Exploring Deceptive Domain Transfer Strategies: Mitigating the Differences among Deceptive Domains

Sadat Shahriar
University of Houston,
Texas, USA
sshahriar@uh.edu

Arjun Mukherjee
University of Houston,
Texas, USA
arjun@cs.uh.edu

Omprakash Gnawali
University of Houston,
Texas, USA
gnawali@cs.uh.edu

Abstract

Deceptive text poses a significant threat to users, resulting in widespread misinformation and disorder. While researchers have created numerous cutting-edge techniques for detecting deception in domain-specific settings, whether there is a generic deception pattern so that deception-related knowledge in one domain can be transferred to the other remains mostly unexplored. Moreover, the disparities in textual expression across these many mediums pose an additional obstacle for generalization. To this end, we present a Multi-Task Learning (MTL) based deception generalization strategy to reduce the domain-specific noise and facilitate a better understanding of deception via a generalized training. As deceptive domains, we use News (*fake news*), Tweets (*rumors*), and Reviews (*fake reviews*) and employ LSTM and BERT models to incorporate domain transfer techniques. Our proposed architecture for the combined approach of domain-independent and domain-specific training improves the deception detection performance by up to 5.28% in F1-score.

1 Introduction

With the advent of the digital age came a deluge of textual content online, which also contains an enormous amount of deceptive text. The deceptive text uses a variety of strategies to trick readers based on the information delivery medium. Fake news, for instance, spreads false information about a person or organization in order to harm their reputation, while fake reviews intentionally exaggerate the positive or negative aspects of a product or service in order to gain attention. Despite the technical differences in deception, all deceptive texts have the same objective of deceiving people; hence, a generic pattern of deception may exist (Shahriar et al., 2021). Identifying the underlying generic pattern of deception may unravel useful information about textual deception. Furthermore, such a system will enable a more effective detection approach through the intermingling of multiple deception domains.

There has been a good number of work done to combat textual deception in domain-specific situations, but a powerful detection system requires a lot of labeled data, which is dependent on things like trustworthy annotators, resources, time, and money. Consequently, the learning paradigm wherein multiple domains may support each other can be a promising solution in deception detection. In this paper, we explore the feasibility of generalizing deception across domains such as News, Reviews, and Tweets, and we present a Multi-Task Learning (MTL) based deceptive domain transfer strategy to mitigate the domain differences and improve deception detection capacity beyond a standard single domain learning approach with LSTM and BERT models.

Researchers dealt with deceptive domain adaptation problems in cross-dataset learning settings, e.g., Opinion Spam on different entities (Li et al., 2014; Sánchez-Junquera et al., 2020), and Fake News on different topics (Pérez-Rosas et al., 2018). Although attempts have been made to generalize the deception for better detection, these efforts have been constrained so far (Gröndahl and Asokan, 2019). Shahriar et al. explored the problem of holistic deception detection, where they used single deep learning networks to detect deception from a holistic perspective (Shahriar et al., 2021). While this approach provides a domain-agnostic system, it is possible that the intrinsic variations between deception domains mean that a single network cannot give an effective solution. A feature-augmentation-based soft domain transfer approach using the last layer of learned models was proposed in (Shahriar et al., 2022). However, the last layers are prone to capturing the domain-specific noise which may

Proceedings of Recent Advances in Natural Language Processing, pages 1076–1084
Varna, Sep 4–6, 2023
https://doi.org/10.26615/978-954-452-092-2_115

have adverse effects in deceptive transfer. Consequently, the lack of a sufficiently robust system that can account for domain differences and leverage the generic deception signal constitutes a significant research gap.

