in this study due to vocabulary and other factors, we test whether the method of Murayama et al. can be used to reduce the domain bias.

3 Dataset and Preliminary Experiment

In this section, we conduct cross-domain fake news detection experiments on the six domains of FakeNewsAMT by using BERT. We use the same method as Murayama et al. [8] to try to mitigate the bias between domains, and verify whether there is a change in accuracy compared to training with normal data.

3.1 Data and Preprocessing

We use FakeNewsAMT in this paper, which is news data consisting of titles and body texts. Here is an exmple of a news item in the dataset:

```
Robots Taking Over the World

Robots are slowly taking over the
workforce of the world. Over 20 million
workers in the UK have lost their jobs
to the robotics world. The consultancy
Firm PwC has found...
```

The first sentence is the title of the news, and the next block is the body texts. A total of 80 news datas—both fake and real—were collected for each domain. The basic statistics are shown in Table 1.

Table 1: FakeNewsAMT statistics: average number of words and sentences per news

label	No. of news	Avg. words	Avg. sentences
Fake	240	132	5
Legit	240	139	5

As preprocessing, the publication date and time of the news and the URL were removed. In addition, there are several news articles without titles. In order to put these data into BERT, "No title" was added to the title of the data.

3.2 Bias Mitigation Experiments with Masking

In this section, we use BERT as a pre-trained classification model and conduct the cross-domain detection using FakeNewsAMT, according to Saikh et al. The authors also consider that the distribution of noun phrases vary greatly across domains, and the phrases can be one of the factors contributing to domain bias. We tried to mitigate the bias by masking noun phrases and comparing the cross-domain detection accuracy when using each normal data and masked data.

In this experiment, POS (Part-of-speech) tagging was used before masking noun phrases. It is a requirement for estimating the part-of-speech of words in a sentence. Tokens estimated as proper nouns were replaced with [NNP] labels and those estimated as nouns with [NN] labels. An example of the masking results for the actual data is shown below.

```
Original data:
Trump's next legislative target:tax reform
Masked data:
[NNP]'s next legislative [NN]:[NN] [NN]
```

We used flair[1], a Python framework, to do the POS tagging.

In this study, we employ BERT as a fake news detection model. BERT can be used for variety of tasks such as classification problems and sentence generation, and due to its generality and high performance, BERT has gained popularity in the field of natural language processing. Also, fine-tuning BERT—which has been trained on a large dataset—can perform well on a small dataset.

BERT is given two sentences or one sentence as input, where the input format is "[CLS] 1st-sentence [SEP] 2nd-sentence [SEP]", where [CLS] is the special token indicating the beginning of a sentence, and [SEP] is the special token indicating the end of a sentence. The embedding of [CLS] tokens is sometimes used in classification problems. The input to BERT is the sum of the embedded representation of the word, the representation indicating whether it is the first or second sentence, and the representation with positional information.

We use a pre-trained model published on HuggingFace[2]. The title and body of a news are given as two input sentences. An overview diagram of the model is shown in Fig. 1.

The embedding of the [CLS] tokens in the final layer of BERT ($T_{[CLS]}$) is given as input to the fully connected layer (*FFN*). In the output layer of the FFN, the Softmax function is used for binary classification as fake or non-fake. The number of neurons in each FFN layer is 768, 10, and 2. AdamW is used as the optimizer and the learning rate is set to 1e-05. To prevent overfitting, we drop out 20% of the output of the input layer. The special tokens for masking, [NN] and [NNP], are added as tokens for BERT and we train the FFN layer and fine-tuning BERT.

Four of the six domains in the FakeNewsAMT are used for training, validation is performed in another domain, and the remaining one is tested with the least lossy parameter in the validation domain. There would be five models for one testing domain, and the domain accuracy is the average of the five accuracies. Fig. 2 shows

