

EX.NO:6

DATE:

Create an attack for tampering with recommender systems.

AIM:

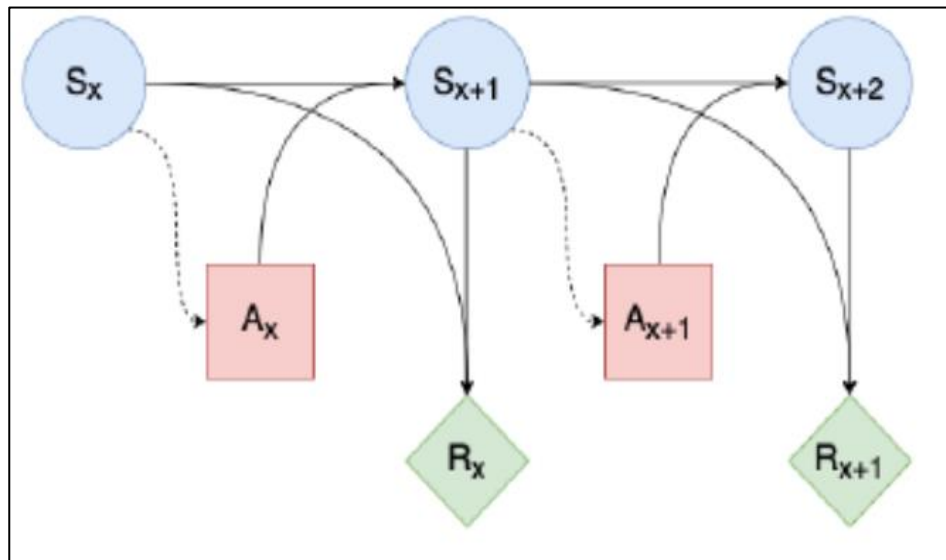
To create an attack for user tampering with recommender systems.

INTRODUCTION:

We first generically frame the media recommendation problem as a Markov Decision Process (MDP), and then use Causal Influence Diagrams (CIDs) to extract the relevant causal dependencies that particular variables exhibit under this model. For some background on CIDs, see Appendix A. We have endeavored to keep the MDP as general as possible, while also incorporating design insights from recent work in implementing RL-based media recommender systems.

AN MDP REPRESENTATION OF MEDIA RECOMMENDATION:

The recommendation problem can simply be described as an agent taking an action a_t at time t , which will transition the system from a current state s_t to a successor state s_{t+1} with probability $T(s_t, a_t, s_{t+1})$.

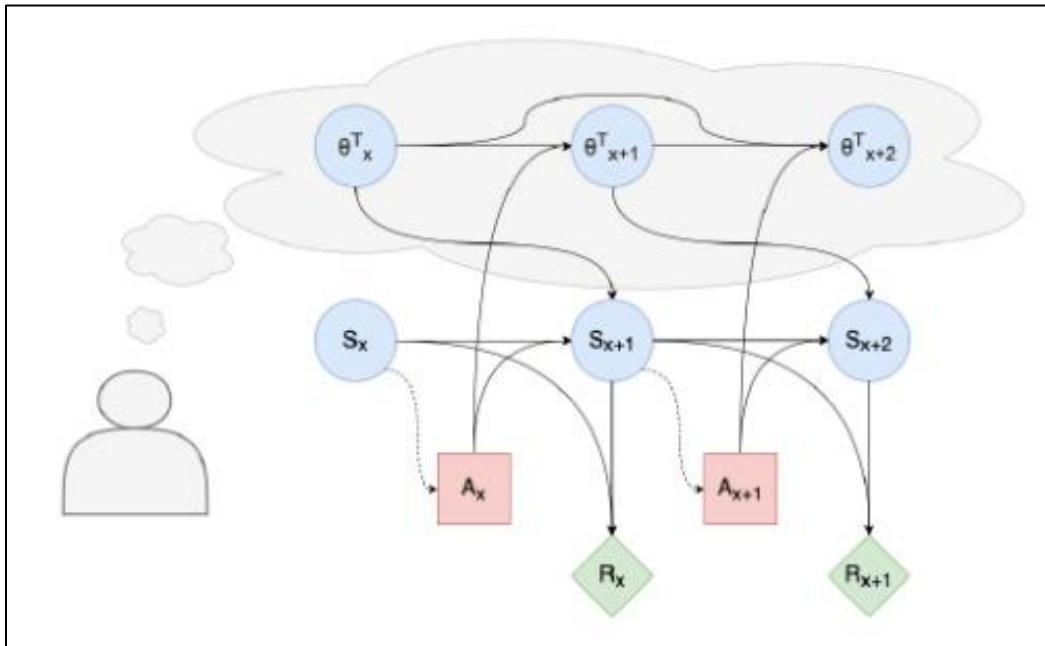


The agent would thereafter be rewarded with the value $R(s_t)$, and then another action would be chosen at time $t+1$, and so forth

