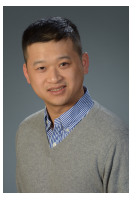
SaTC: CORE: Small: Collaborative: Learning Dynamic and Robust Defenses Against Co-Adaptive Spammers

Sihong Xie (Lehigh University)
Philip S. Yu (University of Illinois at Chicago)

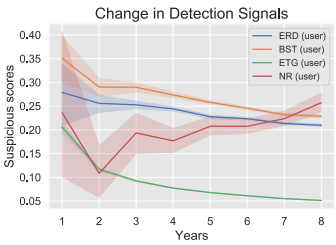


http://www.cse.lehigh.edu/~sxie/satc_19.html

- Fake online reviews swaying user opinions on many platforms.
- Intelligent spammers vs. uninformed users.
- Spammers can learn to adapt and attack static defenses.

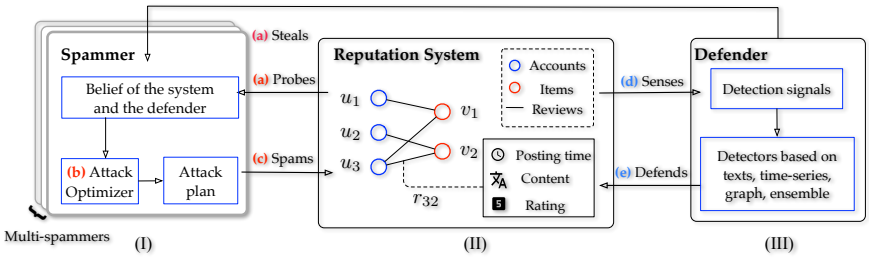
Technical Challenges

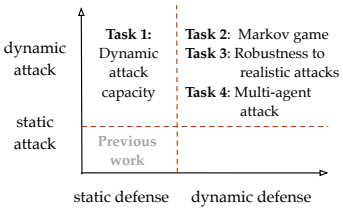
- Dynamic and less predictable spamming activities.
- Detection on multi-modal data.
- Long-term defense and security cost and benefits?
- Real spammers' goals and modi operandi?
- Multiple diverse spamming strategies.



On a Yelp review dataset, various spam detection signals undergo temporal shifting due to changing spamming footprints.

Overview of Research Challenges and Tasks

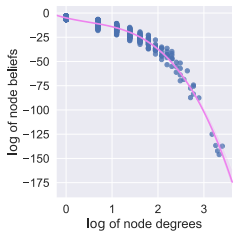


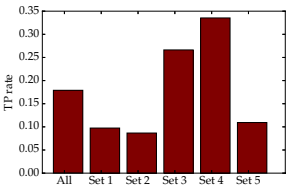


Research Tasks

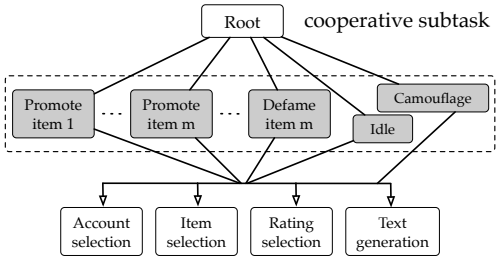
- Design dynamic attacking algorithms on multi-modal data.
- Deep reinforcement learning aiming to find defense policy for long-term security benefits.
- Inverse RL to mine real spammers' goals and modi operandi.

- Hierarchical RL to defend against diverser dynamic spamming strategies.





True Positive detection rates under different sets of attacking strategies.



Proposed Hierarchical Reinforcement Learning for modeling cooperative spamming strategies.

Scientific and Broader Impacts

- ML security on multi-modal, discrete, and structural domains.
- Large-scale deep, inverse, hierarchical RL algorithms.
- Gamified software to educate end-users about spam spotting.