Considering the above aspects, We formulate the research problem as: Given a set of deceptive domains $\{D_i\}_{i=1}^{n}$, how to construct a generic feature set f_g which can help improve detecting deception in all n domains, rather than the domain-specific feature set $\{f_i\}_{i=1}^{n}$. To address this problem, we use a Multi-Task Learning (MTL) approach with the LSTM and BERT model, where the part of the model is shared across all domains to capture the generic information, and multiple branches downstream to account for domain-specific information. We compare our approach with an **A**ll **F**or **O**ne (OFA) mode of deception generalization and **I**ntermediate **L**ayer **C**oncatenation (ILC) mode of domain transfer (Shahriar et al., 2021, 2022).

This study has a wide range of implications. At the outset, this study seeks to characterize the interconnectedness of various forms of deception and to identify the underlying generic pattern. Learning deception across different domains together allows for the development of a more robust system. It will also take into account the labeled data shortage issue in many deception domains. On top of that, the appearance of a new event can frequently lead to more deceptive data in one domain than the other. In such instances, the generalization of deception studies can be extremely valuable. Finally, the MTL-based simultaneous learning of generic and domain-specific deception will incorporate fewer parameters to be trained than separately learning from the domains.

Our research shows that for all domain-transfer and generalization experiments, MTL outperforms the ILC and OFA mode. Our main contribution can be summarized as follows:

- We explore the deceptive domain transfer strategies and compare them with our proposed MTL-based approach for an improved deception detection system by simultaneously capturing the generalized deception while also preserving the domain differences.
- We show the potential association between the domains by comparing the performance improvement, which may provide useful research direction while performing domain transfer.

2 Datasets

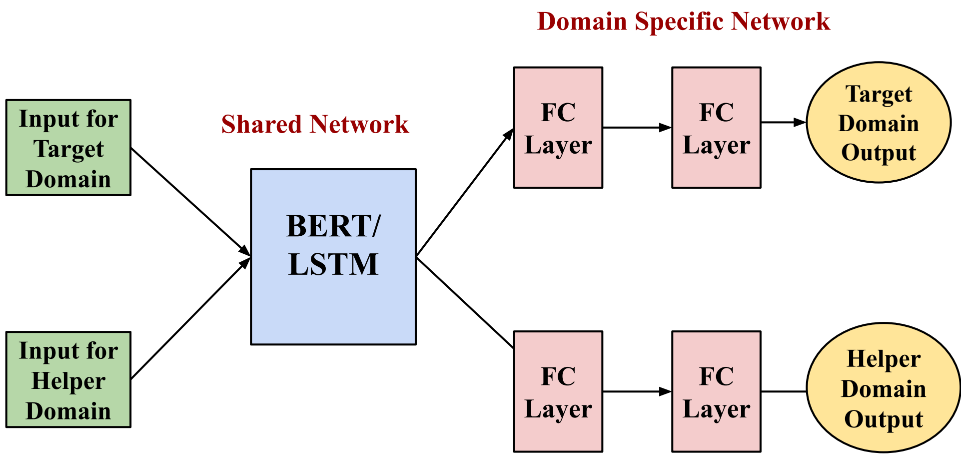
For the three domains, we use six datasets for this paper. For the News domain deception, LIAR dataset contains data from Politifact, and each data is labeled with one of them: True, Mostly-True, Half-True, Mostly-False, False, Pants-on-Fire False. Following the work of Upadhayay and Behzadan 2020, we label the first two as non-fake and the latter four as fake news. Another News dataset, Nela-GT-2021 (NELA) is a source-based labeled news dataset collected from January 2021 to December 2021 and labeled by Media Bias Fact Check (MBFC) (Gruppi et al., 2022). We labeled the news with MBFC *factuality* score 0 as Fake and 5 as non-Fake, and we collect the news sources from the US only. The news domain contains 43,168 news with 64.29% as fake. In the Tweets domain of deception, data comes from PHEME and a collection of *Newly Emerged Rumors in Twitter* (NERT) from 2016 to 2018 (Zubiaga et al., 2016; Bodaghi, 2019). In total we have 20,893 tweets with 49.77% as rumors. For the Reviews domain, we use the Yelp restaurant (RES) and hotel (RES) dataset, which 67,395 reviews, where 13.19% of them are labeled as fake (Mukherjee et al., 2013).

