DAFD: Domain Adaptation Framework for Fake News Detection

Yinqiu Huang[1], Min Gao[1(✉)], Jia Wang[1], and Kai Shu[2]

[1] School of Big Data and Software Engineering, Chongqing University, Chongqing 401331, China
yinqiu@cqu.edu.cn
[2] Illinois Institute of Technology, Chicago, IL 60616, USA

Abstract. Nowadays, social media has become the leading platform for news dissemination and consumption. Due to the convenience of social media platforms, fake news spread at an unprecedented speed, which has brought severe adverse effects to society. In recent years, the method based on deep learning has shown superior performance in fake news detection. However, the training of this kind of model needs a large amount of labeled data. When a new domain of fake news appears, it usually contains only a small amount of labeled data. We proposed a novel Domain Adaptation framework for Fake news Detection named DAFD. It adopts a dual strategy based on domain adaptation and adversarial training, aligns the data distribution of the source domain and target domain during the pre-training process, and generates adversarial examples in the embedding space during the fine-tuning process to increase the generalization and robustness of the model, which can effectively detect fake news in a new domain. Extensive experiments on real datasets show that the proposed DAFD achieves the best performance compared with the state-of-the-art methods for a new domain with a small amount of labeled data.

Keywords: Fake News · Domain Adaptation · Adversarial Training.

1 Introduction

With the continuous development of the internet, all kinds of social media gradually enter people's lives. However, while these social media platforms bring convenience, they also become a breeding ground for fake news. The wide dissemination of fake news has brought severe adverse effects to society, such as weakening the public's trust in the government and journalism, and causing severe economic losses. Thus, timely detection of fake news has great significance.

However, the identification of fake news usually has many challenges because the data is large-scale and multi-modal. To solve this problem, researchers have proposed fake news detection models based on deep learning [8, 13–15]. A large number of manual labeled data is the premise of training a deep learning network. However, when a new domain appears, only domain experts can make

accurate manual annotations. Expensive and time-consuming manual labeling leads to a lack of labeled data in a new domain. Directly training the deep networks on a small amount of labeled data usually causes the models to overfit. In response to this problem, some researchers use pre-trained models to help detection. However, these pre-trained models do not have good detection performance because of the discrepancy between source data and target data.

To tackle these challenges, we propose a Domain-Adaptive Fake news Detection framework (DAFD) based on a dual strategy to improve detection performance. First, we perform domain adaptation operations in the pre-training process. The domain adaptation loss is introduced to align the data distributions of the two domains to learn a news representation with semantic information and domain alignment. We use a domain adaptation loss based on the maximum mean difference (MMD) [1] to learn a representation that optimizes fake news detection and domain invariance. Second, we use adversarial training to generate adversarial examples in the embedding space for additional training during the fine-tuning process to enhance the model robustness and generalization capabilities. Our main contributions are summarized as follows:

• The proposed DAFD can automatically align the data distribution of the source domain and the target domain during the pre-training process to ensure the model's detection performance in the target domain.

• We use adversarial training to make the model more robust and generalized in fine-tuning and further improve the performance of detection.

• We have conducted extensive experiments to show the performance of DAFD on the detection of fake news with a small amount of labeled data and analyzed why adversarial training and domain adaptation work.

2 Related Work

2.1 Fake news detection

Recently, researchers have proposed many fake news detection technologies, which can be roughly divided into methods based on statistical learning and methods based on deep learning. Statistical learning methods usually extract features from news, and then train the model through these features. Tacchini et al. [17] tried to determine the authenticity of a news article based on the users who interacted with it or liked it. The deep learning model often has better performance because of its strong ability to automatic learning information representation. Ma et al. [8] proposed a new method to capture news features based on rumor life cycle time series. Shu et al. [15] developed an interpretable fake news detection framework using the mutual attention of news and corresponding comments.

However, these methods all require many labeled data and cannot detect with only a little labeled data. Therefore, We pay more attention to how to detect fake news in a new domain.

