A Framework to Choose Trust Models for Different E-Marketplace Environments

Athirai A. Irissappane, Siwei Jiang, Jie Zhang
School of Computer Engineering
Nanyang Technological University, Singapore
{athirai001, sjiang1, zhangj}@ntu.edu.sg

Abstract

Many trust models have been proposed to evaluate seller trustworthiness in multiagent e-marketplaces. Their performance varies highly depending on environments where they are applied. However, it is challenging to choose suitable models for environments where ground truth about seller trustworthiness is unknown (called *unknown environments*). We propose a novel framework to choose suitable trust models for unknown environments, based on the intuition that if a model performs well in one environment, it will do so in another similar environment. Specifically, for an unknown environment, we identify a similar simulated environment (with known ground truth) where the trust model performing the best will be chosen as the suitable solution. Evaluation results confirm the effectiveness of our framework in choosing suitable trust models for different environments.

1 Introduction

In multiagent e-marketplaces, self-interested selling agents may act maliciously by not delivering products with the same quality as promised. It is thus important for buying agents to reason about the quality (trustworthiness) of sellers and determine which sellers to do business with. However, in such open and large environments, buyers often encounter sellers with which they have no previous experience. In this case, the buyers often obtain advice (i.e. ratings) about the sellers from other buyers (called *advisors*). However, some advisors may also be dishonest and provide unfair ratings, to promote some sellers or bad-mouth others.

Many trust models [Sabater and Sierra, 2005; Jøsang *et al.*, 2007] have been proposed to assess seller trustworthiness, some of which, such as BLADE [Regan *et al.*, 2006], also address the unfair rating problem. However, their performance is often highly affected by the environments where they are applied. Specifically, Fullam and Barber [2007] found out that the accuracy of trust models is influenced by environmental settings such as the frequency of transactions, the honesty of sellers, and the accuracy of advisors' ratings. In addition, almost all trust models rely on certain tuning parameters which may significantly affect their performance.

Further, most trust models have only been evaluated in simulated e-market environments, where ground truth about agents' malicious behavior is known upfront, such as whether sellers deliver products with lower quality than what they promised and whether advisors provide unfair ratings. In simulated environments, the performance of trust models with specific parameter values can be evaluated, and the best models can then be easily chosen. However, for a real e-market environment, it is difficult to obtain ground truth because it is expensive or time consuming to manually inspect every transaction. Thus, choosing suitable trust models for unknown environments is challenging and not well addressed, but important for practical applications.

In this paper, we propose a novel framework to choose suitable trust models for unknown e-market environments. The intuition is that if a model performs well in one environment, it will also perform well in another similar environment. More specifically, we first find out the best models with their best parameter settings in a set of simulated environments. Then for an unknown environment, we identify the most similar simulated environment by calculating the similarity between each simulated environment and the unknown environment based on a set of carefully selected features. The model performing the best in the identified environment will be chosen as the one for the unknown environment. Experimental results show that with a very high probability, our framework can choose the most suitable trust models to evaluate seller trustworthiness for different unknown environments. Seller trustworthiness evaluated using trust models chosen by our framework in a set of different e-market environments is more accurate than applying any specific trust model with its best parameter values in those environments.

2 Related Work

Some studies, e.g. [Wang and Singh, 2010], use data from real-world e-commerce systems (e.g. eBay.com and Amazon.com) to evaluate the performance of trust models by their accuracy of predicting ratings of given transactions. However, the ground truth about whether the ratings of those transactions are unfair may be unknown. One may argue that we can rely on buyers themselves to choose trust models because they know their true experience with sellers, but it is costly for buyers to try each trust model with various parameters. Some other studies, e.g. [Hang *et al.*, 2009], make use of explicitly indicated trust relationships by users in some real-world systems to evaluate trust models. However, users may lie about their trust relationships.

Closely related to our work is the Personalized Trust Framework (PTF) [Huynh, 2009] that selects an appropriate trust model for a particular environment based on users' own choice. Here, users can specify how to select a trust model based on information about whose trustworthiness is to be evaluated and the configuration of trust models. PTF relies entirely on human intervention, but it is impossible for human users to figure out which models will perform the best in complex e-market environments. Our framework provides an automated approach to address this issue.

The idea of our framework bears similarity to the underlying principle of Case-Based Reasoning (CBR) [Sormo *et al.*, 2005] which is the process of solving new problems based on the solutions of similar past problems. The major challenge in CBR resides in the retrieval of existing cases that are sufficiently similar to a new problem. In contrast, in our framework, e-market environments with ground truth (existing cases) may not exist and we have to create them by simulations. In addition, in our framework, the features used to represent e-market environments are not known beforehand. We thus have to come up with an exhaustive list of potential features and carefully select the most relevant ones for the framework to measure the similarity of environments.