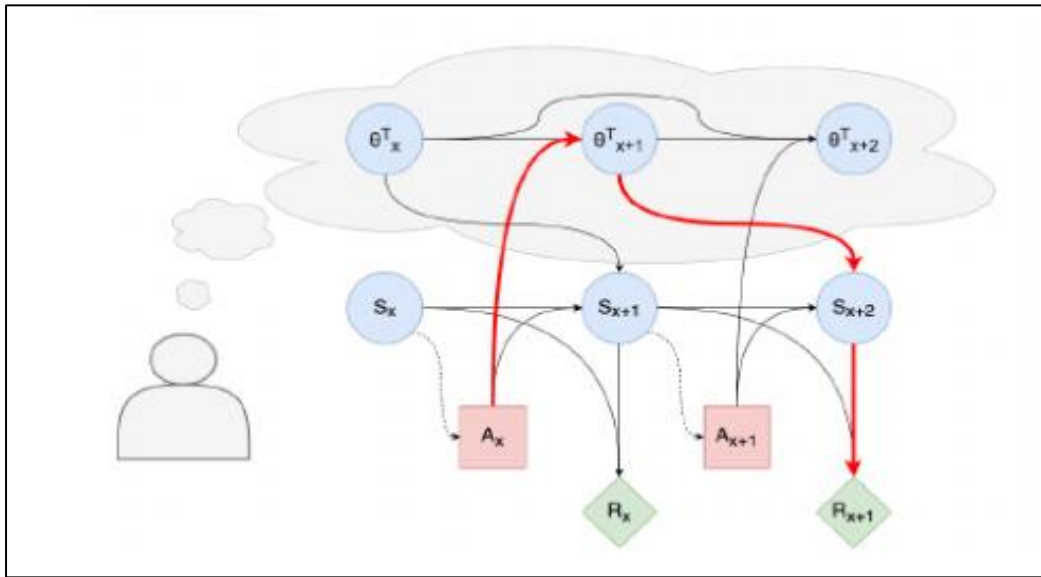
EXTRACTING A CID FROM THE MDP:

A simple thought experiment can demonstrate that this, CID underspecifies the causal relationships in the actual problem by leaving key variables external to the MDP unacknowledged. Consider the following: Alice and Bob are two university students who have just created accounts on some media platform, who have so far both been recommended the same three articles about the student politics at their university, and who have both clicked on all three articles. Within our general definitions, it is quite plausible that the states of the system have been identical thus far from the agent's perspective. However, what if Bob is uninterested in politics and is just clicking on the articles because his friends feature prominently in the cover photos of all three, whereas Alice is clicking out Netherlands of a genuinely strong interest in politics, including student politics? If the recommendation to both Alice and Bob at the next time-step— say, A_{x+1} — is an article about federal politics, it is intuitively untrue that the distribution over possible states at S_{x+1} is the same; Alice is surely more likely to observably engage with this content. Evidently, a random variable exogenous to the MDP must be introduced to properly model the causal properties of the true system. Informally, we argue that this variable can be characterized as the preferences/opinions/interests of the specific user to which the agent is recommending media.

If we introduce the exogenous variable to the system, without changing any other definitions, we arrive at the CID. This CID, we argue, more completely captures the actual causal dynamics of the Media Recommendation MDP. We note that previous literature has acknowledged a similar causal structure to the recommendation process [12]; however, this was not formulated in the CID framework that we have used, which permits sophisticated graphical analysis of the kind developed in the next section.



