Enhancing Society-Undermining Disinformation Detection through Fine-Grained Sentiment Analysis Pre-Finetuning

Tsung-Hsuan Pan,[1] **Chung-Chi Chen,**[2] **Hen-Hsen Huang,**[3] **Hsin-Hsi Chen**[1]
[1] Department of Computer Science and Information Engineering,
National Taiwan University, Taiwan
[2] AIST, Japan
[3] Institute of Information Science, Academia Sinica, Taiwan
b08902138@ntu.edu.tw, c.c.chen@acm.org,
hhhuang@iis.sinica.edu.tw, hhchen@ntu.edu.tw

Abstract

In the era of the digital world, while freedom of speech has been flourishing, it has also paved the way for disinformation, causing detrimental effects on society. Legal and ethical criteria are insufficient to address this concern, thus necessitating technological intervention. This paper presents a novel method leveraging pre-finetuning concept for efficient detection and removal of disinformation that may undermine society, as deemed by judicial entities. We argue the importance of detecting this type of disinformation and validate our approach with real-world data derived from court orders. Following a study that highlighted four areas of interest for rumor analysis, our research proposes the integration of a fine-grained sentiment analysis task in the pre-finetuning phase of language models, using the GoEmotions dataset. Our experiments validate the effectiveness of our approach in enhancing performance significantly. Furthermore, we explore the application of our approach across different languages using multilingual language models, showing promising results. To our knowledge, this is the first study that investigates the role of sentiment analysis pre-finetuning in disinformation detection.

1 Introduction

The advent of digitalization has significantly impacted societal discourse, notably manifesting in the phenomenon of "Fake News," a term so ubiquitous that it was selected as The Macquarie Dictionary Word of the Decade.[1] Fake news and disinformation have infiltrated every aspect of our lives, from politics and elections (Grinberg et al., 2019), to financial markets (Clarke et al., 2020; Kogan et al., 2020), and public health narratives (Hansen and Schmidtblaicher, 2021; Loomba et al., 2021). Based on the intentions behind their dissemination, false information can be classified into two categories: misinformation and disinformation (Hernon, 1995). The former results from honest mistakes, while the latter is deliberately spread with malicious intent.

Judgement	Example
Punishable	The underworld member kept beating the victims in the private guest house.
Impunity	The government holds a ministerial meeting and order expensive lunch box from the restaurant with Michelin star.

Table 1: Example from Court Orders

However, our contention is that this classification system fails to account for the varying degrees of severity inherent in disinformation instances. As illustrated in Table 1, although both examples represent false information, one instance, as judged by the court, is considered society-undermining disinformation and punishable, while the other is not. This distinction underscores our argument that detecting and combating society-undermining disinformation should be a priority focus, and such a task warrants substantial attention. Therefore, this paper seeks to contribute to this area by performing experiments on a real-world dataset. The labels for this dataset are uniquely derived from court orders and provided by judges, thereby granting a legal perspective on what constitutes society-undermining disinformation.

The question of what constitutes society-undermining rumors was initially probed by Chen et al. (Chen et al., 2021). They identified four key research directions: (1) the intention of the writer, (2) the tone of the writer, (3) the sentiment of the reader, and (4) the topic of the post. Building upon this analysis, we propose a research question: how much can the performance of a language model be enhanced in detecting society-undermining disinformation if it is trained to better understand sen-

[1] https://www.macquariedictionary.com.au/blog/article/780/

Findings of the Association for Computational Linguistics: EACL 2024, pages 1371–1377
March 17-22, 2024 ©2024 Association for Computational Linguistics

timent? In an attempt to answer this, we adopt a pre-finetuning strategy wherein we equip language models with a fine-grained sentiment analysis task, utilizing the GoEmotions dataset (Demszky et al., 2020). Furthermore, we conduct experiments under various settings to test the effectiveness of our approach. Encouragingly, our experimental results indicate that the proposed pre-finetuning method significantly improves performance.