3 Our Framework

Fig. 1 illustrates the procedural design of the framework. We first simulate a large set of e-market environments with the ground truth about the honesty of agents' behavior. Given a set of available trust models with specific values of their parameters (referred to as *candidate trust models*), we evaluate their performance in each simulated environment where the best model is identified for each environment and forms a *best environment-model pair*. For each environment, the framework then extracts a set of carefully selected (most relevant) features, based on which we can calculate the similarity between the unknown environment and each simulated environment. Finally, the trust model, which performs the best in the most similar simulated environment, is chosen as the most suitable trust model for the unknown environment. The major components and the detailed procedures will be described in the following subsections.

E-Market Environments An e-market environment mainly consists of a set of sellers, buyers, transactions (each of which is between a seller and a buyer with a certain monetary value) and ratings (each of which is given by a buyer to a seller at specific time indicating whether the buyer is satisfied with the transaction). A rating can be binary (e.g. 0 or 1), multinominal (e.g. 1 - 5) or real (e.g. in the range of $[0, 1]$). A dishonest seller may advertise its products having high quality but actually deliver low quality ones or not deliver at all. A dishonest buyer may lie about its satisfaction level of a transaction by providing an unfair rating.

There are two types of environments in our framework. One is those where the ground truth about seller and buyer deception is known (called *known environments*). The ground truth may consist of information about whether a seller is deceptive in a transaction or whether a buyer gives an unfair rating. However, in reality, it is difficult to obtain such ground truth. Hence, we mainly depend on simulations to artificially generate these known environments (called *simulated environments* in Fig. 1). In the framework, we will simulate a large number of environments, to cover as many scenarios as possible and closely depict real-world environments. For example, we may simulate an environment with many sellers but fewer buyers or with many buyers but fewer sellers. We may simulate a very sparse environment with few ratings provided by buyers, and a very dense environment where each seller is flooded with a large number of ratings. We may also simulate different scenarios where buyers are active or inactive in providing ratings. In these environments, we also simulate sellers with different levels of honesty in delivering products, and buyers with different unfair rating attacking scenarios, including for example, unfair ratings to only reputable or disreputable sellers, a lot or few unfair ratings, unfair ratings given in a short or long time period, etc. Note that ratings in simulated environments are of the real type for being easily mapped to other types.

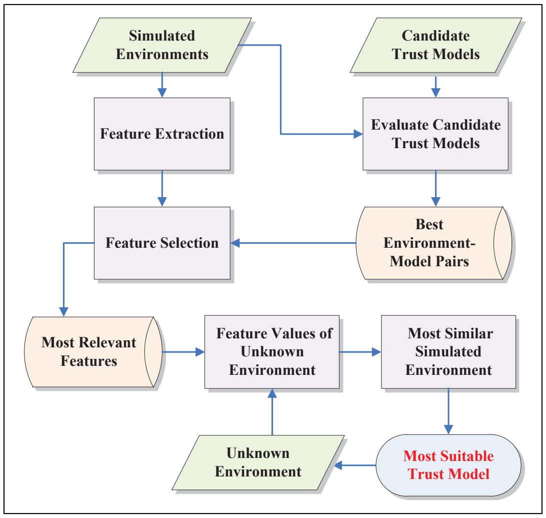

Simulated Environments
Candidate Trust Models
Feature Extraction
Evaluate Candidate Trust Models
Feature Selection
Best Environment-Model Pairs
Most Relevant Features
Feature Values of Unknown Environment
Most Similar Simulated Environment
Unknown Environment
Most Suitable Trust Model

Figure 1: Procedural Design of the Framework

Another type of environments is those where the ground truth about seller and buyer deception is unknown, referred to as *unknown environments* in Fig. 1. We will choose the most suitable trust models for the unknown environments.

Candidate Trust Models Many trust models have been proposed to evaluate seller trustworthiness in e-marketplaces. New trust models will also likely be proposed in the future. All these trust models can be considered as candidate trust models in our framework. In addition, most of them have some parameters to tune, which may result in different performance. Thus, a candidate trust model is defined as a trust model with a specific value for each of its parameters. For a parameter varying in a range, we divide its range into a number of equal intervals and randomly choose a value in each interval. Ideally, the larger number of intervals is better.

Best Environment-Model Pairs Given a set of candidate trust models and a set of simulated (known) environments, we find out in each environment, which candidate model performs the best. The results are a set of best environment-model pairs. If several models perform equally the best in an environment, we keep all of them in the best environment-model pair. The way of identifying the best candidate model for a specific environment is to evaluate the average performance of each candidate model in the environment and choose the one that achieves the highest performance.

