CS6140: Machine Learning

CS6140 Final Project - Multi-Agent Social Reinforcement

Assigned: 10/15/2024   *Due: 10/29/2024, 11:59pm, through Canvas*

Chase Coogan - coogan.c@northeastern.edu
Kevin Tang - tang.kevi@northeastern.edu

Section C

This project aims to model and analyze the evolution of opinions within a social network when influenced by false information. We will leverage reinforcement learning to train agents to resemble the general public, malicious falsehood spreaders, real-information spreaders and fact-checkers to understand the role each plays in facilitating opinion dynamics. By simulating these social interactions, the underlying impact of accurate and misleading information can be revealed. This is significant as it addresses critical challenges in today's digital age, where false information can rapidly spread across social media platforms, impacting both public opinion and decision-making processes.

Our motivation stems from applying reinforcement learning to a pressing real-world issue. The social discourse and political polarization from false information could have a detrimental effect on real-world decisions. As a result, we hope our project will provide some insight on how fact-checkers can effectively counter false information and influence opinion dynamics. This in turn will result in better algorithms or policies for combating the spread of false information.

Section D

Opinion dynamics in social networks is essentially the study of how an individual's opinion forms, evolves, and spreads over time within an interconnected network of other individuals. In the digital age, all social media is a form of a social network, which facilitates an exchange of information for opinion making. This information can be classified into one of two categories, truths or falsehoods. Here falsehoods include misinformation and disinformation, which both mislead the public and complicate decision making. To simulate this information exchange, we have decided to use a reinforcement learning approach. This approach will allow us the ability to create various agents, including the general public (who form and evolve opinions based on interactions), malicious falsehood spreaders (who spread misinformation or disinformation), fact-checking agents (who attempt to counteract falsehoods), and real-information agents (who spread factual information). These agents will then interact and adapt to the changing environment, allowing for realistic simulations of how opinions

spread. This will then hopefully reveal existing and hidden concepts in social influence theory like confirmation bias (favoring information that aligns with one's preexisting beliefs) and echo chambers (groups where similar opinions reinforce each other, often excluding contrary views).

Research on opinion dynamics has been done previously and several different models have been created to better understand how these opinions spread and update. One of these models is the DeGroot model, where individuals update their beliefs based on the opinions of their neighbors until a consensus is reached. Similarly, the Hegselmann-Krause model considers agents that only interact with those whose opinions are within a certain range of their own, thereby simulating opinion polarization. Additionally, this topic does not solely stem from computer science and various fields like sociology and psychology have investigated to create a comprehensive model.

Our project, however, is distinct from the others in a few regards. The first, is by using a reinforcement learning model we can have a dynamic and ever evolving system which more closely mimics the interactions in the real world. By leveraging the malicious and benevolent agents, each group will adapt and find strategies to complete their own specific objectives. From here lies the second distinct element which is how we are incorporating both misinformation spreaders and fact-checkers. Most models focus on the spread of information between individuals based on their current opinions, but we wanted to focus on the impact of a third party. By introducing the separate entities in charge of spreading information, we can simulate the effect corporations and social media have on the public. This way we can create algorithms and policies to regulate the spread of false information to the public.

Section E

We will be generating our own data for this experiment as our project will rely heavily on simulations where agents invoke a set of actions and socialize across our network as they try to maximize their Q-values. Q-values are the data that each agent gathers as it performs a set of actions in the environment while it learns what is right or wrong to do based on a reward/penalty assigned to its actions outcome. For instance, as they spread information, the agents individual attributes will be change. For example, as a fake-news agent spreads misinformation and it tricks an agent into believing it its trust-level will increase which will affect the next iteration as it now has more influence across the network. The agent will use this updated Q-value as input to its next iteration cycle to continue its learning process.

In order to carry out this experiment, we plan to create a social network using a graph structure where we will have various agents representing real-information, fake-information, fact-checkers and regular information-consuming agents. Real-information agents are responsible for sharing factual based information. Fake-information agents will do the opposite, spreading misinformation. Fact-checkers will verify both real and fake news that gets sent across the network, and regular information-consuming agents will be a consumer of all types of news presented to it. These agents will be the nodes in our graph, where edges represent a social relationship amongst them. With this type of environment, we can construct graph

networks that allow us to change how many of each type of agent we want introduced to the network to see how a differing size of agents may change the outcome. Agents will "communicate" with each other by passing news around where each agent will have to decide to trust or not trust the source. Agents will be limited to a set number of interactions they can have per iteration per news source.

Agents will be of type 'real-information, fake-information, fact-checker, regular information-consumer' that define what their mission is. Each agent will have its own mission to optimize its Q-values, other than the information-consumers who are utilized solely to spread information across the network. Agents will be rewarded or penalized based on the amount of influence they create across the network based on our objective function. Real-information and fake-information agents are responsible for spreading their news where they will be rewarded for the amount of influence they invoke. Fact-checkers will have the responsibility to check the legimitacy behind the news being spread. A fact-checker will be rewarded for correctly identifying fake-information sources and will be penalized for incorrectly labeling real-information as fake. Fake-information agents will be penalized when fact-checkers catch them. The accruacy of the news being sent around will be rated on a scale of 0-100%, based on truthfullness. For example, a real-news agent will always send a news article above 50%, while a fake-news agent will always send news below 50%.

When information is spread, the recieving agent will store the information passed to them, the agent that sent it, and the source it originated from. They will also store the overall trust level in that information source. Each agent will store its reward and penalty amount to help train them which will be dependent on either the number of agents influenced, or number of information sources correctly debunked in the case of the fact-checker.