To verify the universality of our approach across different languages, we conduct further experiments with multilingual language models. Additionally, we translate all instances into another language for comparison. The results of these cross-lingual experiments corroborate that the proposed pre-finetuning strategy is indeed beneficial across multiple language application scenarios. As far as we know, this study represents the first attempt to investigate the potential of a sentiment analysis pre-finetuning task in enhancing society-undermining disinformation detection capabilities.

2 Related Work

Though the role of sentiment features in fake news detection has been examined extensively (Castillo et al., 2011; AlRubaian et al., 2015; Popat et al., 2017; Ajao et al., 2019; Anoop et al., 2020; Zhang et al., 2021; Alonso et al., 2021; Yang et al., 2023), it is noteworthy that little attention has been directed towards severe instances of disinformation, particularly society-undermining disinformation. Chen et al. (2021) delineated a research agenda for society-undermining disinformation detection, but did not propose a specific method to address this concern. Dharawat et al. (2022) introduced the concept of harmfulness assessment in relation to COVID-19 misinformation. Our research represents a pioneering effort to tackle society-undermining disinformation. Further distinguishing our work is our exploration of the role of sentiment pre-finetuning tasks within this context, a topic which, to our knowledge, has not been previously explored.

3 Dataset

In this study, our primary focus lies in the identification of disinformation that judges deem detrimental to societal harmony. Consequently, we align our approach with the previous study (Chen et al., 2021), using court orders as our primary data source. Our dataset[2] has been amassed by the news vendor, READr,[3], extracting information from the government's Law and Regulations Retrieving System,[4] and is shared under the CC0 License.

	2020-2019	2018-2007
Impunity/Innocent	360	38
Punishable	103	19
# of Court Orders	463	57

Table 2: Statistics of Court Orders.

The instances in our dataset revolve around lawsuits filed under Paragraph 5, Article 63 of Taiwan's Social Order Maintenance Act, which condemns:

> Spreading rumors in a way that is sufficient to undermine public order and peace.

The dataset statistics, related to Paragraph 5, Article 63 of the Social Order Maintenance Act in Taiwan, are illustrated in Table 2. The data reveals a remarkable increase in cases during 2019-2020, corresponding to the period of the 2020 presidential election. A notable observation is the high rate of impunity, reflecting the "chill effect" concerns (Schauer, 1978) as indicated by Chen et al. (2021). The chill effect posits that the fear of potential legal backlash may inhibit individuals from expressing their opinions, eventually leading to a reluctance in sharing information.

In order to mitigate the potential for such stifling of free speech on future social media platforms, we propose that only severe disinformation, capable of undermining societal harmony, should be promptly identified and removed from the platform. Other posts, like the impunity examples in Table 1, should be allowed to remain part of the discourse and can be clarified through ongoing discussion. Accordingly, our experimental setup is geared towards a binary classification scenario: determining whether a given text would be deemed punishable by a judge under Paragraph 5, Article 63 of Taiwan's Social Order Maintenance Act.

Given the unique nature of our dataset and the difficulty of reproducing similar scenario-based annotations, we utilize all available instances, embracing the real-world challenges of few-shot learning and class imbalance. To ensure a substantial test

[2]https://github.com/readr-media/readr-data/tree/master/fake_news

[3]https://www.readr.tw/

[4]https://law.judicial.gov.tw/LAWENG/default.aspx

	Input Language	Accuracy	Precision	Recall	F1
BERT-Chinese	Chinese	0.35	0.62	0.35	0.30
+ Pre-Finetuning with Fine-grained SA	Chinese	**0.72**	**0.99**	**0.72**	**0.83**
mBERT	Chinese	0.31	0.52	0.31	0.23
+ Pre-Finetuning with Fine-grained SA	Chinese	**0.72**	0.92	**0.72**	0.80
BERT	English	0.71	0.58	0.71	0.60
+ Pre-Finetuning with Fine-grained SA	English	0.70	0.91	0.70	0.78
mBERT	English	0.28	0.08	0.28	0.13
+ Pre-Finetuning with Fine-grained SA	English	0.67	0.80	0.67	0.72

Table 3: Experimental Results.