2.2 Transfer learning

Our research is related to transfer learning, which can be roughly divided into four categories. Instance based Transfer Learning generates rules based on some weights and reuses data samples. Feature based Transfer Learning reduces the distance between different domains by feature transformation. For example, Zhang et al. [21] proposed to train different transformation matrices between different domains to achieve better results. Model based Transfer Learning finds the shared parameter information between different domains for information transfer. Relation Based Transfer Learning focuses on the relationship between samples in different domains, and researchers usually use Markov Logic Net to explore the similarity of relationships between different domains [3].

With the idea of feature based transfer learning, we propose a pretraining-finetuning framework to continuously align the data distribution during the pre-training process to improve the performance of fake news detection in a new domain.

3 Methodology

We will simply introduce the domain adaptation framework of fake news detection in this section and then introduce each component in detail.

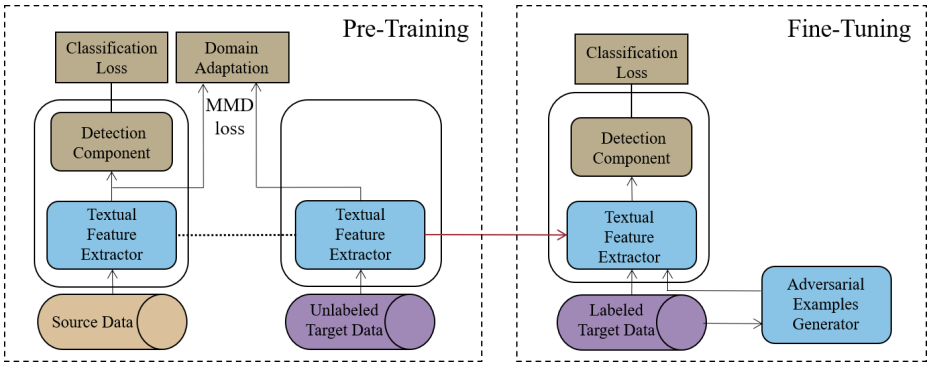

Fig. 1. The architecture of the DAFD, including pre-training and fine-tuning.

3.1 Overview

The question is set up as follows. We regard fake news detection as a binary text classification problem. That is, each news article can be real ($y = 0$) or fake ($y = 1$), and the target domain data used for fine-tuning contains only a small amount of labeled data. Our goal is to predict the labels of other news data in the target domain.

Figure 1 shows the structure of the DAFD. It can be divided into two parts: pre-training and fine-tuning. The former comprises three parts: textual feature extractor, domain adaptation, and detection part. The latter consists of adversarial examples generator, textual feature extractor, and detection part.

The pre-training model's input is source data and unlabeled target data, and then the textual feature extractor is used to model the news text from language features to hidden feature space. Meanwhile, we use domain adaptation to learn a classification representation that aligns the data distribution and can extract semantics. Finally, the fake news detection component performs fake news detection. The fine-tuning module takes labeled target data as input, extracts features through a pre-trained textual feature extractor, and then detects fake news through a fake news detection component that initializes parameters randomly, and then uses discriminative fine-tuning and gradual unfreezing [5] to finetune the model parameters. To enhance the robustness and generalization capabilities, we also generate adversarial examples for adversarial training. In this paper, the fake news detection component is an MLP network.

3.2 Textual feature extractor

Some researchers have found that the hierarchical attention neural network [19] has great advantages for learning document representations that emphasize important words or sentences. Inspired by this, we propose to learn document representation through a hierarchical structure. Concretely, we first learn sentence vectors through a word encoder and then learn documents vectors through a sentence encoder.

Word encoder In order to consider the context information of words, we use bidirectional GRU to capture the features of word sequences.

Given a sentence $s_i = \{w_1^i, \ldots, w_{M_i}^i\}$ contains M_i words, we use both forward GRU $\overrightarrow{f}$ and backward GRU $\overleftarrow{f}$ to model sentences from two directions:

$$\begin{aligned} \overrightarrow{\mathbf{h}_t^i} &= \overrightarrow{GRU}\left(\mathbf{w}_t^i\right), t \in \{1, \ldots, M_i\}, \\ \overleftarrow{\mathbf{h}_t^i} &= \overleftarrow{GRU}\left(\mathbf{w}_t^i\right), t \in \{M_i, \ldots, 1\}. \end{aligned} \tag{1}$$