[1] https://huggingface.co/flair
[2] https://huggingface.co/models

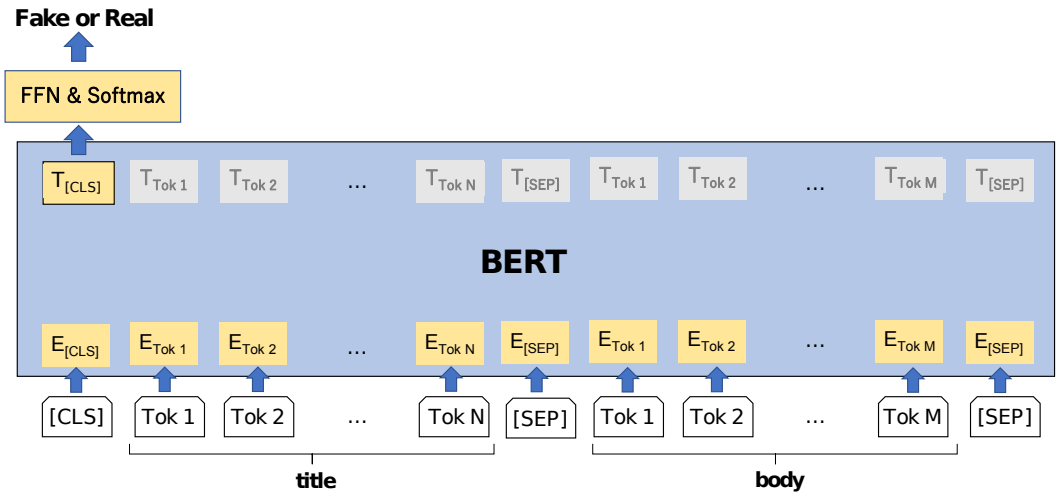

Figure 1: Figure of classification model in this study

an example of how to split the data for training when the business is the testing domain.

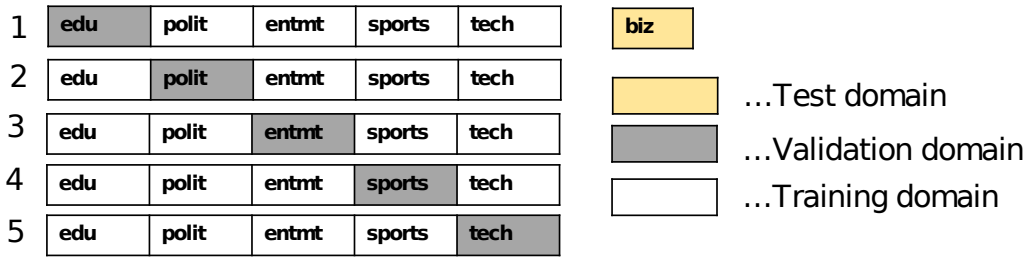

Figure 2: How to split the training data. In line 1, edu is used for validation and others are used for training.

Python 3.8.10 and AllenNLP 2.8.0[3]—a framework for natural language processing—are used in this experiment, where OS is Ubuntu 20.04 LTS, the CPU is AMD Ryzen 7 3700x (3.6GHz), and the GPU is NVIDIA GeForce RTX 3080 (10GB).

These tables show the accuracy when trained with normal data (Table 2) and with noun-masked data (Table 3). The rows represent the test domain and the columns represent the validation domain, and the value of each cell is the accuracy when testing with the classification model trained with the four training domains. The rightmost column is the average accuracy for each test domain.

Firstly, looking at the results for each domain, the average accuracy increases in the politics and entertainment domains due to masking. On the other hand, the average accuracy decreases for the business, education, and technology domains, and remains the same for the sports domain. In terms of overall results, the average accuracy for the six domains is 0.815 for the normal data and 0.801 for the masked data. Masking result in a 1.4% decrease in accuracy.

In conclusion, for FakeNewsAMT, masking noun phrases is not likely to have

[3] https://allenai.org/allennlp

Table 2: Accuracy when training with normal data

Test \Validation	biz	edu	polit	entmt	sports	tech	**Average**
biz		0.89	0.93	0.85	0.86	0.90	**0.886**
edu	0.84		0.84	0.80	0.85	0.80	**0.826**
polit	0.82	0.82		0.79	0.80	0.84	**0.814**
entmt	0.76	0.78	0.70		0.68	0.79	**0.742**
sports	0.86	0.85	0.74	0.79		0.86	**0.820**
tech	0.88	0.70	0.84	0.78	0.82		**0.804**