set, we divide the dataset into two halves, with 50% of instances assigned to the training set and the remaining to the test set. We make our dataset publicly available for replication and further investigation, adhering to the same licensing terms as used by READr.[5]

4 Methods

Drawing inspiration from the logic that judges apply in their courtroom decisions:

> Although it is improper for the transferred person to post without verification and judgment, this post does not cause the listeners to fear or panic due to the untruth.

we observed that negative sentiments, such as "fear" and "panic," play a significant role in society-undermining disinformation. Proceeding with this understanding, we adapt the concept of pre-finetuning to enhance the sensitivity of language models to fine-grained sentiment analysis (henceforth denoted as fine-grained SA). The pre-finetuning approach has demonstrated utility in extensive multi-task learning contexts (Aghajanyan et al., 2021) and for particular applications (Chen et al., 2023). For the proposed task, we utilize the GoEmotions dataset (Demszky et al., 2020), consisting of 58k comments sourced from Reddit, to pre-finetune BERT, BERT-Chinese, and multilingual BERT (mBERT) (Devlin et al., 2019). The instances in GoEmotions are annotated with 27 distinct emotion labels.

To facilitate pre-finetuning of BERT-Chinese using the GoEmotions dataset, we translate all instances to Chinese using the Google Translation API. As the nature of these court orders is unique, we aim to replicate application scenarios in other languages to identify potential performance gaps and assess the performance of a universal language model, i.e., mBERT. For pre-finetuning mBERT, we explore two settings: (1) using the original GoEmotions dataset, and (2) using the translated GoEmotions dataset.

[5]https://github.com/TsungHsuan-Pan/Undermine-Society-Rumor-Detection

5 Experiment

5.1 Model Comparison

We assess our results based on various metrics, including accuracy, precision, recall, and F1 score. Table 3 outlines the experimental outcomes of different language models with and without the proposed pre-finetuning strategy.

Firstly, we observe an improvement in the detection of society-undermining disinformation when applying our pre-finetuning method, regardless of the model or language used. Secondly, the pre-finetuned BERT-Chinese model outperforms all other models, aligning with our expectation considering the original dataset is in Chinese. However, this finding reinforces that translating the GoEmotions dataset is a viable approach for the task at hand. Thirdly, an intriguing observation is that when we translate all instances in the court orders to English, the original BERT model (without pre-finetuning) shows the best performance among all original models. We hypothesize that this could be due to the simplification of instances post-translation, potentially reducing noise in the input data. Hence, translation might present a promising avenue for future work in this area. Lastly, the results of the pre-finetuned mBERT with English input data suggest that the court order dataset can be applied for detecting disinformation in other languages.

5.2 Performance on Fine-Grained SA

Table 4 presents the performance metrics for various models tasked with fine-grained sentiment analysis (SA). More specifically, these metrics pertain to the models in their pre-finetuning state, evalu-

	Accuracy	Precision	Recall	F1
BERT-Chinese	0.46	0.55	0.48	0.51
mBERT-Chinese	0.40	0.50	**0.55**	**0.52**
BERT	0.45	0.54	0.47	0.49
mBERT-English	**0.49**	**0.57**	0.51	**0.52**

Table 4: Performance on Fine-grained SA. mBERT-Chinese and mBERT-English denote the mBERT with Chinese and English input data, respectively.

	Accuracy	Precision	Recall	F1
BERT-Chinese	0.75	**0.73**	0.75	0.69
mBERT-Chinese	0.64	0.60	0.64	0.61
BERT	0.69	0.62	0.69	0.62
mBERT-English	**0.78**	**0.73**	**0.78**	**0.72**

Table 5: Performances on negative sentiment identification.