We concatenate $\overrightarrow{h_t^i}$ and $\overleftarrow{h_t^i}$ to get the annotation $h_t^i = \left[\overrightarrow{h_t^i}, \overleftarrow{h_t^i}\right]$ of w_t^i, which contains the contextual information centered around w_t^i. Not every word has the same influence on a sentence. Hence, we propose an attention mechanism to measure word importance, and the sentence vector v^i is

$$\mathbf{v}^i = \sum_{t=1}^{M_i} \alpha_t^i \mathbf{h}_t^i, \quad \mathbf{u}_t^i = \tanh\left(\mathbf{W}_w \mathbf{h}_t^i + \mathbf{b}_w\right), \quad \alpha_t^i = \frac{\exp\left(\mathbf{u}_t^i \mathbf{u}_w^\top\right)}{\sum_{k=1}^{M_i} \exp\left(\mathbf{u}_k^i \mathbf{u}_w^\top\right)}, \tag{2}$$

where α_t^i represents the importance of t^{th} word for the sentence s_i, u_t^i is a hidden representation of h_t^i, and u_w is a parameter matrix representing a word-level context vector, which will be initialized randomly and updated with other parameters.

Sentence encoder Similar to word encoder, we learn the document representation by capturing the context information in the sentence-level. Given the sentence vectors v^i, we use a bidirectional GRU to encode the sentences:

$$\begin{array}{l} \overrightarrow{\mathbf{h}^i} = \overrightarrow{GRU}\left(\mathbf{v}^i\right), i \in \{1, \ldots, N\}, \\ \overleftarrow{\mathbf{h}^i} = \overleftarrow{GRU}\left(\mathbf{v}^i\right), i \in \{N, \ldots, 1\}. \end{array} \tag{3}$$

We concatenate $\overrightarrow{h^i}$ and $\overleftarrow{h^i}$ to get the annotation $h^i = \left[\overrightarrow{h^i}, \overleftarrow{h^i}\right]$, which captures the context from neighbor sentences around sentence s_i. Similarly, an attention mechanism is used to measure sentence importance, and we calculate the document vector v by

$$\mathbf{v} = \sum_i \alpha^i \mathbf{h}^i, \quad \mathbf{u}_i = \tanh\left(\mathbf{W}_s \mathbf{h}^i + \mathbf{b}_s\right), \quad \alpha_i = \frac{\exp\left(\mathbf{u}_s \mathbf{u}_i^\top\right)}{\sum_i \exp\left(\mathbf{u}_s \mathbf{u}_i^\top\right)}, \tag{4}$$

where α^i represents the importance of i^{th} sentence for the document, u_s is the weight parameter that represents the sentence-level context vector. It also will be initialized randomly and updated with other parameters.

3.3 Domain adaptation component

In the new domain scenario, the labeled data of the target domain is usually very scarce, so we cannot directly train the model with the target data. Besides, news in the new domain (target domain) usually has a different data distribution. If it is simply pre-trained, it will lead to the classifier overfits the source data distribution. Suppose we can get a representation that aligns the source data distribution with the target data distribution during the pre-training process, we can be better compatible with the target data.

We use MMD distance to measure the distance between two domains. We calculate the distance by a specific representation $\varphi(\cdot)$. We use this representation $\varphi(\cdot)$ to operate on source data, $x_s \in X_S$, and target data, $x_t \in X_T$. In our work, we use the Gaussian kernel function, then an empirical approximation to this distance is computed as:

$$MMD\left(X_S, X_T\right) = \left\| \frac{1}{|X_S|} \sum_{x_s \in X_S} \phi\left(x_s\right) - \frac{1}{|X_T|} \sum_{x_t \in X_T} \phi\left(x_t\right) \right\|, \tag{5}$$

As shown in Figure 1, we not only need to minimize the distance between the source domain and the target domain (or align the data distribution of the two domains) but also need an effective classification representation that is conducive to detection. Such a classification representation will enable us to transfer fake news detection models across domains easily. We achieve this goal by minimizing the following loss during the pre-training process:

$$L = L_C\left(X_S, Y_S\right) + \lambda MMD^2\left(X_S, X_T\right), \tag{6}$$

where $L_C(X_S, Y_S)$ represents the classification loss of the labeled data X_S and the real label Y_S in the source domain, $MMD(X_S, X_T)$ represents the distance between the two domains. The hyperparameter λ represents the degree to which we want to align the data distribution.