Table 3: Accuracy when training with masked data

Test \Validation	biz	edu	polit	entmt	sports	tech	**Average**
biz		0.89	0.89	0.85	0.84	0.84	**0.862**
edu	0.82		0.79	0.79	0.86	0.72	**0.796**
polit	0.88	0.79		0.84	0.84	0.80	**0.830**
entmt	0.76	0.74	0.78		0.70	0.75	**0.746**
sports	0.85	0.81	0.79	0.81		0.84	**0.820**
tech	0.78	0.72	0.76	0.76	0.75		**0.754**

effect on domain bias. However, it is noteworthy that although cross-domain fake news detection is considered difficult, some domains have very high accuracy when trained on regular data, as in the experiments of Saikh et al.

This result suggests that the classification model may not been affected by bias for some reason, i.e., the noun phrases—a major source of bias—may not have affected the detection. The fact that no significant differences in accuracy are observed despite the masking of noun phrases also suggests this possibility.

We examine this possibility in detail in the next section.

4 Property to Mitigate Domain Bias

In this section, we quantitatively examine properties of the dataset, and test the impact of these properties on the classification model.

4.1 Lexical Overlap between Paired Data

In this section, we quantitatively analyze the properties of FakeNewsAMT to determine if there are any factors that may contribute to the results in the Section 3.2.

FakeNewsAMT consists of both correct news and crowd-sourced fake news based on the correct news. In other words, this dataset always contains pairs of fake news and real news on the same topic. At this point, it can be assumed that news in the same pair have similar noun phrases, and in fact, FakeNewsAMT shows overlapping noun phrases between the paired news data. An example is shown in Fig. 3. It can be seen that the noun phrases are somewhat similar between the fake news and real news.

We test whether overlap of noun phrases between paired data is found across the entire dataset.

"Alex Jones , **purveyor** of the independent investigative news website Infowars and **host** of The Alex Jones Show , has been vindicated in his **claims** regarding the so-called "Pizzagate" **controversy** . Jones and **others** uncovered **evidence** last **year** that top Democratic **Party officials** were involved in a bizarre , satanic **child sex cult** and **pornography ring** using the Washington D .C . **pizza parlor** Comet Ping Pong Pizza as a **front** . The **allegations** rocked the Democratic **Party** and may have caused serious **damage** to the **Hillary Clinton** presidential **campaign .** Top **U.S.** federal **investigators** have now confirmed that they have verified many of these **claims** after executing raids on the **offices** of several of the key **players** . **Charges** are expected to be filed in the coming **days** .

(a) Fake news

Alex Jones a prominent **conspiracy theorist** and the **host** of a popular right-wing **radio show** has apologized for helping to spread and promote the **hoax** known as Pizzagate . The **admission** on **Friday** by Mr . Jones the **host** of "The Alex Jones Show" and the **operator** of the website Infowars was striking . The Pizzagate **theory** which posited with no **evidence** that top Democratic **officials** were involved with a satanic **child pornography ring** centered around Comet Ping Pong a **pizza restaurant** in Washington D .C **.** grew in online **forums** before making its **way** to more visible **venues** including Mr . Jones's **show.**

(b) Real news

Figure 3: Paired news data
(Bold: Noun phrase, Blue: Overlapping noun phrase)

Before calculating the overlap rate of noun phrases between the paired data, we preprocessed the data using the method of Juan et al. [9]. Firstly, all letters are converted to lowercase. Secondly, frequently used words such as “I”, “a”, and “of”, called stop words, are removed. Finally, lemmatization is performed to convert words into headwords. For example, “dogs” is converted to “dog”, and “met” to “meet”.

The overlapping noun phrases shown in blue in Fig. 3 are often unique to the news, such as proper nouns. We suspect that noun phrases unique to that news may have more overlap in the paired data, so in addition to calculating the overlap rate for all noun phrases, we also calculate the overlap rate for characteristic noun phrases. We use the TF-IDF method to check whether the noun phrases are news-specific or not.