		Accuracy	P	R	F1
BERT-Chinese	Negative	0.58	0.56	0.58	0.57
	DFS	0.47	0.60	0.47	0.49
	CDFS	0.65	0.54	0.65	0.58
mBERT-Chinese	Negative	0.32	0.52	0.32	0.27
	DFS	0.61	0.58	0.61	0.59
	CDFS	0.70	0.62	0.70	0.62
BERT	Negative	0.54	0.60	0.54	0.56
	DFS	0.52	0.60	0.52	0.55
	CDFS	0.67	0.59	0.67	0.61
mBERT-English	Negative	0.56	0.59	0.56	0.57
	DFS	0.65	0.58	0.65	0.61
	CDFS	0.70	0.63	0.70	0.62
XLM-RoBERTa-English	Negative	0.53	0.58	0.53	0.55
	DFS	0.31	0.76	0.31	0.19
	CDFS	0.72	0.62	0.72	0.61
PFT BERT-Chinese	CDFS	0.68	0.64	0.68	0.65
PFT BERT-Chinese	All	**0.72**	**0.99**	**0.72**	**0.83**

Table 6: Results based on different sentiment labels. P and R denote precision and recall. PFT denotes Pre-finetuned.

ated on the fine-grained SA task. From an F1 score standpoint, we observe that the models yield comparable performances. However, it's noteworthy that multilingual BERT models (mBERTs) achieve higher F1 scores than their BERT counterparts for specific languages.

Considering the criticality of negative sentiments in the society-undermining disinformation detection task, we conduct a more nuanced performance analysis on this aspect. As per Demszky et al. (2020), eleven sentiment labels—anger, annoyance, disappointment, disapproval, disgust, embarrassment, fear, grief, nervousness, remorse, and sadness—are classified as negative sentiments. Table 5 details the comprehensive performance on these labels. Among the models, mBERT-English outperforms BERT in negative sentiment identification. By juxtaposing the F1 scores from Table 4 and Table 5, it becomes clear that mBERT performs superiorly to BERT in English fine-grained SA. This observation, however, is not mirrored in the results obtained from Chinese data.

5.3 Role of Negative Sentiments

Delving deeper into the role of fine-grained SA in the society-undermining disinformation detection task, we propose using sentiment labels as markers to identify potential society-undermining disinformation content. In our view, utilizing all negative labels for this purpose is an overly broad approach, which may not suitably align with the specificities of the proposed task. Therefore, we propose two subsets of sentiment labels, both potentially indicative of society-undermining disinformation: (1) **DFS**: disgust, fear, and sadness, and (2) **CDFS**: confusion, disappointment, fear, and sadness.

Table 6 lays out the experimental results of the society-undermining disinformation detection task. Firstly, we observe that sentiment labels belonging to the CDFS group facilitate superior performance in both Chinese and English scenarios compared to the DFS group. This supports our contention that relying solely on a generic negative label is an overly simplistic approach for optimizing performance in the task at hand. Secondly, a comparison of results highlights the performance gap between label-based methods and the pre-finetuning scheme (as evidenced by pre-finetuned BERT-Chinese). Thirdly, our results using XLM-RoBERTa (Conneau et al., 2020) confirm the stability of our findings across various cross-lingual models. We additionally pre-finetune BERT-Chinese exclusively using CDFS. While its performance is inferior to that of BERT-Chinese pre-finetuned with all sentiment labels, it exhibits significant improvement over standard language models. These results suggest that improving society-undermining disinformation detection through pre-finetuning with fine-grained SA is a promising avenue for further research and development.

5.4 Exploration with LLM

Large Language Models (LLMs) exhibit robust general performance and possess multilingual capabilities. This section details the results obtained with GPT-3.5. We examine two prompts for comparative analysis. The first prompt (P1) inquires, "Is the following statement guilty or not?" The second prompt (P2) adds context, stating, "Spreading

	Accuracy	Precision	Recall	F1
Chinese (P1)	0.53	0.29	0.50	0.37
English (P1)	0.54	0.25	0.31	0.28
Chinese (P2)	0.35	0.28	0.81	0.42
English (P2)	0.40	0.28	0.71	0.40

Table 7: Performances of GPT-3.5.

rumors that are sufficient to disturb public peace" constitutes guilt, otherwise it does not. Table 7 presents these findings. The F1 score of GPT-3.5 is similar in both Chinese and English when employing P2, and its performance surpasses that of standard pre-trained language models. Nonetheless, there remains a notable disparity compared to the pre-finetuned models. This observation emphasizes the value of adopting a tailored approach for specific tasks.

