Towards the Design of Robust Trust and Reputation Systems

Siwei Jiang
School of Computer Engineering
Nanyang Technological University, Singapore
sjiang1@ntu.edu.sg

1 Research Problem

Our research is within the area of artificial intelligence and multiagent system. More specifically, we are interested in addressing robustness problems in trust and reputation systems so that buying agents are able to accurately model the reputation of selling agents even with the existence of various unfair rating attacks from other dishonest buyers (called advisors) [Zhang and Cohen, 2008].

In multiagent-based e-marketplaces, trust and reputation systems are designed for buyers to model seller reputation based on ratings shared by advisors. However, unfair rating attacks from dishonest advisors render trust and reputation systems ineffective to mislead buyers to transact with dishonest sellers [Jøsang, 2012]. Typical unfair rating attacks include **Constant** where dishonest advisors constantly provide unfairly positive/negative ratings to sellers; **Camouflage** where dishonest advisors camouflage themselves as honest advisors by providing fair ratings to build up their trustworthiness first and then gives unfair ratings; **Whitewashing** where a dishonest advisor is able to *whitewash* its low trustworthiness by starting a new account with the initial trustworthiness value; **Sybil** where a dishonest buyer creates several accounts to constantly provide unfair ratings to sellers; **Sybil Camouflage** (the combination of Sybil and Camouflage) where a number of dishonest advisors perform Camouflage attacks together; and **Sybil Whitewashing** where a number of dishonest advisors perform Whitewashing attacks together. The challenge is thus how to design robust trust models to against various strategic attacks.

Various trust models [Zhang and Cohen, 2008; Jøsang, 2012] have been proposed to cope with unfair ratings. However, these models are not completely robust against various attacks. In particular, when dishonest advisors occupy a large proportion in e-marketplaces (i.e., Sybil), BRS becomes inefficient and iCLUB is unstable because they both employ the "majority-rule". When dishonest advisors adopt strategic attacks, TRAVOS does not work well because it assumes an advisors' rating behavior is consistent. ReferralChain assigns trust value 1 to every new buyer (advisor) which provides a chance for dishonest advisors to abuse the initial trust (i.e., Whitewashing). Personalized is vulnerable when buyers have insufficient experience with advisors and the majority of advisors are dishonest (i.e., combination of Whitewashing and Sybil). Thus, we need more robust trust models.

2 Progress to Date

The research progress includes two major parts. At first, we combine different existing trust models to cope with unfair rating attacks [Zhang *et al.*, 2012]. Secondly, a multiagent evolutionary trust model (MET) is proposed for buyers to construct robust trust networks [Jiang *et al.*, 2013].

2.1 Combination of Trust Models

The existing trust models can be generally classified into two major categories: *Filtering-based* (e.g., BRS and iCLUB) and *Discounting-based* (e.g., TRAVOS, ReferralChain and Personalized). Under comprehensive experimental evaluation, none of the single trust model is completely robust against the six typical unfair rating attacks [Zhang *et al.*, 2012]. However, different trust models show their unique characteristics for various testing scenarios due to their special designs.

We begin with a novel insight that combining the advantages of existing trust models is able to cope with various unfair rating attacks. Two feasible methods are proposed.

Approach 1—Filter-then-Discount:

1. Discard dishonest advisors by a Filtering-based model.
2. Use a Discounting-based model to aggregate discounted advisor ratings and calculate seller reputation.

Approach 2—Discount-then-Filter:

1. Calculate each advisor A's trustworthiness $T(A)$ by a Discounting-based model.
2. If $T(A) < threshold$, remove A's all ratings.
3. Calculate seller reputation by a Filtering-based model.

To verify effectiveness of different trust models, we design a multiagent-based e-marketplace testbed [Zhang *et al.*, 2012; Jiang *et al.*, 2013]. The robustness of a trust model (defense, Def) against an attack model (Atk) is defined as follows:

$$\mathcal{R}(Def, Atk) = \frac{|Tran(S^H)| - |Tran(S^D)|}{|B^H| \times Days \times Ratio} \quad (1)$$

where $|Tran(S^D)|$ and $|Tran(S^H)|$ are transaction volumes of the dishonest and honest duopoly sellers, respectively. $\mathcal{R}(Def, Atk) = 1$ or -1 means Def is *complete robust* or *complete vulnerable* to Atk, respectively. The larger value indicates the trust model is more robust against the attack.

From results in [Zhang *et al.*, 2012], the robustness of single trust models can be enhanced by combining different categorial models, and Discount-then-Filter is most robust than Filter-then-Discount against typical unfair rating attacks.

2.2 Construction of Robust Trust Network

To date, most trust models are designed based on statistical techniques. In [Jiang *et al.*, 2013], we propose a novel multiagent evolutionary trust model (MET) where each buyer constructs its trust network (information about which advisors should be include in the network and their trustworthiness) by the evolutionary model.

Assume that in an e-marketplace, the set of buyers is denoted as $B = \{B_i | i = 1, \cdots, l\}$ and the set of sellers is denoted as $S = \{S_j | j = 1, \cdots, m\}$. We also denote the trustworthiness of an advisor $A_k \in B$ from the view of a buyer B_i as $T_{B_i}(A_k) \in [0, 1]$. In the buyer B_i's trust network TN_{B_i}, the trustworthiness values of advisors connected with B_i is then denoted as $T_{B_i}(A) = \{T_{B_i}(A_k) | A_k \in TN_{B_i}\}$.

