A Testbed to Evaluate the Robustness of Reputation Systems in E-Marketplaces

(Demonstration)

Athirai A. Irissappane Jie Zhang
School of Computer Engineering, Nanyang Technological University, Singapore
{athirai001, zhangj}@ntu.edu.sg

ABSTRACT

Existing testbeds to evaluate reputation systems are mainly simulation based and are not flexible to perform robustness evaluations against unfair rating attacks. In this paper, we present a novel comprehensive testbed, which can evaluate reputation systems using both simulations and real data. The testbed incorporates sophisticated deception models and unfair rating attack models, and introduces several performance metrics to fully test and compare the effectiveness and robustness of different reputation systems.

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence - Intelligent Agents, Multiagent Systems

Keywords

Testbed; Reputation System; Unfair Ratings; Robustness; Electronic Marketplaces

1. INTRODUCTION AND MOTIVATION

Reputation systems strengthen the quality of electronic marketplaces by providing incentives for good behavior of sellers and their high quality services and by sanctioning their bad behavior and low quality services. However, the performance of reputation systems may be affected by unfair rating attacks (misleading opinions) from dishonest buyers (also called advisors), to promote or demote a seller.

Several reputation systems [1] (BRS, TRAVOS, etc.) have been proposed to detect unfair ratings and to assist buyers in accurately evaluating the reputation of sellers. However, majority of them have only been evaluated using experimental frameworks of their authors' own and compared with only a very few other approaches. Also, most of the experimental frameworks are based on simple simulated scenarios, which cannot be considered as reliable evidence for how the reputation systems would perform in realistic environments.

Some unified testbeds [2] (ART and TREET) have been proposed to evaluate reputation systems. However, they are mainly simulation-based and cannot reflect real environmental settings. Besides, they are not specifically designed for evaluating the robustness of reputation systems in coping with unfair rating attacks. Also, the testbeds often propose only a single performance metric to evaluate trust models.

Appears in: *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*

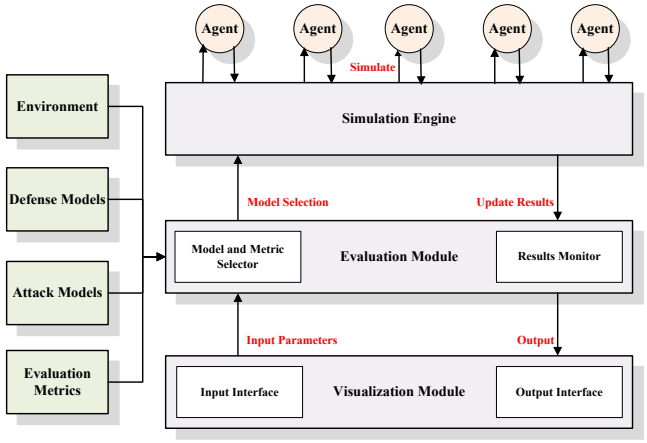

Figure 1: High level architecture of the testbed

We have developed a comprehensive testbed to evaluate and compare reputation systems, to fully analyze their effectiveness and robustness in coping with unfair rating attacks. The unique features of the testbed are: 1) it provides a variety of testing environments using simulations as well as real data, with the flexibility to choose the different environmental settings and experimental parameters; 2) it can evaluate the robustness of reputation systems in handling unfair ratings; 3) it is a comprehensive unified platform, supporting a variety of reputation systems (centralized/decentralized, single/multi-criteria) and attack models (simple as well as complex attacks like collusion attacks); 4) the testbed easily allows to add new reputation systems and attack models; 5) it allows to visually compare the performance of different reputation systems in a variety of experimental settings.

2. TESTBED ARCHITECTURE

Fig. 1 shows the high level architecture of the testbed. The major components of the testbed are described below.