3 Methodology

As the baseline text classification models, we use attention-based LSTM, and BERT models, followed by a FC layer and a softmax layer (Vaswani et al., 2017; Devlin et al., 2018). We explore Intermediate Layer Concatenation (ILC) and Multi-Task Learning (MTL) for deceptive domain transfer strategies.

3.1 Intermediate Layer Concatenation (ILC)

First, the baseline self-domain models are individually trained for detecting deception. The trained models are used as kernels to obtain the target domain's feature representation. Next, the obtained features are concatenated and fed to a Fully-Connected (FC) layer to detect deception. The intuition behind this approach is that by obtaining the feature representation in different domains' high-level latent space, the deceptive text may obtain richer information to detect deception than in its own domain only. The training strategy is adopted from Shahriar et al. 2022.



(a)

Figure 1: Multi-Task Learning (MTL) based deceptive domain transfer. The shared network captures the generalized deceptive pattern and the domain-specific Network accounts for the domain differences.

3.2 Multi-Task Learning

Multi-Task Learning (MTL) aims to exploit the potential information from different training objectives and build a more robust learner (Zhang and Yang, 2017). If we have n learning tasks, where each task is presented as F_i, and i=1 to n, MTL helps ameliorate the task F_i by utilizing the learned knowledge from the n tasks. Based on different learning objectives, MTL can have a different spectrum of supervision, parameters can be hard-shared or soft-shared, and different architectures are employed to account for different task categories (Zhang and Yang, 2017; Caruana, 1997; Ruder, 2017). In this paper, we use a hard parameter shared-based supervised MTL to improve the deception detection performance using a deep-learning-based sequence classification approach.

Our MTL-based domain transfer architecture is depicted in Figure 1. The Target deceptive domain T and the helper deceptive domain H are fed to the MTL architecture. The LSTM, or BERT model is used as the shared network where the model jointly learns the domain-independent hard-shared parameters θ^s. Next, we have a two-layer Fully-Connected (FC) network, followed by a sigmoid layer in two levels for a domain-specific network, which is used to learn the domain-specific parameters θ^T and θ^H. We use cross-entropy loss for each domain and form a combined loss by adding the target domain loss with the weighted (λ) loss of the helper domain. The algorithm for this approach is demonstrated in Algorithm 1.

There are several reasons for MTL being a promising mode of deceptive domain transfer. First, text data is inherently noisy, and so are deceptive domains (Subramaniam et al., 2009; Agarwal et al., 2007). A model trained on self-domain data only can be prone to overfitting due to being modeled on the domain-specific noise. Since two different domains have different noise patterns, training them jointly would achieve better representation by implicit data augmentation. Furthermore, since the deception domains are closely related by their same intention of deceiving the reader, the similarity allows the model to focus on important features than the noise, and the helper domain can provide additional support for the relevance or irrelevance of the focused features (Ruder, 2017). Next, due to the complex nature of deceptive textual data, feature interactions in some domains might be more difficult to learn than others. Hence, MTL allows to eavesdrop on the complex learning process and helps transfer the knowledge from one domain to another. Finally, MTL can act as a regularizer by reducing inductive and representation bias.

4 Experiments and Results

We use 80-20 split for train and test, and 20% from the train set as validation with three random splits. For the Reviews domain, we train with a balanced proportion of fake and non-fake reviews. We compare accuracy and binary F1-score for performance.

Algorithm 1 Multi Task Learning for Domain Tranfer

1: **Input:** Deceptive target domain T, deceptive helper domain H, loss weight of helper domain λ
2: **Output:** Hard-shared parameters θ^s, target domain paramters θ^T, and helper domain parameters θ^H
3: Compute loss for Target domain $L_T(T; \theta^T, \theta^s)$
4: Compute loss for Helper domain $L_H(H; \theta^H, \theta^s)$
5: Combine the losses $L = L_T + \lambda L_H$
6: Update θ^s based on combined loss L
7: Update θ^T based on loss L_T
8: Update θ^H based on loss λL_H

The batch size, learning rate, hidden layers and epochs are chosen by validation set performance.