Existing studies have shown that the meanings of features extracted from different layers of neural networks are different [20]. Specifically, some neural networks in the first few layers capture relatively general features, while some neural networks in the latter layers capture more specific features. Tzeng et al. [18] proved that domain adaptation in the previous layer of the classifier could achieve the best results, so we perform domain adaptation operations after feature extraction.

3.4 Adversarial examples generator

Since there is usually less labeled data in the target domain, direct fine-tuning will result in poor model generalization. Adversarial training is commonly used to build robust deep models. Existing research shows that adversarial training on the language model can increase generalization ability and robustness [2, 11].

Pre-trained models such as Bert [4] and Albert [7] have been proved to impact downstream tasks positively. Our goal is to further the generalization ability of fake news detection in a new domain by enhancing the robustness of model embeddings. It is challenging to create actual adversarial examples for the language directly because the context determines the semantics of each word [22]. Some scholars have proposed that embedding-based confrontation is more effective than text-based confrontation [22]. Therefore, we create appropriate adversarial examples in the embedded space and then update the parameters on these adversarial examples to implement adversarial training.

Specifically, we use a gradient-based method to add norm-bounded adversarial perturbations to the embedding of the input sentence. We use $W = [w_1, w_2, \cdots, w_n]$ to represent the sequence of one-hot representations of the input words. V represents the embedding matrix, $y = f_\theta(X)$ represents the language model function, $X = VW$ is the word embeddings, and y is the output (class probabilities). θ represents all the parameters of the model. We get new prediction $\hat{y} = f_\theta(X + \delta))$ by adding adversarial perturbations δ to the embedding. To maintain the original semantics, we limit the norm of δ and assume that the detection results of the model will not change after perturbation.

For any δ within a norm ball, we expect to minimize the maximum risk as:

$$\min_{\theta} E_{(Z,y)\sim\mathcal{D}} \left[\max_{\|\delta\| \leq \epsilon} L\left(f_\theta(X+\delta), y\right) \right], \tag{7}$$

where D is the data distribution, L is the loss function. Madry et al. [10] demonstrated that SGD and PGD could reliably solve the saddle-point problem in neural networks.

With the idea of PGD, we perform the following steps in each iteration:

$$x_{t+1} = \Pi_{x+\mathcal{S}}\left(x_t + \alpha g\left(x_t\right) / \left\|g\left(x_t\right)\right\|_2\right), \tag{8}$$

where α is the step size, $S = r\epsilon R^d : \|r\|_2 \leq \varepsilon$ is the constraint space of disturbance, and $g(x)$ is the gradient of embedding. The loss function during the fine-tuning process is as follow:

$$L = L_C\left(X_T, Y_T\right) + L_C\left(X_{T_{adv}}, Y_T\right), \tag{9}$$

where X_{Tadv} is the adversarial examples generated by equation 8.

4 Experiments

Table 1. The statistics of datasets

	PolitiFact	GossipCop	Covid
True news	145	3,586	5,600
Fake news	270	2,230	5,100
Total	415	5,816	10,700

Our complete code can be obtained from the link[1]. In this section, we conducted a lot of experiments on three datasets to validate and analyze the performance of DAFD on fake news detection. Specifically, we aim to answer the following evaluation questions:

• EQ1 Compared with other existing methods, can DAFD improve the fake news detection performance in a new domain?

• EQ2 How effective are domain adaptation and adversarial training in improving DAFD detection performance?

• EQ3 What is the impact of the amount of labeled data on the detection performance of DAFD?

4.1 Experiments setup

In this experiment, We use the FakeNewsNet [16] dataset, which is a comprehensive dataset for fake news detection. It contains the data of two fact verification platforms: PolitiFact and GossipCop, including news content, labels, and social information. In addition, we also use a dataset Covid [12] for pretraining. Our model only uses the news text in the dataset, and some baseline methods use the news comment data additionally. The detailed statistics of these datasets are shown in Table 1. We use four metrics commonly used in classification tasks to measure the effect of fake news detection.