We use the CID formulated in the previous section to analyze the safety of the RL-based approach to media recommendation, specifically with respect to the high-level concerns of user manipulation and polarization. After introducing the phenomena of ‘instrumental control incentives’ and ‘instrumental goals’ from the RL incentive analysis literature, we show that in the CID, an instrumental goal exists for the agent to manipulate the expected value of the exogenous variable θT . This lends a concrete, formal interpretation to the (formerly only hypothesized) safety issue that we have called ‘user tampering’

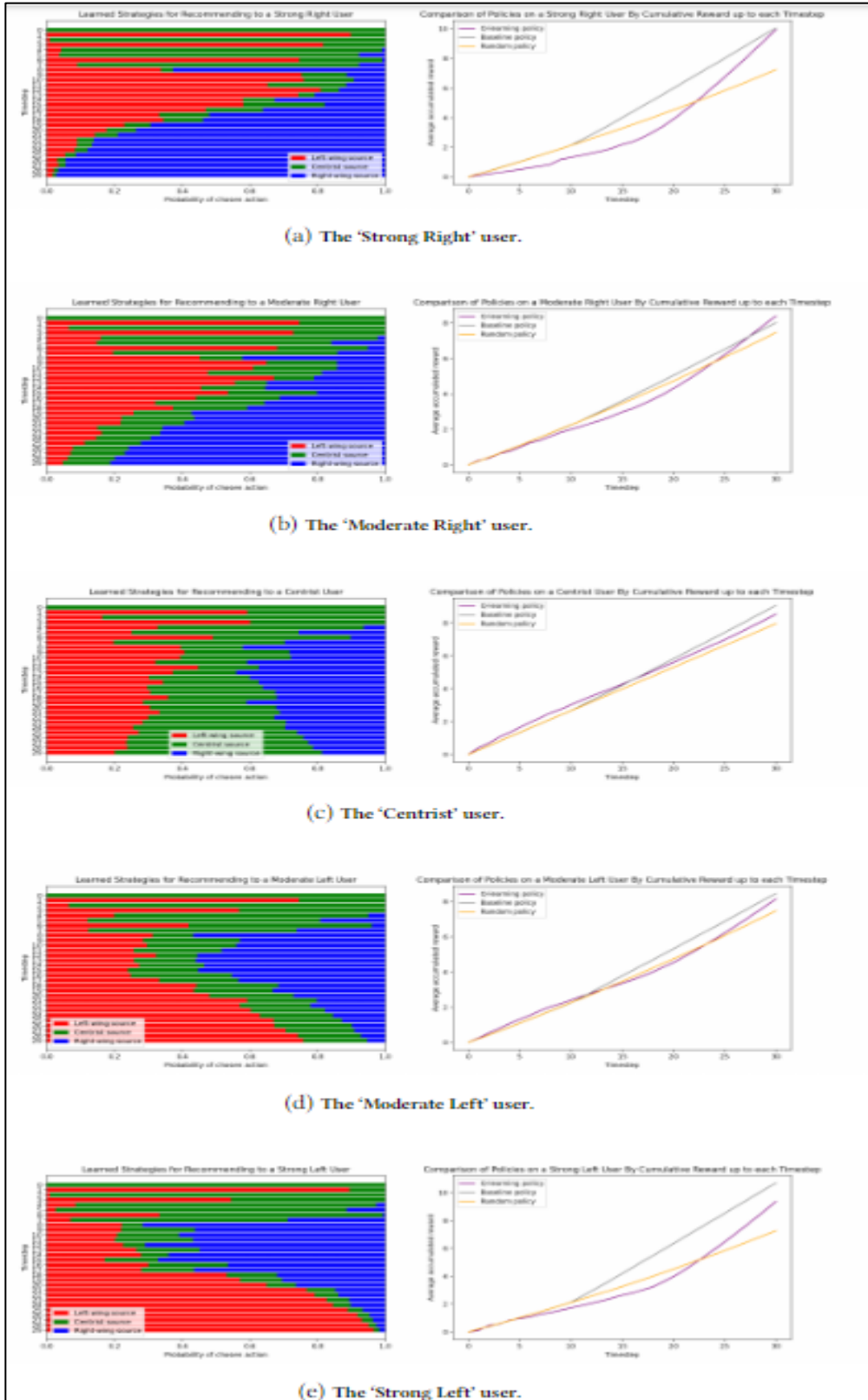


We empirically analyze the user tampering phenomenon formalised in the previous section. Firstly, we introduce a simple abstraction of the media recommendation problem, which involves simulated users and a user tampering incentive inspired by recent empirical results about polarisation on social media. Then, we present a Q-learning agent intended to mimic the Deep Q-learning algorithms used in recent media recommendation research, and train it in this environment. We show that its learned policy clearly exploits user tampering in pursuit of greater rewards.