4.1 Cross-Domain Deception Detection

In the cross-domain (CD) setting, we experiment to see whether deception trained in one domain can be generalized enough to detect deception in another domain. Table 1 shows that deception detection performs best when trained and tested on the same domain. For all three domains, performance drops significantly (p-value<0.05) when tested on different domains. However, in the CD setting, performance being better than the chance implies that there is some domain association present, which can be leveraged for improved deception detection.

4.2 Deceptive Domain Transfer

4.2.1 ILC-based Domain Transfer

In the ILC mode of domain transfer, we utilize the deception information captured in the intermediate layers of the model. While we concatenate the post-attention layers for LSTM, we experiment with different combinations of the last six layers of the BERT model and report the best result in each transfer.

Table 2 shows the ILC mode of domain transfer for the LSTM model. For the News domain, Tweets help the most as a single domain by improving the F1-score by 0.94%, and the combination of all three domains improves the F1-score by 1.26%. For the tweets domain, News helps the most with improvement by .73% and for the Reviews domain, Tweets help the most by .54%. However, the overall performance improvement is less than 1% for all cases. Although the model captures some non-domain deceptive information, the last layer being highly focused on domain-specific deception, the transfer of deceptive information is rather minimal.

In the ILC mode of BERT model, the best performance is found when all three domains of deception are concatenated (Table 3), with improvement over the single domain deception by 2.11%, 2.09%, and 1.23% respectively for News, Tweets and Reviews domain. As individual helper domains, News and Tweets are most helpful to each other and Tweets help the reviews most. It should be emphasized, however, that in none of the scenarios is last layer concatenation useful. For News as helper domain, 9th and 7th layer concatenations were most helpful for Tweets and Reviews respectively. The 9th layer of Tweets was helpful in both cases. For the Reviews as helper domain, 6th, 7th and 8th layers have similar transfer performance but decline significantly from the 9th to the last layer.

4.2.2 MTL-based Domain Transfer

The use of MTL-based domain transfer ensures that the model captures generalized deceptive information at shared layers while accounting for domain-specific knowledge in domain-specific layers. To accommodate for the inter-domain data imbalance, we employ two training strategies: **regular** training where every batch will retain its original data distribution, and **balanced** training in which we upsample or downsample other domains to the target domains training size. We perform the experiments with different combinations of the loss function and report the results with the best validation set performance.

The table 2 and 3 show that MTL-based domain transfer outperforms the ILC-based domain transfer in all cases. For the LSTM model, combining all three domains helps in performance boost from the single-domain model by 1.96%, 4.38% and 0.70% in News, Tweets, and Reviews respectively. The improvement is on average 1.63% more than the ILC mode.

The performance boost with MTL-based BERT model is higher than in the LSTM model. We find the average F1-score improvement with combined domains to be 4.63%, 3.65%, and 5.28% respectively over the single-domain models, and an average of 2.70% more than the ILC model. The best helper domain for each target domains are consistent with ILC mode for both BERT and

		TO: News (acc/f1)	TO: Tweets (acc/f1)	TO: Reviews (acc/f1)
LSTM	News	**72.57/80.13**	59.03/71.01	63.01/77.31
	Tweets	50.11/65.52	**68.72/67.22**	49.84/64.53
	Reviews	48.79/22.76	52.21/23.19	**63.43/32.58**
BERT	News	**80.19/86.21**	63.01/77.31	62.05/76.28
	Tweets	52.09/65.65	**75.83/75.39**	52.48/63.53
	Reviews	76.19/23.62	64.98/26.31	**56.62/31.76**

Table 1: Cross-Domain deception detection while **T**rained **O**n (TO) different deception domains. Performance drops significantly when trained on one domain but tested on a different domain.