In order to evaluate how our agents are performing, we will invoke Q-Learning to help train them properly based on a reward/penalty system. Real-information and fake-information agents will be rewarded based on the number of other agents who recieve their information as well as the total number of agents they convince at the end of each iteration. Additionally, once an agent recieves some news, that agent can then choose to continue passing it to other agents or not. This demonstrates the amount of influence and credibility they are imposing across the network. Penalties will be awarded to the misinformation agent if a fact-checker correctly identifies misinformation being sent across the network that originated from them. The general public agents do not have any reward/penalty, they will only be used to recieve and send information across the network. The update rule for Q-Learning that we will implement is as follows:

general objective function

$$Q_i(s, a) \leftarrow Q_i(s, a) + \gamma \cdot \alpha \left[R(s`, a`) + max\{Q_i(s`, a`)\} \right]$$

$$let \begin{cases} s = \text{current state} \\ a = \text{action taken in state } s \\ r = \text{reward received after taking action } a \\ s' = \text{next state resulting from action } a \\ \alpha = \text{learning rate} \\ \gamma = \text{discount factor (importance of future rewards vs immediate rewards)} \end{cases}$$

We will run an iteration across some δt time, initially 1 year, and track our simulations across time t of 1 day. The state that we will be measuring are the number of consumers who believe in the fake articles vs the number of consumers who believe in the real articles as a ratio.

fake-news objective function

$$R_{\text{fake}}(s, a)^{\delta t} \leftarrow \text{ratio of consumers agreeing with the fake-news agent}$$

real-news objective function

$$R_{\text{real}}(s, a)^{\delta t} \leftarrow \text{ratio of consumers agreeing with the real-news agent}$$

fact-checker objective function

$$R_{\text{fact}}(s, a)^{\delta t} \leftarrow \text{-(num. of mislabeled articles)}$$

Each agent will need to act independently to update its Q-value based on its interactions with other agents. Each agent will also need to hold a set of actions it can take which we can define for each agent. The agent will choose what action to perform based on the epsilon greedy exploration algorithm.

agent actions:

Misinformation agent:
1. Spread misinformation
2. Do nothing

Real-information agent:
1. Spread truthful information
2. Do nothing

Fact-checking agent:
1. Investigate the message/agent and label it real/fake
2. Ignore the message (do nothing)

Regular information-consuming agent:
1. Believe and share a message
2. Do not believe message/do not share

We can use the epsilon-greedy strategy to balance exploration to exploitation. Meaning, we will assign some probability ϵ as the probability the agent chooses a random action (exploration). And, we can assign 1 - ϵ as the probability the agent chooses the action with the highest Q-value. After an agent performs an action they will go through a state change that increases or decreases their influence in the network.

We expect that the fake-news agents to try to learn strategies to act in its favor to spread as much misinformation as possible, and possibly try to avoid being detected by the fact-checkers. We suspect they will try to target agents that have low skepticism. We also epect echo-chambers to form amongst the network where misinformation is being shared continously. We expect the real-news agents to act similarly, where their main goal is to maximize the amount of real information they spread. We expect the real-news agents to focus on areas where fact-checkers are more dominant. We expect fact-checkers to invoke an optimal strategy for uncovering fake-news. We are hoping that fact-checkers are able to effectively differentiate between real and fake news such that it is able to correctly label them. If we run into any issues where failure may occur, we plan to reduce the number of actions that each agent can take, and possible reduce the nubmer of types of agents we use. We will also look into how to optimally reward our agents so that they behave how we have expected them to.

Section F

Chase will work on creating the graph network and set up the infrastructure for each of the agents along with the set of actions they can take and the probabilities of them. Chase will setup the agent attributes such as trust levels, belief models, and setup the initial state of the simulation based on each agent. Chase will work with libraries like NetworkX (or similar) to find unique ways to visualize the data. This is important because our entire simulation is dependent on having a graph network where the agents are able to communicate with one another to pass the data around. Chase will work to make sure the network is setup efficiently, and that agents are setup to correctly store, and transfer data to one another.

Kevin will work to implmement the Q-Learning algorithm and run the simulations. Kevin will be responsible for implementing the update rules for each of the different agents so that their reward/penalty system works efficiently. Kevin will track training performance to make sure the agents are continously learning throughout simulation iterations. This is important because our agents need a sound-proof model so that they train effectively.

Both team-members will collaborate towards helping debug code when necessary and provide

additional necessary support. No one person will handle running all simulations, as we are unsure how computationally heavy it may be. Both team-members will provide equal contribution towards making sure the overall model for the simulation is working correctly.

Section G

1. M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, U. R. Acharya, V. Makarenkov, and S. Nahavandi, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," Information Sciences, vol. 546, pp. 555-588, 2021. Available: https://doi.org/10.1016/j.ins.2019.12.005

2. A. Dosovitskiy, L. Beyer, A. Kolesnikov, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," arXiv preprint arXiv:2010.03050, 2020. [Online]. Available: https://arxiv.org/abs/2010.03050

3. Q. Zhang, W. Zhou, S. Wang, Q. Shen, F. Xu, and J. Sun, "Reinforcement-Learning-Based Dynamic Opinion Maximization in Social Networks," IEEE Transactions on Information Forensics and Security, vol. 17, pp. 290-303, 2022. [Online]. Available: https://ieeexplore.ieee.org/document/9676712

4. K. Dagenais and I. David, "Opinion-Guided Reinforcement Learning," arXiv preprint arXiv:2405.17287, 2024. [Online]. Available: https://arxiv.org/abs/2405.17287