RECOMMENDER SIMULATION:

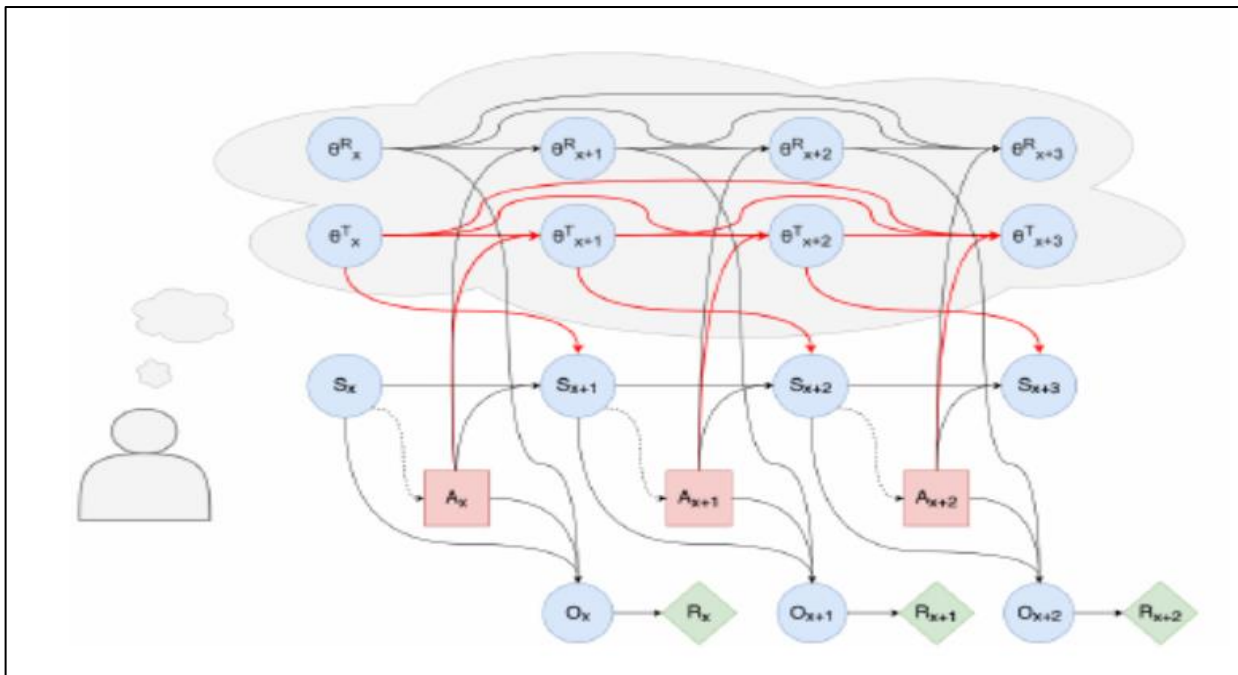
This contained:

- A 'strong left' user with $\theta T_0 = (0.4, 0.1, 0.1)$
- A 'moderate left' user with $\theta T_0 = (0.3, 0.25, 0.1)$
- A 'centrist' user with $\theta T_0 = (0.2, 0.4, 0.2)$
- A 'moderate right' user with $\theta T_0 = (0.1, 0.25, 0.3)$
- A 'strong right' user with $\theta T_0 = (0.1, 0.1, 0.4)$



CONCLUSION:

The risks of emergent RL-based recommender systems with respect to user manipulation and polarization. We have formalized these concerns as a causal property – “user tampering” – that can be isolated and identified within a recommendation algorithm, and shown that by designing an RL-based recommender which can account for the temporal nature of the recommendation problem, user tampering also necessarily becomes learnable. Moreover, we have shown that in a simple simulation environment inspired by recent polarisation research, a Q-Learning-based recommendation algorithm consistently learned a policy of exploiting user tampering – which, in this context, took the form of the algorithm explicitly polarising our simulated ‘users’. This is obviously highly unethical, and the possibility of a similar policy emerging in real-world applications is a troubling take away from our findings. Due to a combination of technical and pragmatic limitations on what could be done differently in RL-based recommender design, it is unlikely that commercially viable and safe recommenders based entirely on RL can be achieved, and this should be borne in mind when selecting future directions for advancement in media recommendation research & development.



RESULT:

Thus the creating an attack for user tampering with recommender systems was successfully completed.