		News (acc/f1)	Tweets (acc/f1)	Reviews (acc/f1)
	self-domain	72.57/80.13	68.72/67.22	63.43/32.58
ILC	Tweets+Reviews	62.18/75.82	68.83/67.55	64.08/33.12
	News+Tweets	73.15/81.07	68.93/67.95	55.88/24.01
	News+Reviews	73.11/80.89	57.11/64.08	63.98/32.66
	News+Tweets+Reviews	73.99/81.39	68.49/67.89	64.12/32.80
MTL	Tweets+Reviews	61.97/76.45	69.56/71.35	66.58/**33.31**
	News+Tweets	74.80/81.26	**71.07**/71.55	51.27/24.17
	News+Reviews	74.03/81.11	57.82/64.10	61.55/32.01
	News+Tweets+Reviews	**75.83/82.09**	69.78/**71.60**	**67.05**/33.28

Table 2: Deceptive domain transfer using LSTM model. We observe the MTL mode performing better than ILC in almost all cases.

		News (acc/f1)	Tweets (acc/f1)	Reviews (acc/f1)
	self-domain	80.19/86.21	75.83/75.39	56.62/31.76
ILC	Tweets+Reviews	65.19/76.92	75.91/77.28	57.24/32.83
	News+Tweets	83.07/88.25	76.03/76.52	65.60/23.67
	News+Reviews	81.01/86.35	53.04/63.93	55.67/31.94
	News+Tweets+Reviews	83.10/88.32	76.17/77.48	56.12/32.99
MTL	Tweets+Reviews	65.95/77.11	76.36/78.32	68.29/36.26
	News+Tweets	87.79/90.45	74.80/78.09	65.51/22.09
	News+Reviews	86.27/89.41	54.38/66.51	67.47/35.39
	News+Tweets+Reviews	**88.39/90.84**	**77.08/79.02**	**74.81/37.04**

Table 3: Deceptive domain transfer using BERT model. The best performance across all domains are achieved with MTL mode and while trained with all three domains.

LSTM model. We further find that balanced training works best for News and Tweets, and regular training works best for Reviews.

4.3 Generalized Deception Detection

In the generalized deception detection setting, we simultaneously learn deception in different domains. We use two architectures for that. First, in the **O**ne **F**or **A**ll (OFA) mode, we mix the training data from all different domains and use a single network (BERT or LSTM) for all domains without the model being aware of the domain differences. In the MTL mode, the shared layers are used for generalization and the task-specific layers are used to account for the domain differences. Note that we do not tune the loss weight parameter and use the same value for each domain.

Table 4 shows the generalized deception detection performance. In the OFA mode, there is only a slight performance boost in News and Tweets, while declining in Reviews for the LSTM model. Since the OFA mode presents a domain-agnostic view, while the model achieves a generalized representation of deception, it fails to capture the domain-specific distinction. The remedy is achieved in the MTL-based generalization by employing the task-specific layer on the top of the shared generalized layer, and thus, outperforms the OFA mode by 2.00% on average.

5 Result Analysis and Discussion

Overall, the ILC mode of domain transfer exhibits less improvement than the MTL mode. This is because, in the MTL mode, the deception domains share a latent space in the upstream layers and are only distinguished by the domain-specific layers on the later levels. On the contrary, the ILC mode can access high-level representations only. Hence MTL mode has a higher chance of learning underlying representations than ILC modes by leveraging information from other domains.

We investigate the performance improvement of the MTL-based LSTM model by plotting the validation loss in the first 10 epochs. Figure 2 shows that the MTL mode achieves better generalizability for all three deceptive domains, whereas self-domain modes tend to overfit quickly. Thus, the improved performance of MTL mode might be attributed to better generalization.