TF is the frequency of a word in a document.

$$tf(t,d) = \frac{n_{t,d}}{\sum_{s \in d} n_{s,d}}, \tag{1}$$

where $n_{t,d}$ is the frequency of a word t in document d and $\sum_{s \in d} n_{s,d}$ is the sum of the frequencies of all words in document d. In this experiment, a document d refers to a news article. The IDF value for the word t is defined as

$$idf(t) = \log \frac{N}{df(t)} + 1, \tag{2}$$

where N is the number of all documents and $df(t)$ is the number of documents in which the word t appears. In this experiment, all documents refers to 480 news data including fake and real news.

Finally, the product of TF and IDF is TF-IDF. Document-specific words are assigned a higher TF-IDF value.

Lexical overlap rates are calculated between all paired data of fake and real news. The results are shown in Table 4. In all domains, the overlap rate for noun phrases only is higher than that for the entire vocabulary. We also calculate the overlap rate of noun phrases with the top 20 TF-IDF values, which is higher in most domains. The top TF-IDF words include many words that are likely to be major sources of

domain bias, such as named entities and nouns specific to that domain. The results show that many of these words overlaps between pairs of data.

Table 4: Average percentage of lexical overlap between paired data

	Whole vocabulary	Noun only	TF-IDF Top 20 Noun
biz	0.300	0.367	0.397
edu	0.247	0.308	0.391
polit	0.343	0.409	0.402
entmt	0.238	0.319	0.405
sports	0.278	0.340	0.343
tech	0.210	0.279	0.408

There is a lot of overlap of noun phrases between pairs of FakeNewsAMT data. The similarity of noun phrases between the fake and non-fake data suggests that the model may have learned to make judgments without noun phrases. It is very important for cross-domain fake news detection that the model is not influenced by noun phrases, which can be a source of domain bias.

4.2 The Effect of Training Data Properties on Accuracy

In this section, we examine whether training on paired data with overlapping noun phrases affects domain bias and improves the accuracy of cross-domain fake news detection. We create training data consisting of paired data only and, conversely, training data consisting of unpaired data. We evaluate the accuracy on the test domain of models trained on these datasets.

The paired dataset consists of 80 randomly selected fake and non-fake pairs of data from each of the four training domains. Conversely, the unpaired dataset consists of 160 randomly selected (paired data not included) data from each domain (Fig. 4). The amount of data for both datasets is 160, and 10 training datasets were created for each dataset.

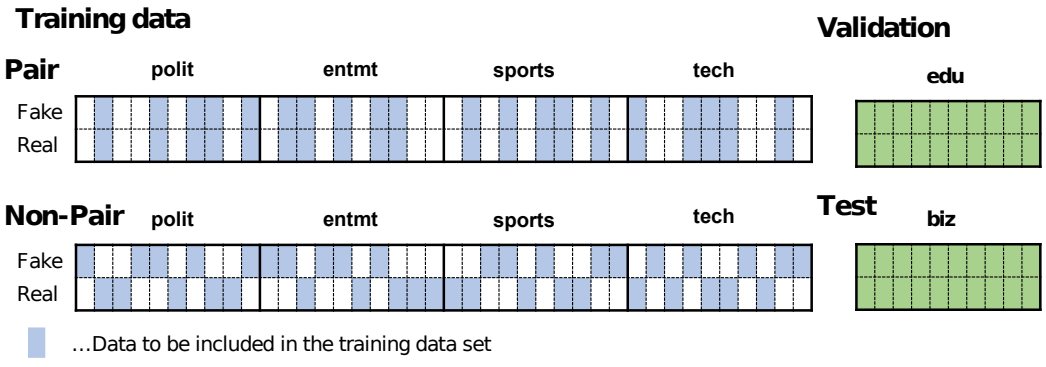

Figure 4: An example of how to create a data set

The same classification model from Section 3.2 was used. Table 5 shows the results when trained on the dataset constructed with or without paired data, and showing 5% to 7% higher accuracy with paired data. This improvement's cause

is explored through a comparative experiment visualizing BERT's saliency scores, depicted in Fig. 5. The Deeper colored areas indicate higher saliency score. When trained on unpaired data, proper nouns like "Nintendo" have the highest saliency score, which means the word has the greatest affect and may mislead the model. When it comes to paired data, the saliency score is more even, allowing for more comprehensive feature learning.

