Erik Rodriguez
Arielle Posadas
Dr. Karen Mazidi
CS 4395.001

Which side are you on? Insider-Outsider classification in conspiracy-theoretic social media

https://aclanthology.org/2022.acl-long.341.pdf

Author list and their affiliations

Pavan Holur: Electrical and Computer Engineering Ph.D. student, UCLA
Tianyi Wang: Electrical and Computer Engineering Ph.D. student, UCLA
Shadi Shashsvari: Electrical and Computer Engineering Ph.D. student, UCLA
Timothy Tangherlini: Department of Scandinavian Professor, UC Berkeley
Vwani Roychowdhury: Department of Electrical and Computer Engineering Professor, UCLA

The problem addressed by the paper

The emergence of social media narratives has led to the rise of conspiracy theories. These narratives typically adopt a semantic structure featuring insider and outsider groups. Insiders are individuals or groups who share commonalities with the authors, while outsiders are perceived as threatening insiders. The challenge in this problem lies in correctly identifying insiders and outsiders due to the abundance of poorly constructed posts. Moreover, the classification of these groups is highly contextual and often implicit, making it transitive. To complicate matters, common negative sentiment patterns do not always align with phrases that imply an individual's status as an outsider.

Prior work

Pavan Holur, Tianyi Wang, and Shadi Shashsvari, the three authors of this paper, have previously worked together to research automatic detection methods of emerging COVID-19 conspiracy theories. They presented various narrative frameworks to analyze how they attach to the broader scope of the pandemic. They concluded that studying how people perceived events such as COVID-19 is helpful as it can aid in devising methods to disrupt the spread of misinformation. [1]

In addition to the authors, the research was guided by the expertise of Dr.Tangherlini and Dr. Roychowdhury. Dr. Tangherlini is a professor at UC Berkeley who has written multiple publications for the Association of Computing Machinery. He has an interest in computational approaches to problems in the study of literature and culture. Dr. Roychowdhury is a professor at UCLA and has considerable research on adaptive learning algorithms. [2][3]

The unique contributions of this paper.
The authors address the challenges involved in identifying the members of groups on social media platforms that promote conspiracy theories. The paper indicates that it is possible to create algorithms that can accurately identify semantically-labeled narratives, even if there are several overlapping ones. As a result, this helps make sense of fragmented conversational threads seen on social media.

How the authors evaluated their work.
The authors developed a dataset named Conspiracy Theory 5000 (CT5K) by collecting a corpus of social media posts using web crawlers. The posts were centered on the topic of vaccine hesitancy during the COVID-19 pandemic.

The model they created was the Noun-Phrase-to-Insider-Outside (NP2IO) model. The model uses a token classification design, which includes a BERT-like pre-trained model as its primary structure and a softmax classifier placed on top of the pre-trained model. The pre-trained model serves as the backbone of the architecture, while the softmax classifier is responsible for assigning labels to each token in the input text. First, the label of a noun phrase is extended to all of its individual tokens. But when the model makes predictions, the label assigned with the most votes from the tokens of a noun phrase is used to classify the entire noun phrase.

The authors used baseline models to compare the performance of the NP2IO model. These models include random, deterministic, naive Bayes models with lemmatized and non-lemmatized words and three different contextual GloVe+CBOW+XGBoost (CBOW) models. They created a table that compared the metric scores (accuracy, precision, recall, F1 score, and weighted F1 score) for each model. The results show that random and deterministic models perform poorly, while naïve Bayes and its lemmatized form were better but worse than the two contextual models. NP2IO had the highest metric scores, which were noticeably higher than the CBOW models.

A table also tested the models' zero-shot performances when the noun phrases were not seen during training. This tests out the generalization of each model. The performance of the naive Bayes models drops significantly to where it’s seen as random. However, the models CBOW and NP2IO kept their strong performance, indicating both are effective in identifying context to predict labels accurately.

Finally, the authors checked if NP2IO can accurately classify noun phrases that have biased labels in the CT5K dataset. They chose three terms ("microchip," "government," and "chemical") that are mostly labeled as outsiders and created new samples that suggest they are insiders. Next, they checked how well NP2IO could detect these insider labels. The results showed that NP2IO could classify insiders moderately well, implying that the dataset's biased labels did not significantly affect the model's performance.

Impact of their work

The authors of this paper have a related publication from 2020 called “Conspiracy in the Time of Corona: automatic detection of Emerging COVID-19 conspiracy theories in social media and the News.” which highlights their impact on their field and has received 280 citations [1]. This new paper is a follow-up to their previous work and is significant on its own. Their latest research has not yet received any citations in Google Scholar. However, the discoveries presented in this paper have the potential to allow us to analyze the current climate of what different groups of people are thinking, which in turn can help reduce the spread of misinformation. By understanding the beliefs and attitudes of distinct communities, we can create effective strategies to contribute to a more informed and educated society.

Citations

[1] Shahsavari S, Holur P, Wang T, Tangherlini TR, Roychowdhury V. Conspiracy in the time of corona: automatic detection of emerging COVID-19 conspiracy theories in social media and the news. J Comput Soc Sci. 2020;3(2):279-317. doi: 10.1007/s42001-020-00086-5. Epub 2020 Oct 28. PMID: 33134595; PMCID: PMC7591696.

[2] T. Tangherlini, "Timothy Tangherlini - Berkeley Research," UC Berkeley, 2021. [Online]. Available: https://vcresearch.berkeley.edu/faculty/timothy-tangherlini. [Accessed: Apr. 06, 2023].

[3] V. P. Roychowdhury, "Vwani P. Roychowdhury - Electrical and Computer Engineering - UCLA," UCLA Electrical and Computer Engineering Department, 2021. [Online]. Available: https://www.ee.ucla.edu/vwani-p-roychowdhury/. [Accessed: Apr. 06, 2023].