We further explore how deception generalization is conducted on the attention head level with MTL-based BERT model. Table 5 shows an example where most of the attention heads on last four layers focus on “Chief”, “Suspended” and “Helping”. Notably, although all words got some attention scores, none of the heads on the last four layers have the highest attention scores on the word “vaccine”, which might be a key phrase for deception detection on COVID-19 events. In contrast, the baseline BERT model features four attention heads in the final four layers that give the term “Vaccine” the most weight. Important proper nouns, like “Trump” and “Obama” were also analyzed; whilst on the baseline models, these terms receive an average of 21.87% of attention on the last four layers of heads, for the MTL model, this number drops to 12.19%. Thus, rather than focusing on domain-specific deception characteristics, our proposed architecture for the MTL mode may be able to generalize deception.

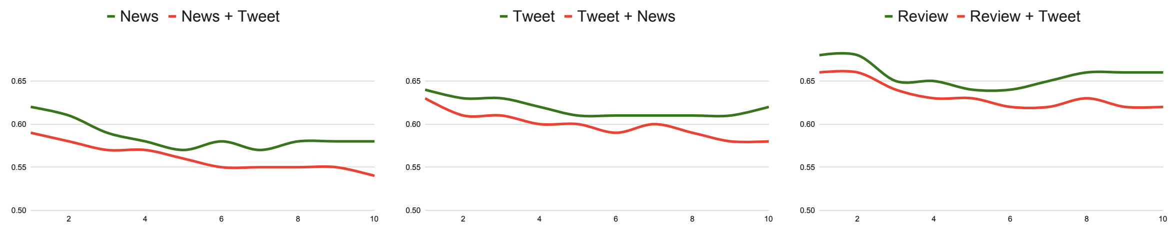
6 Related Works

Most of the previous works in domain transfer dealt with cross-dataset knowledge transfer on different topics from similar information sources. For example, fake news from different news sources and topics are shown to have different word usage and propagation pattern (Silva et al., 2021; Huang and Chen, 2020). Janicka et al. showed that stylometric and psycholinguistic features in different fake news varies widely and results in the performance drop to 20% when train and test sources are different (Janicka et al., 2019). Silva et al. addressed the challenge by storing domain-specific and cross-domain knowledge in embedding representation. (Silva et al., 2021). Sicilia et al. explored how the differences in topics between train and test set affect the performance in rumor detection in the health domain (Sicilia et al., 2018). Ren et al. linearly combined a set of vector representations on different topics with the textual features and obtained an Attention network-based cross-topic solution for rumor detection (REN et al., 2021). In the field of Fake Review detection, Hernández-Castañeda et al. performed a cross-domain fake review detection using three opinion datasets with LDA, SVN, and WSM-based features (Hernández-Castañeda et al., 2017). They also measured the domain association by training on one domain and testing on the other. Sànchez-Junquera et al. proposed a model where they performed a filtering approach for masking domain-specific terms and transformed the original text to a domain-agnostic form (Sánchez-Junquera et al., 2020). Similar works in cross-domain fake review detection was done in (Li et al., 2014) and (Abri et al., 2020).

The existing works on the cross-dataset domain transfer technique suggests that a robust model should exploit both domain-aware and domain-independent attributes for a successful deception detection task. Our proposed method of MTL-based domain transfer technique builds up on shared and domain-specific layers to account for the aforementioned strategy. Nevertheless, the comparative study of deceptive medium-based domain transfer was not explored in previous work to the best of our knowledge. Hence, our method is the first one to address this problem.

		News (acc/f1)	Tweets (acc/f1)	Reviews (acc/f1)
LSTM	OFA	72.95/80.86	65.64/68.54	65.83/30.36
	MTL	73.21/81.14	69.89/71.12	63.88/32.49
BERT	OFA	82.11/86.97	76.94/76.02	70.58/34.18
	MTL	85.35/88.86	77.17/78.71	73.96/36.57

Table 4: Generalized deception detection using OFA and MTL architecture. In all cases, MTL mode outperforms the OFA mode.



(a) (b) (c)

Figure 2: Loss function curve with increase in epoch for validation set for LSTM model. Green represents the self-domain loss and red represents the MTL-based domain transfer loss (a) validation loss for News domain (b) validation loss for Tweet domain, (c) validation loss for Review domain.