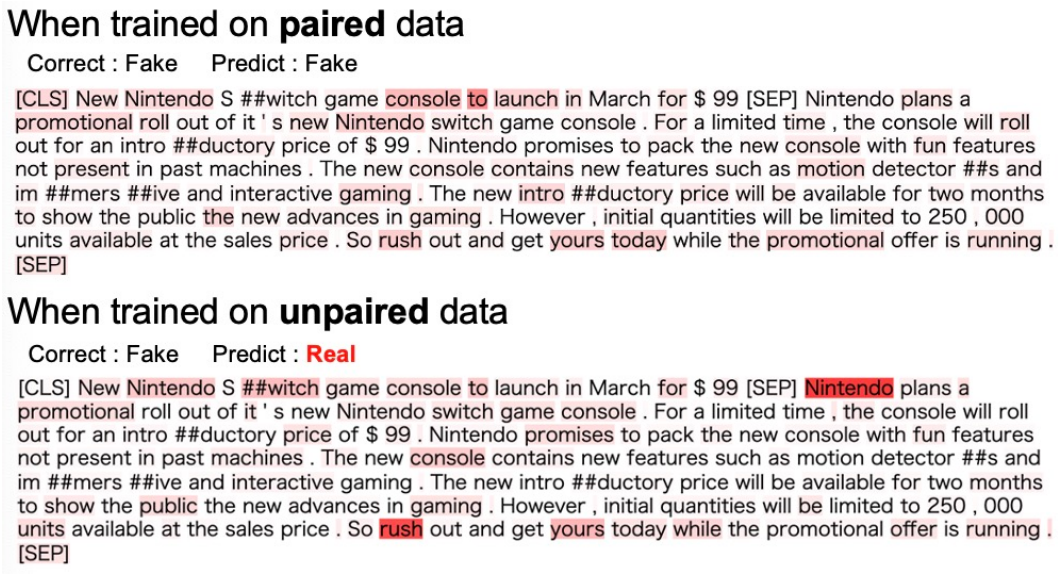

Figure 5: A result of the comparative experiment

However, despite the difference in average accuracy in the sports domain, there is only a small difference in accuracy except in some cases where accuracy rates are extremely low. Therefore, to check whether there is a significant difference in the accuracy between the two groups, we use Mann-Whitney's U test, which is a nonparametric method that tests whether there is a difference in the population representative values between two groups with no correspondence (Table 6).

Table 5: Average accuracy for each domain when training on a dataset built with corresponding fake and real news paired data or without paired data

Training Data	Test	1	2	3	4	5	6	7	8	9	10	Average
Paired Data	**biz**	0.78	0.85	0.81	0.79	0.82	0.81	0.80	0.86	0.81	0.84	**0.817**
	edu	0.70	0.69	0.71	0.71	0.69	0.68	0.68	0.62	0.68	0.69	**0.685**
	polit	0.74	0.68	0.80	0.78	0.66	0.71	0.75	0.79	0.70	0.78	**0.739**
	entmt	0.64	0.69	0.65	0.66	0.74	0.65	0.70	0.68	0.66	0.66	**0.673**
	sports	0.78	0.89	0.82	0.76	0.89	0.80	0.84	0.84	0.79	0.81	**0.822**
	tech	0.79	0.75	0.75	0.78	0.75	0.79	0.79	0.81	0.75	0.70	**0.766**
Non-Paired Data	**biz**	0.81	0.82	0.79	0.61	0.80	0.72	0.68	0.85	0.71	0.66	**0.745**
	edu	0.56	0.72	0.65	0.66	0.69	0.68	0.60	0.57	0.59	0.64	**0.636**
	polit	0.72	0.66	0.56	0.69	0.75	0.76	0.68	0.75	0.59	0.68	**0.684**
	entmt	0.53	0.69	0.62	0.64	0.62	0.57	0.60	0.65	0.69	0.62	**0.623**
	sports	0.80	0.84	0.78	0.80	0.61	0.86	0.85	0.78	0.61	0.80	**0.773**
	tech	0.70	0.78	0.76	0.71	0.70	0.70	0.71	0.57	0.72	0.68	**0.703**

The result of the test shows that the differences at the 5% significance level in the business, education, and entertainment domains, and at the 1% significance level in the technology domain. No significant differences are found in the politics and sports domains in this experimental data.