[1] https://github.com/964070525/DAFD-Domain-Adaptation-Framework-for-Fake-News-Detection

4.2 Baselines

To validate the effectiveness of the DAFD, we choose both traditional machine learning algorithms and deep learning models as baseline methods. The representative advanced fake news detection baselines compared in this paper are as follows:

• Traditional machine learning: TF-IDF is a statistical method, which is often used for feature extraction. Based on TF-IDF features, we detect fake news with different traditional machine learning algorithms, including Naive Bayesian (NB), Decision Tree (DT) and Gradient Boosting Decision Tree (GBDT).

• Text-CNN [6]: It uses the convolution neural network to extract news features and capture different granularity feature information through different size filters.

• TCNN-URG [13]: It uses two components to extract different information features, uses a convolutional neural network to learn the representation of content, and uses a variational automatic encoder to learn the feature information of user comments.

• dEFEND [15]: It is an interpretable fake news detection framework using the mutual attention of news and corresponding comments.

4.3 Fake news detection performance (EQ1)

To answer EQ1, we compare our method with representative baselines introduced in Section 4.2 under the lack of labeled data in a new domain. To simulate the lack of labeled data, we artificially decrease the ratio of labeled data to 5% in GossipCop dataset and named it GossipCop_05. The experimental results of other ratios are shown in section 4.6. Because the PolitiFact dataset is small, we keep 75% of the labeled data and named it PolitiFact_75. The amount of data used for pre-training should be greater than the target data amount. For PolitiFact_75 data, GossipCop data is used as the source domain for pre-training, and for GossipCop_05 data, Covid data is used for pre-training. The average performance is shown in Table 2.

Among the traditional machine learning methods, GBDT has the best effect because GBDT continuously fits the model residuals of the previous stage, making the model have a stronger generalization ability. In the deep learning methods, dEFEND and TCNN-URG use the comments of each news. Among them, dEFEND has the best effect. dEFEND uses two separate encoders and a co-attention layer to capture the article's representation to achieve better detection results. The effects of Text-CNN and TCNN-URG are relatively poor, which also show that traditional deep learning models cannot effectively detect with a small amount of labeled data.

We can observe that DAFD has achieved the best results in almost all indicators. All baselines are trained directly on the target data. When we use dEFEND to pretraining on GossipCop and then finetuning on PolitiFact_75, the F1 score of DAFD is still better than dEFEND (7.4%), which prove the DAFD can effectively detect fake news in a new domain.

Table 2. The performance comparison between DAFD and baselines

Datasets	Metric	NB	DT	GBDT	Text-CNN	TCNN-URG	dEFEND	DAFD
PolitiFact_75	Accuracy	0.760	0.712	0.808	0.653	0.712	0.904	**0.966**
	Precision	0.942	0.768	0.928	0.678	0.711	0.902	**0.965**
	Recall	0.756	0.791	0.810	0.863	0.941	0.956	**0.960**
	F1	0.838	0.779	0.864	0.760	0.810	0.928	**0.962**
GossipCop_05	Accuracy	0.627	0.630	0.681	0.630	0.591	0.689	**0.703**
	Precision	0.156	0.502	0.438	0.505	0.428	0.604	**0.614**
	Recall	0.556	0.521	**0.622**	0.604	0.196	0.548	0.610
	F1	0.244	0.511	0.514	0.516	0.269	0.575	**0.602**

4.4 Effects of framework components (EQ2)

To answer EQ2, we evaluate the dual strategy of domain adaptation and adversarial training. Specifically, we investigate the effects of these strategies by defining three variants of DAFD:

• DAFD\D: DAFD without domain adaptation. In the process of pre training, it removes the domain adaptation part, only uses the source domain data, uses cross entropy as the loss function.

• DAFD\A: DAFD without adversarial training. In the process of fine-tuning, it removes the adversarial training part and keeps the rest components.

• DAFD\All: DAFD without the pre-training and fine-tuning parts. It uses the feature extractor to extract the features and then directly performs classification detection, using cross entropy as the loss function.

The best performances are shown in Figure 2, we make the observations:

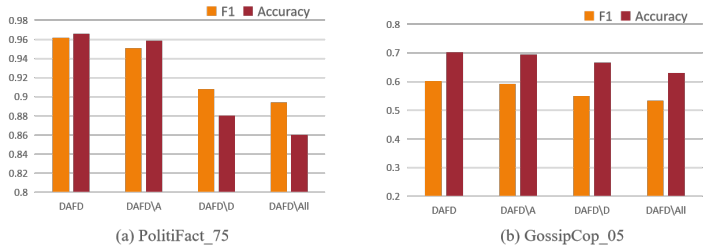

Fig. 2. Impact analysis of framework components for fake news detection.