6 Conclusion

This paper has shed light on a critical, yet often overlooked, aspect of the discourse around false information: the detection of society-undermining disinformation. By conducting a series of rigorous experiments, we have established a notable connection between such disinformation and fine-grained sentiment labels. Our innovative pre-finetuning approach equips language models with enhanced capabilities to detect such disinformation, improving their performance significantly across multiple language scenarios. Moreover, the cross-lingual applicability of our pre-finetuning methodology underscores its robustness and versatility. It sets the stage for future investigations that could further refine this approach for different languages.

However, we recognize that this is only the beginning. The insights and results obtained in this study represent a preliminary step towards a comprehensive understanding of society-undermining disinformation and the development of robust detection strategies. Future research should continue to delve deeper into the complex interplay between disinformation, sentiment, and societal impact, exploring the diverse avenues we have outlined.

Acknowledgments

This work was supported by National Science and Technology Council, Taiwan, under grants MOST 110-2221-E-002-128-MY3, NSTC 112-2634-F-002-005 -, and Ministry of Education (MOE) in Taiwan, under grants NTU-112L900901. The work of Chung-Chi Chen was supported in part by JSPS KAKENHI Grant Number 23K16956 and a project JPNP20006, commissioned by the New Energy and Industrial Technology Development Organization (NEDO).

Limitations

Building on the findings and limitations of this paper, we propose several directions for future work. (1) **Cross-country studies**: This study was limited by the availability of court orders fitting the proposed application scenarios only from one country. Future research could extend this study by analyzing similar cases across different countries. An understanding of how different countries approach the concept of society-undermining disinformation could significantly enrich the current body of knowledge. (2) **Reader sentiment analysis**: In our work, we focused on the sentiment of the writer as it played a crucial role in the disinformation classification. However, the sentiment of the reader may also hold valuable insights in understanding and detecting society-undermining disinformation. Future research could consider constructing a dataset similar to GoEmotions to capture and analyze reader sentiment. (3) **Ethical considerations in application**: As we noted in our ethical considerations, there's a delicate balance between freedom of speech and the need to mitigate the spread of harmful disinformation. Future research should consider this balance, particularly when developing models and tools designed to detect and filter such disinformation. It's essential to ensure that these tools are not used to unjustly limit freedom of speech. (4) **Deepening sentiment analysis**: This paper made strides in applying sentiment analysis for the pre-finetuning of models to detect society-undermining disinformation. Future research could further explore this area, delving deeper into the nuances of sentiment and emotion expressed in disinformation instances. More complex sentiment analysis could uncover subtle cues and patterns that could be instrumental in enhancing detection methods. (5) **Multilingual and multicultural studies**: We found that the proposed pre-finetuning strategy was beneficial across multiple language application scenarios. Future research could extend this line of inquiry, examining the application of this approach in a variety of languages and cultural contexts. Such research could provide valuable insights into the universal and language-specific aspects of society-undermining disinformation.

By exploring these directions, we can continue to build on the contributions of this paper, advancing our understanding of society-undermining disinformation and improving our methods for detecting and combatting it.

Ethical Note

Freedom of speech is one of the core universal values. The trade-off between the scope of freedom of speech and the limitation to the freedom of speech is discussed for a long time but is still an open question. This paper proposes a research direction that may have a risk of limiting the freedom of speech, but it could also prevent the harmful disinformation from spreading in our society. Since things could be double edged sword, we argue that understanding the properties of society-undermining disinformation from different aspects is always a good topic for improving the utility of our society and discussing potential threatens in our society.

References

Armen Aghajanyan, Anchit Gupta, Akshat Shrivastava, Xilun Chen, Luke Zettlemoyer, and Sonal Gupta. 2021. Muppet: Massive multi-task representations with pre-finetuning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 5799–5811, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics.

Oluwaseun Ajao, Deepayan Bhowmik, and Shahrzad Zargari. 2019. Sentiment aware fake news detection on online social networks. In *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2507–2511. IEEE.