In MET, a fitness function is designed for buyers to measure the quality of their trust networks. Formally, the fitness value of buyer B_i's trust network $T_{B_i}(A)$ is calculated as:

$$f(T_{B_i}(A)) = \frac{1}{m'} \sum_{j=1}^{m'} |R_{B_i}(S_j) - \tilde{R}_{B_i}(S_j)| \qquad (2)$$

where $R_{B_i}(S_j)$ and $\tilde{R}_{B_i}(S_j)$ are reputation of a seller S_j based on B_i's personal experience and advisors' ratings from $T_{B_i}(A)$, respectively. In addition, $m' \leq m$, indicating that sellers with which either buyer B_i or its advisors have no experience will not be considered in the fitness evaluation.

A smaller fitness value indicates that the buyer's trust network is in higher quality, because the combination of advisors' ratings is more similar to the buyer's own opinions regarding common sellers. In other words, the fitness function measures the suitability of selected advisors and the accuracy of the trust values assigned to these advisors simultaneously.

In each generation, buyer B_i choose advisor A_r to take interaction only when the following condition is satisfied:

$$\begin{aligned} &(\text{diff}(T_{B_i}(A), T_{A_r}(A)) - 0.5) \qquad (3) \\ &\quad \times (\text{diff}(f(T_{B_i}(A)), f(T_{A_r}(A))) - 0.5) > 0 \end{aligned}$$

After choosing three[1] advisors by trust network comparison, the buyer will generate a candidate trust network using evolutionary operators [Jiang *et al.*, 2012]. Two widely used evolutionary operators (DE crossover and polynomial mutation) are adopted to produce candidate trust works in MET. By comparing the candidate trust network with the buyer's own trust network, the one with higher fitness value measured by Eq. 2 will survive to the next generation.

From Table 1, experimental results show that iCLUB and the Personalized approach have large perturbation under Sybil attacks. BRS, TRAVOS and ReferralChain are vulnerable to Sybil, Camouflage and Whitewashing, respectively. It demonstrates that MET is more robust than these other trust models against typical unfair rating attacks.

[1]If insufficient advisors satisfy Eq. 3, some advisors will be randomly selected from the buyer's trust network.

Table 1: Robustness of Trust Models vs. Attacks

	Constant	Camouflage	Whitewashing
BRS	0.87±0.03	0.89±0.02	-0.18±0.07
iCLUB	0.98±0.02	**0.99±0.02**	0.77±0.13
TRAVOS	0.97±0.02	0.82±0.03	0.87±0.03
ReferralChain	0.89±0.04	0.69±0.04	-0.95±0.08
Personalized	**0.99±0.03**	**0.99±0.03**	**0.98±0.03**
MET	0.98±0.02	**0.99±0.02**	**0.98±0.04**
	Sybil	Sybil Cam*	Sybil WW*
BRS	-0.99±0.08	-0.47±0.07	-0.30±0.07
iCLUB	0.23±0.35	0.90±0.09	0.20±0.29
TRAVOS	0.16±0.09	-0.57±0.07	-0.98±0.07
ReferralChain	0.82±0.06	0.63±0.08	-0.98±0.07
Personalized	0.74±0.45	**0.94±0.08**	-1.00±0.08
MET	**0.87±0.15**	**0.94±0.06**	**0.82±0.11**

*Sybil Cam: Sybil Camouflage; Sybil WW: Sybil Whitewashing

3 Future Research

In future, we will investigate the impact of various parameters and components on our trust models, i.e., the *threshold* in Discount-then-Filter, different types of evolutionary operators in MET, etc. In our MET model, the trust network information for buyers is encoded as the fixed length of trust values for advisors. We plan to test different encoding methods, such as varying length of trust networks for different buyers and graph encoding to represent buyers' trust networks.

We also plan to build a comprehensive testbed to evaluate the robustness of trust models based on simulated and real e-commerce environments. In the current work, we only investigated the typical unfair rating attacks from advisors. In future, we will incorporate more intelligent attacks from advisors. For instance, dishonest advisors firstly adopt learning techniques to analyze a buyer's rating pattern, and then inject a small number of unfair ratings into e-marketplaces to mislead the buyer to transact with dishonest sellers. We will also combine sellers' cheating behaviors with advisors' unfair ratings, and evaluate their impact on different trust models.

References

[Jiang *et al.*, 2012] S. Jiang, J. Zhang, and Y.S. Ong. A multiagent evolutionary framework based on trust for multiobjective optimization. In *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2012.

[Jiang *et al.*, 2013] S. Jiang, J. Zhang, and Y.S. Ong. An evolutionary model for constructing robust trust networks. In *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2013.

[Jøsang, 2012] A. Jøsang. Robustness of trust and reputation systems: Does it matter? In *Proceedings of the 6th IFIP International Conference on Trust Management (IFIPTM)*, pages 253–262, 2012.

[Zhang and Cohen, 2008] J. Zhang and R. Cohen. Evaluating the trustworthiness of advice about seller agents in e-marketplaces: A personalized approach. *Electronic Commerce Research and Applications*, 7(3):330–340, 2008.

[Zhang *et al.*, 2012] L. Zhang, S. Jiang, J. Zhang, and W. Ng. Robustness of trust models and combinations for handling unfair ratings. In *Proceedings of the 6th IFIP International Conference on Trust Management (IFIPTM)*, 2012.