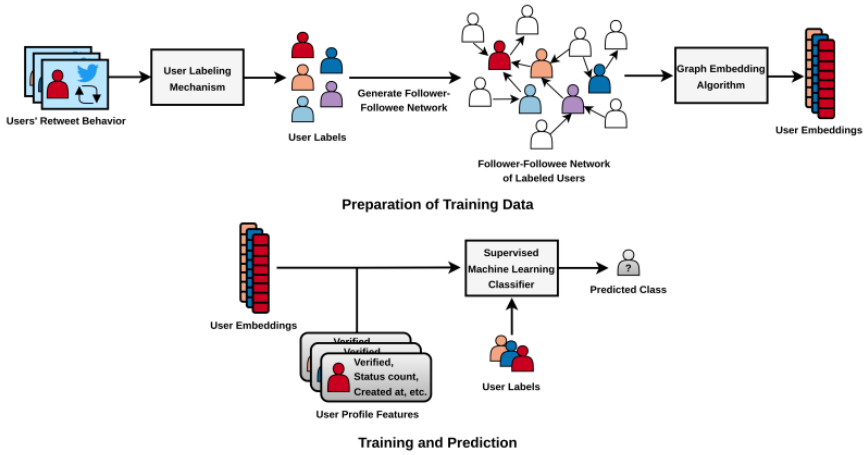
UNIVERSITY OF MINNESOTA
Driven to Discover℠

Machine learning model to mitigate the spread of false information on social networks

A machine learning model to identify and classify misinformation spreaders on social networks.



IP Status: US Patent Pending; Application No. 18/540,131

Applications

- Counteract the spread of false information on social networks
- Identify spreaders of false information

Key Benefits & Differentiators

- **Broadly applicable:** Can be applied to any social network
- **Avoids binary classification:** Behavioral data is used to label users into one of five categories based on user intent and level of informedness
- **Predictive behavioral label :** Network and profile features are used to predict behavioral label for users without any behavioral data

Technology Overview

In recent times, the ease of access to online social networks and the extensive reliance on them for news has increased the dissemination of false information. The spread of misinformation (unintentional) and disinformation (intentional) can have severe impacts on our lives, so they are important to detect along with those that spread the false information. Currently, a combination of AI and human intervention is used to combat the sharing of false information. Despite this, it is still difficult to determine the intent behind those that spread false information on social networks. Typically, accounts deemed to be objectionable are flagged or banned

Technology ID
2022-285

Category
Software & IT/Algorithms
Software & IT/Artificial Intelligence
Software & IT/Data Mining
Software & IT/Simulation & Modeling

View online page

without investigation of the intent of those sharing the false information. By labeling users as only true or false information spreaders, accounts that were deceived by misinformation are unfairly banned.

Researchers at the University of Minnesota have developed a machine learning model that identifies and classifies different types of false information spreaders on social networks. This model detects if a user spreads false information intentionally or unintentionally. By observing users' behavior when exposed to misinformation and its refutation, users can be categorized based on their intent and level of informedness. By labeling users in one of five categories, proper action can be taken to ban accounts that are maliciously sharing disinformation, and educate users that were naively spreading misinformation. This model has been evaluated on a real-world Twitter dataset and shown to be effective in detecting the malicious actors.

Phase of Development

TRL: 5-6
Performance of the working prototype has been tested on real world data.

Desired Partnerships

This technology is now available for:

- License
- Sponsored research
- Co-development

Please contact our office to share your business' needs and learn more.

Researchers

- **Jaideep Srivastava, PhD** Professor, Department of Computer Science and Engineering
- **Euna Mehnaz Khan** Graduate Student, Department of Computer Science and Engineering
- **Ayush Ram** Undergraduate Student, Department of Computer Science and Engineering

References

1. Euna Mehnaz Khan, Ayush Ram, Bhavtosh Rath, Emily Vraga, Jaideep Srivastava(May 1, 2023) , https://doi.org/10.48550/arXiv.2305.00957, https://arxiv.org/abs/2305.00957