Miguel A Alonso, David Vilares, Carlos Gómez-Rodríguez, and Jesús Vilares. 2021. Sentiment analysis for fake news detection. *Electronics*, 10(11):1348.

Majed AlRubaian, Muhammad Al-Qurishi, Mabrook Al-Rakhami, Sk Md Mizanur Rahman, and Atif Alamri. 2015. A multistage credibility analysis model for microblogs. In *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 1434–1440. IEEE.

K Anoop, Deepak Padmanabhan, and VL Lajish. 2020. Emotion cognizance improves health fake news identification. In *24th International Database Engineering & Applications Symposium*. Association for Computing Machinery (ACM).

Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In *Proceedings of the 20th international conference on World wide web*, pages 675–684.

Chung-Chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2021. Which kind of rumors may undermine society: perspectives from court orders. In *Proceedings of the 2021 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, pages 14–17.

Chung-Chi Chen, Hiroya Takamura, Ichiro Kobayashi, and Yusuke Miyao. 2023. Improving numeracy by input reframing and quantitative pre-finetuning task. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 69–77, Dubrovnik, Croatia. Association for Computational Linguistics.

Jonathan Clarke, Hailiang Chen, Ding Du, and Yu Jeffrey Hu. 2020. Fake news, investor attention, and market reaction. *Information Systems Research*, 32(1):35–52.

Alexis Conneau, Kartikay Khandelwal, Naman Goyal, Vishrav Chaudhary, Guillaume Wenzek, Francisco Guzmán, Edouard Grave, Myle Ott, Luke Zettlemoyer, and Veselin Stoyanov. 2020. Unsupervised cross-lingual representation learning at scale. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 8440–8451, Online. Association for Computational Linguistics.

Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. GoEmotions: A dataset of fine-grained emotions. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4040–4054, Online. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. 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, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Arkin Dharawat, Ismini Lourentzou, Alex Morales, and ChengXiang Zhai. 2022. Drink bleach or do what now? covid-hera: A study of risk-informed health decision making in the presence of covid-19 misinformation. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 16, pages 1218–1227.

Nir Grinberg, Kenneth Joseph, Lisa Friedland, Briony Swire-Thompson, and David Lazer. 2019. Fake news on twitter during the 2016 us presidential election. *Science*, 363(6425):374–378.

Peter Reinhard Hansen and Matthias Schmidtblaicher. 2021. A dynamic model of vaccine compliance: how fake news undermined the danish hpv vaccine program. *Journal of business & economic statistics*, 39(1):259–271.

Peter Hernon. 1995. Disinformation and misinformation through the internet: Findings of an exploratory study. *Government information quarterly*, 12(2):133–139.

Shimon Kogan, Tobias J Moskowitz, and Marina Niessner. 2020. Fake news in financial markets. *Available at SSRN 3237763*.

Sahil Loomba, Alexandre de Figueiredo, Simon J Piatek, Kristen de Graaf, and Heidi J Larson. 2021. Measuring the impact of covid-19 vaccine misinformation on vaccination intent in the uk and usa. *Nature human behaviour*, 5(3):337–348.

Kashyap Popat, Subhabrata Mukherjee, Jannik Strötgen, and Gerhard Weikum. 2017. Where the truth lies: Explaining the credibility of emerging claims on the web and social media. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 1003–1012.

Frederick Schauer. 1978. Fear, risk and the first amendment: Unraveling the chilling effect. *BUL rev.*, 58:685.

Sin-han Yang, Chung-chi Chen, Hen-Hsen Huang, and Hsin-Hsi Chen. 2023. Entity-aware dual co-attention network for fake news detection. In *Findings of the Association for Computational Linguistics: EACL 2023*, pages 106–113, Dubrovnik, Croatia. Association for Computational Linguistics.

Xueyao Zhang, Juan Cao, Xirong Li, Qiang Sheng, Lei Zhong, and Kai Shu. 2021. Mining dual emotion for fake news detection. In *Proceedings of the Web Conference 2021*, pages 3465–3476.