Words	**Attention Heads**
Police	H_{10}^{12}, H_{12}^{1}
Chief	$H_{9}^{1,6,8,11}$, $H_{10}^{2,5}$, H_{12}^{2}
Suspended	$H_{9}^{3,7}$, H_{11}^{12}, $H_{12}^{6,9,11}$
For	$H_{10}^{7,8}$, H_{12}^{3}
Helping	$H_{10}^{1,11}$, $H_{12}^{5,10,12}$
Officers	H_{9}^{4}, H_{10}^{9}
Dodge	$H_{11}^{1,2,7,11}$
Vaccine	
Mandate	H_{11}^{9}

Table 5: Attention Heads on MTL-mode of BERT in the last four layers. The heads pay “attention” to different words on the rumor (deception) text *Police Chief Suspended For Helping Officers Dodge Vaccine Mandate*. H_L^N indicates the highest attention in layer L for head number N.

7 Conclusion

Although distinct deception domains have their own methods and characteristics for disseminating deceptive information, they all have the same objective: to deceive individuals. Hence, the generalized detection approach can be immensely useful for addressing labeled data shortage issues in numerous domains. Here, we compare state-of-the-art domain transfer strategies and present an MTL-based method for transferring information across deceptive domains for enhanced deception detection. Our experiments demonstrate that learning deception in multiple domains simultaneously results in improved generalization and performance. In any case, with MTL-based architecture showing promise as a possible option for universal deception detection, we can investigate different hybrid structures of textual parameter sharing and weighted-loss methods for deception detection. Furthermore, the continual learning approach of deception detection can be a promising research direction due to its capability of catastrophic forgetting prevention and knowledge transfer (Biesialska et al., 2020). In addition, we plan to incorporate email and Facebook post deceptions into our future research.

8 Ethics and Broader Impact

Our work has its limitations, considering the complexities inherent in the deceptive content and the ever-evolving landscape of deception. Consequently, the data used in this research may not represent every category of deception and does not consider the cultural nuances of all forms of deception, especially in the current age of *chatGPT* and other LLM-based text generation techniques (ChatGPT; Hacker et al., 2023). We recognize that our study has important ethical implications, particularly with regard to the potential misuse of deception detection techniques. While our research aims to improve the performance of deception detection in various domains, we acknowledge that these techniques could be used to invade individuals' privacy or unfairly target certain groups. Therefore, we urge researchers and practitioners to use these techniques responsibly and with consideration for the potential consequences.

Our research sets the stage for broader implications. The proposed deception detection approach and domain transfer strategies can be extended beyond the domains explored in this paper. We envision their potential application in combating deception in diverse contexts, including online forums, and chat platforms, and addressing the challenges posed by misinformation contents.

Acknowledgments

This research was supported in part by grant ARO W911NF-20-1-0254. The views and conclusions contained in this document are those of the authors and not of the sponsors.

References

Faranak Abri, Luis Felipe Gutierrez, Akbar Siami Namin, Keith S Jones, and David RW Sears. 2020. Fake reviews detection through analysis of linguistic features. *arXiv preprint arXiv:2010.04260*.

Sumeet Agarwal, Shantanu Godbole, Diwakar Punjani, and Shourya Roy. 2007. How much noise is too much: A study in automatic text classification. In *Seventh IEEE International Conference on Data Mining (ICDM 2007)*, pages 3–12. IEEE.

Magdalena Biesialska, Katarzyna Biesialska, and Marta R Costa-Jussa. 2020. Continual lifelong learning in natural language processing: A survey. *arXiv preprint arXiv:2012.09823*.

Amirhosein Bodaghi. 2019. Newly emerged rumors in twitter.

Rich Caruana. 1997. Multitask learning. *Machine learning*, 28(1):41–75.

ChatGPT. Chatgpt, may 3 version.