• Without the adversarial training part, the performance will be reduced, which indicates that the adversarial training helps to improve the generalization ability of the model.

• Without the domain adaptation process,the performance reduces 5.6% and 8.9% in terms of F1 and Accuracy metrics on PolitiFact, 8.8% and 5.4% on GossipCop.

• Without the pre-training and fine-tuning parts, the performance degradation is the biggest, which indicates the importance of DAFD framework in fake news detection in a new domain.

Through the component analysis of DAFD, we conclude that (1) both components of domain adaptation and adversarial training are conducive to improve the performance of DAFD; (2) it is necessary to use a pre-training and fine-tuning framework to detect fake news in a new domain because it can effectively use other knowledge to assist detection.

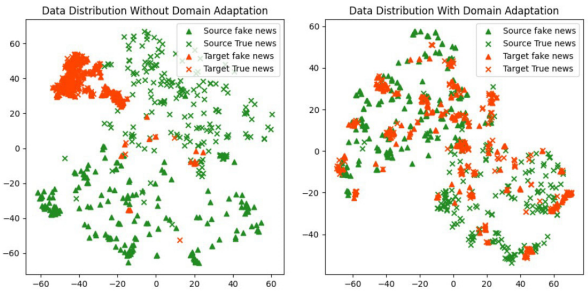

Fig. 3. The distribution of news embedding with domain adaptation and without domain adaptation

4.5 Analysis of domain adaptation effectiveness

To observe the result of domain adaptation more intuitively, we use t-Distribution Stochastic Neighbor Embedding (t-SNE) [9] to decrease the dimension of news' embedding to 2 and draw them in Figure 3. To show the results better, we randomly selected 200 samples for each category. As can be seen in Figure 3, the two data distributions without domain adaptation (left) are quite different, and the model parameters learned in the source domain may not be well applied to the target domain. The two data distributions with domain adaptation (right) are roughly aligned, and the parameters learned in the source domain can be well applied to the target domain.

4.6 Impact of the amount of labeled data on the DAFD (EQ3)

This section answers EQ3, we explore the impact of the amount of labeled data on the model performance. We decrease the dataset GossipCop again and name it GossipCop_10 according to the proportion of training samples of 10%, in the same way, we get GossipCop_15 and GossipCop_75. We also use Covid data for pre-training. The average performance is reported in Figure 4.

As shown in Fig. 4, when the amount of labeled data is 5%, the F1 value increases by the framework is the largest (9.6%). With the increasing amount

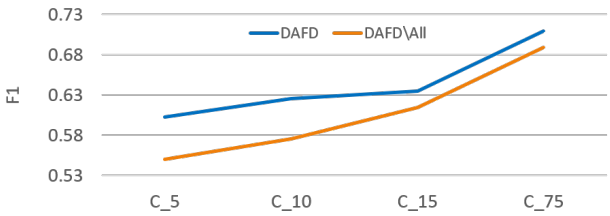

Fig. 4. The influence of labeled data amount on the performance

of labeled data, the F1 value benefits gradually decrease. Until the amount of labeled data reaches 75%, the F1 value benefit is the lowest (3.1%). This is because when there are enough labeled data in the new domain, the deep learning model can learn enough information from the data in the new domain, and the benefit of the pretraining model will decrease.

5 Conclusions and Future Work

In this paper, we investigate an important problem of fake news detection. Due to the suddenness of news in a new domain, it is challenging to obtain a large amount of labeled data, and it is difficult for models based on deep learning to cope with this situation. Therefore, we propose a new framework that can use knowledge in other domains to assist detection, align data distribution through domain adaptation technology, and enhance model robustness and generalization capabilities through adversarial training. Experiments on real-world datasets demonstrate the effectiveness of the proposed framework. In the future, we will capture the specificity of each domain in the process of domain adaptation to further improve the detection performance in each domain. Besides, we will introduce social information to assist detection.