Table 6: The result of Mann-Whitney U test

Test domain	p-value
biz	**0.045**
edu	**0.028**
polit	**0.076**
entmt	**0.014**
sports	**0.307**
tech	**0.008**

There are significant differences in four of the six domains, suggesting that it is possible to improve the accuracy of unknown domains by composing data sets with paired data. Although we were not able to quantitative analyze the causes, the sports domain contains several fake news that seem to reverse the wins and losses of games. This suggests that it is quite difficult to determine whether these news are fake or not, and some of the unique properties of fake news in the sports domain may have influenced the results.

5 Conclusion

In this study, we validated the property of the dataset for cross-domain detection.

First, we tried reducing bias by training on data with masked noun phrases. Contrary to expectations, this didn't improve accuracy, but training with normal data achieved high accuracy. We then examined FakeNewsAMT's structure, where fake and real news pairs cover identical topics with overlapping noun phrases. Our hypothesis was that the model, due to this property, might learn to disregard noun phrases in determining news authenticity, thus reducing domain bias. To verify this, we compared the accuracy of models trained on datasets with only paired or unpaired data. The results show that accuracy is higher when trained on the paired dataset, with four of the six domains showing higher accuracy at the 5% significant level rather than differences due to randomness. With the result of the additional experiment, we conclude that the reason for the accuracy improvement is that training on paired data can diminish the misleading by proper nouns.

We showed that in order to improve the accuracy of cross-domain detection, it may be important to collect pairs of real and fake news with similar noun phrases on the same topic.

This study presents two future challenges. The first is devising a method for actual dataset creation. FakeNewsAMT employs crowdsourcing for generating fake news, but when compiling a dataset from real news and fake news, a strategy for gathering paired news data is needed. The second challenge is to quantitatively analyze why the sports domain shows fewer differences. Understanding the unique nature of sports-related fake news could offer valuable insights for research in fake news detection.

6 Acknowledgement

This work was supported by JSPS KAKENHI Grant Number 23H03459.

References

[1] Xinyi Zhou and Reza Zafarani. A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities. *ACM Comput. Surv.*, 53(5), 2020.

[2] Washington Post. A new study suggests fake news might have won Donald Trump the 2016 election. `https://www.washingtonpost.com/news/the-fix/wp/2018/04/03/a-new-study-suggests-fake-news-might-have-won-donald-trump-the-2016-election/`, 2018.

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, NAACL-HLT, Volume 1 (Long and Short Papers)*, pages 4171–4186, 2019.

[4] Amila Silva, Ling Luo, Shanika Karunasekera, and Christopher Leckie. Embracing Domain Differences in Fake News: Cross-domain Fake News Detection using Multi-modal Data. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021*, pages 557–565. AAAI Press, 2021.

[5] Benjamin Horne and Sibel Adali. This Just In: Fake News Packs A Lot In Title, Uses Simpler, Repetitive Content in Text Body, More Similar To Satire Than Real News. *Proceedings of the International AAAI Conference on Web and Social Media*, 11(1):759–766, 2017.

[6] Shingo Kato, Linshuo Yang, and Daisuke Ikeda. Domain Bias in Fake News Datasets Consisting of Fake and Real News Pairs. *the 12th International Congress on Advanced Applied Informatics (IIAI-AAI), the 14th International Conference on E-Service and Knowledge Management (ESKM 2022)*, pages 101–106, 2022.

[7] Tanik Saikh, Arkadipta De, Asif Ekbal, and Pushpak Bhattacharyya. A deep learning approach for automatic detection of fake news. In *Proceedings of the 16th International Conference on Natural Language Processing*, pages 230–238, 2019.

[8] Murayama Taichi, Wakayama Shoko, and Aramaki Eiji. Diachronic bias in fake news detection datasets (in Japanese). *Proceedings of the Twenty-seventh Annual Meeting of the Association for Natural Language Processing*, pages 1011–1016, 2021.

[9] Juan Pablo Posadas-Durán, Helena Gomez-Adorno, Grigori Sidorov, and Jesús Jaime Moreno Escobar. Detection of fake news in a new corpus for the Spanish language. *Journal of Intelligent and Fuzzy Systems*, 36:4868–4876, 2019.