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

Tommi Gröndahl and N. Asokan. 2019. Text analysis in adversarial settings: Does deception leave a stylistic trace? *ACM Comput. Surv.*, 52(3).

Maurício Gruppi, Benjamin D Horne, and Sibel Adalı. 2022. Nela-gt-2021: A large multi-labelled news dataset for the study of misinformation in news articles. *arXiv preprint arXiv:2203.05659*.

Philipp Hacker, Andreas Engel, and Marco Mauer. 2023. Regulating chatgpt and other large generative ai models. *arXiv preprint arXiv:2302.02337*.

Ángel Hernández-Castañeda, Hiram Calvo, Alexander Gelbukh, and Jorge J. García Flores. 2017. Cross-domain deception detection using support vector networks. *Soft Computing*, 21(3):585–595.

Yin-Fu Huang and Po-Hong Chen. 2020. Fake news detection using an ensemble learning model based on self-adaptive harmony search algorithms. *Expert Systems with Applications*, 159:113584.

Maria Janicka, Maria Pszona, and Aleksander Wawer. 2019. Cross-domain failures of fake news detection. *Computación y Sistemas*, 23(3):1089–1097.

Jiwei Li, Myle Ott, Claire Cardie, and Eduard Hovy. 2014. Towards a general rule for identifying deceptive opinion spam. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1566–1576, Baltimore, Maryland. Association for Computational Linguistics.

Arjun Mukherjee, Vivekanand Venkataraman, B. Liu, and Natalie S. Glance. 2013. What yelp fake review filter might be doing? In *ICWSM*.

Verónica Pérez-Rosas, Bennett Kleinberg, Alexandra Lefevre, and Rada Mihalcea. 2018. Automatic detection of fake news. In *COLING*.

Weijieying REN, Jing Jiang, Ling Min Serena Khoo, and Hai Leong Chieu. 2021. Cross-topic rumor detection using topic-mixtures.

Sebastian Ruder. 2017. An overview of multi-task learning in deep neural networks. *arXiv preprint arXiv:1706.05098*.

Sadat Shahriar, Arjun Mukherjee, and Omprakash Gnawali. 2021. A domain-independent holistic approach to deception detection. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 1308–1317.

Sadat Shahriar, Arjun Mukherjee, and Omprakash Gnawali. 2022. Deception detection with feature-augmentation by soft domain transfer. In *International Conference on Social Informatics*, pages 373–380. Springer.

Rosa Sicilia, Mario Merone, Roberto Valenti, Ermanno Cordelli, Federico D'Antoni, Vincenzo De Ruvo, Patrizia Benedetta Dragone, Sara Esposito, and Paolo Soda. 2018. Cross-topic rumour detection in the health domain. In *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2056–2063. IEEE.

Amila Silva, Ling Luo, Shanika Karunasekera, and Christopher Leckie. 2021. Embracing domain differences in fake news: Cross-domain fake news detection using multi-modal data. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, pages 557–565.

L Venkata Subramaniam, Shourya Roy, Tanveer A Faruquie, and Sumit Negi. 2009. A survey of types of text noise and techniques to handle noisy text. In *Proceedings of The Third Workshop on Analytics for Noisy Unstructured Text Data*, pages 115–122.

Javier Sánchez-Junquera, Luis Villaseñor-Pineda, Manuel Montes y Gómez, Paolo Rosso, and Efstathios Stamatatos. 2020. Masking domain-specific information for cross-domain deception detection. *Pattern Recognition Letters*, 135:122–130.

Bibek Upadhayay and Vahid Behzadan. 2020. Sentimental liar: Extended corpus and deep learning models for fake claim classification.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems*, 30.

Yu Zhang and Qiang Yang. 2017. An overview of multi-task learning. *National Science Review*, 5(1):30–43.

Arkaitz Zubiaga, Geraldine Wong Sak Hoi, Maria Liakata, and Rob Procter. 2016. PHEME dataset of rumours and non-rumours.