Acknowledgments Yinqiu Huang, Min Gao, and Jia Wang are supported by the Natural Science Foundation of Chongqing (cstc2020jcyj-msxmX0690). Kai Shu is supported by the John S. and James L. Knight Foundation through a grant to the Institute for Data, Democracy & Politics at The George Washington University.

References

1. Borgwardt, K.M., Gretton, A., Rasch, M.J., Kriegel, H.P., Schlkopf, B., Smola, J., A.: Integrating structured biological data by kernel maximum mean discrepancy. Bioinformatics (2006)
2. Cheng, Y., Jiang, L., Macherey, W.: Robust neural machine translation with doubly adversarial inputs. arXiv preprint arXiv:1906.02443 (2019)
3. Davis, J., Domingos, P.: Deep transfer via second-order markov logic. ACM (2009)

4. Devlin, J., Chang, M.W., Lee, K., Toutanova, K.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
5. Howard, J., Ruder, S.: Universal language model fine-tuning for text classification. arXiv preprint arXiv:1801.06146 (2018)
6. Kim, Y.: Convolutional neural networks for sentence classification. Eprint Arxiv (2014)
7. Lan, Z., Chen, M., Goodman, S., Gimpel, K., Sharma, P., Soricut, R.: Albert: A lite bert for self-supervised learning of language representations. arXiv preprint arXiv:1909.11942 (2019)
8. Ma, J., Gao, W., Wei, Z., Lu, Y., Wong, K.F.: Detect rumors using time series of social context information on microblogging websites. In: Proceedings of the 24th ACM international on conference on information and knowledge management. pp. 1751–1754 (2015)
9. Van der Maaten, L., Hinton, G.: Visualizing data using t-sne. Journal of machine learning research **9**(11) (2008)
10. Madry, A., Makelov, A., Schmidt, L., Tsipras, D., Vladu, A.: Towards deep learning models resistant to adversarial attacks. arXiv preprint arXiv:1706.06083 (2017)
11. Miyato, T., Dai, A.M., Goodfellow, I.: Adversarial training methods for semi-supervised text classification. arXiv preprint arXiv:1605.07725 (2016)
12. Patwa, P., Sharma, S., PYKL, S., Guptha, V., Kumari, G., Akhtar, M.S., Ekbal, A., Das, A., Chakraborty, T.: Fighting an infodemic: Covid-19 fake news dataset. arXiv preprint arXiv:2011.03327 (2020)
13. Qian, F., Gong, C., Sharma, K., Liu, Y.: Neural user response generator: Fake news detection with collective user intelligence. In: IJCAI. vol. 18, pp. 3834–3840 (2018)
14. Ruchansky, N., Seo, S., Liu, Y.: Csi: A hybrid deep model for fake news detection. In: Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. pp. 797–806 (2017)
15. Shu, K., Cui, L., Wang, S., Lee, D., Liu, H.: defend: Explainable fake news detection. In: Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining. pp. 395–405 (2019)
16. Shu, K., Mahudeswaran, D., Wang, S., Lee, D., Liu, H.: Fakenewsnet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big Data **8**(3), 171–188 (2020)
17. Tacchini, E., Ballarin, G., Della Vedova, M.L., Moret, S., de Alfaro, L.: Some like it hoax: Automated fake news detection in social networks. arXiv preprint arXiv:1704.07506 (2017)
18. Tzeng, E., Hoffman, J., Zhang, N., Saenko, K., Darrell, T.: Deep domain confusion: Maximizing for domain invariance. Computer Science (2014)
19. Yang, Z., Yang, D., Dyer, C., He, X., Smola, A., Hovy, E.: Hierarchical attention networks for document classification. In: Proceedings of the 2016 conference of the North American chapter of the association for computational linguistics: human language technologies. pp. 1480–1489 (2016)
20. Yosinski, J., Clune, J., Bengio, Y., Lipson, H.: How transferable are features in deep neural networks? MIT Press (2014)
21. Zhang, J., Li, W., Ogunbona, P.: Joint geometrical and statistical alignment for visual domain adaptation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1859–1867 (2017)
22. Zhu, C., Cheng, Y., Gan, Z., Sun, S., Goldstein, T., Liu, J.: Freelb: Enhanced adversarial training for language understanding (2019)