Environment The environment is the most important component of the testbed, representing an e-market scenario for evaluating the reputation systems. It is cumbersome to obtain ground truth about which ratings are unfair for data extracted from real-world sites because: 1) it is costly for system managers to find out the ground truth of the data by hiring human subjects to inspect every rating; 2) system managers with such information may not be willing to

share it. Hence, in the testbed, we use three ways to generate environments for experimentation (more details of which can be found in [3]): 1) simulated environments: environments are entirely based on simulations with known ground truth about which ratings are unfair; 2) real environments: data is collected from real environments (*Dianping.com*, *TripAdvisor.com*, etc.) and the performance of reputation systems is compared by their accuracy in predicting the outcome of a transaction with a seller at time t based on all ratings before time t; 3) real environments with simulated attacks: a variety of unfair rating attacks are simulated on real data and the performance of reputation systems is compared based on their ability to filter out the unfair ratings.

Defense Models They represent reputation systems with their unfair rating detection approaches. We have implemented BRS, TRAVOS, iCLUB, BLADE, WMA, ProbCog, Personalized, eBay's reputation system and some multi-criteria defence schemes, which evaluate a seller on multiple criteria.

Attack Models They represent a variety of unfair rating attack scenarios [4] from simple attacks to sophisticated ones such as camouflage, whitewashing and sybil.

Evaluation Metrics Mean Absolute Error (MAE) in estimating sellers' reputation, Matthew's Correlation Coefficient (MCC), precision, recall and F-score are the conventional metrics used in the testbed. Besides, we have introduced the robustness evaluation metric (which is the transaction volume difference between honest and dishonest sellers in the e-marketplace), to specifically evaluate the robustness of reputation systems against unfair ratings.

Visualization Module The input interface provides a convenient way to configure different environmental settings and set up customized experiments. Visual comparison and analysis of results is supported by the output interface.

Evaluation Module The model and metric selector selects the appropriate environment, defence, attack models and evaluation metrics based on the user input and communicates it to the simulation engine which will perform the actual simulation of the multi-agent system. The results monitor updates the global results on the evaluation metrics using the results from the simulation.

Simulation Engine It simulates the e-marketplace with different agent (buyer, seller) behavior for a given time period. In the simulation, the selected defense model predicts agents' behavior for a transaction and the evaluation metrics are updated based on the predicted values.

The testbed (snapshots are shown in Fig. 2) is implemented in JAVA. The source code of the testbed is available at http://reputation-testbed.googlecode.com/svn/trunk/.

3. DEMO

In the demo, we will show how the testbed can be used to compare the reputation systems under various attacking scenarios using simulated as well as real data. The result analysis using graphical displays will also be demonstrated. Other special features of the testbed, such as to easily add defense models and import saved environmental settings will also be shown. A movie showing the testbed functionalities can be found at http://youtu.be/tLKpwAkhsz8.

4. CONCLUSION AND FUTURE WORK

We have developed a comprehensive testbed to evaluate and compare the performance of reputation systems. The testbed supports a variety of environmental settings and can perform thorough evaluation of the robustness of reputation systems against unfair rating attacks. The testbed will be highly beneficial for researchers in the field to analyze and compare the performance of their approaches in different experimental settings. Our future work will be to extend the testbed to adapt to domains other than e-commerce.

Figure 2: Snapshots of the testbed

5. ACKNOWLEDGMENTS

This work is supported by MoE AcRF Tier 2 Grant M4020110.020 and the Institute for Media Innovation at Nanyang Technological University.

6. REFERENCES

[1] Irissappane, A.A., Jiang, S., Zhang, J.: A framework to choose trust models for different e-marketplace environments. In IJCAI'13.

[2] Kerr, R., Cohen, R.: TREET: The trust and reputation experimentation and evaluation testbed. Electronic Commerce Research, 10(3):271–290, 2010.

[3] Irissappane, A.A., Jiang, S., Zhang, J.: Towards a comprehensive testbed to evaluate the robustness of reputation systems against unfair rating attacks. In UMAP TRUM workshop, 2012.

[4] Jiang, S., Zhang, J., Ong, Y.: An evolutionary model for constructing robust trust networks. In AAMAS'13.