Feature Extraction and Selection In the framework, we need to measure similarity between a simulated (known) environment and an unknown environment. It is obviously infeasible to directly compare ratings in two environments. Instead, we define a set of features to represent each environment. Features refer to the statistics describing the characteristics of an environment (e.g. ratio of number of buyers versus sellers, variance of ratings per seller or per buyer, average number of transactions per time period, percentage of rated sellers). An exhaustive list of potential features are extracted from which the most relevant features are identified and used to represent the environment, to reduce the computational cost and increase the efficiency.

In the framework, we simulate another set of e-market environments using different settings than those of the simulated environments to find out best environment-model pairs. We then evaluate the performance of our framework in those simulated environments when using all the possible features. The features whose values significantly correlate to the performance of the framework are considered as relevant features. More specifically, in the framework, five widely used correlation and regression analysis techniques are adopted to measure the correlation between features and the performance of our framework, namely Pearson correlation, Kendall rank correlation coefficient, Spearman rank correlation coefficient, linear regression (backward) and linear regression (stepwise). The results of the correlation are also analyzed by the Paired-Samples T-test to check whether the correlation is statistically significant. Each correlation and regression analysis technique will give a combination of significantly relevant features recognized by that technique. We will then compare the average performance of our framework in the simulated environments (for feature selection) by using each combination of features and select the combination that gives the highest average performance.

Similarity between Environments With the carefully selected features, we can represent an environment by a vector where each element is the value of each feature. Given the assumption that any two environments are independent of each other, we can use Distance Correlation Analysis (e.g. Euclidean distance) to calculate similarity between two environments after feature values are normalized.

Choosing the most Suitable Trust Model Once we compute similarity between the unknown environment and each simulated environment, we can find out which simulated environment is the most similar to the unknown environment. According to the best environment-model pairs discovered earlier, the trust model performing the best in the most similar simulated environment will be chosen as the most suitable trust model for the unknown environment. In the case where some simulated environments are equally and most similar to the unknown environment, we can randomly choose one trust model among their best models.

In summary, our framework is generic. It can be further extended or concretized in the following aspects: 1) whenever the ground truth of any environment is discovered, it can be added into the framework to generate a more complete list of best environment-model pairs; 2) whenever a new trust model is proposed, it can be added into the framework. Our framework is capable of taking the advantage of the trust model to increase the performance of evaluating seller trustworthiness; 3) whenever a new insightful feature is identified, it can be added into the framework to participate in the feature selection process and may further increase the performance of the framework; 4) more promising feature selection methods such as incremental hill-climbers [Wettschereck and Aha, 1995], a wrapper model to measure the importance of features, can be adopted to increase the performance of the framework.

4 Experimentation

We instantiate our framework and conduct a series of experiments to demonstrate its effectiveness. Specifically, 972 e-market environments are simulated, consisting of different numbers of sellers (chosen from $\{10, 25, 50\}$) with different levels of honesty (uniformly distributed over $[0, 1]$). Total number of ratings is chosen from $\{50, 100, 250\}$. The marketplaces operate for 100 days. We simulate different distributions of fair ratings given by honest buyers: 1) *sparse*, where an honest buyer rates a seller at most once; 2) *intensive*, where an honest buyer rates a seller more than once; 3) *mixed*, which is combination of sparse and intensive scenarios. We also simulate different unfair rating attacking scenarios for dishonest buyers by adjusting 4 parameters: 1) *individual attack frequency* denoting the average number of unfair ratings provided by each dishonest buyer which can be sparse, intensive and mixed; 2) *attack period* referring to the period when unfair ratings are given, where 7 and 100 denote that dishonest buyers provide unfair ratings over one week (a concentrated attack) and 100 days (a distributed attack), respectively. (3) *attack target* taking a value of 0 or 1, indicating that attack targets are sellers with low reputation (below 0.5) or high reputation (greater than 0.5), respectively; 4) *overall attack rate* denoting the ratio of numbers of unfair ratings to fair ratings, chosen from $\{0.25, 1, 4\}$. Through the parameters of individual attack frequency and overall attack rate, the numbers of dishonest and honest buyers are determined. We also limit the total number of ratings to $\{50\}$ and $\{50, 100\}$ to simulate 324 and 648 environments respectively, to examine the influence of number of simulated environments on the effectiveness of our framework.

The framework includes 7 representative trust models: BRS [Whitby *et al.*, 2004], iCLUB [Liu *et al.*, 2011], TRAVOS [Teacy *et al.*, 2006], Personalized [Zhang and Cohen, 2007], Referral Networks [Yu and Singh, 2003], BLADE [Regan *et al.*, 2006] and Prob-Cog [Noorian *et al.*, 2011]. The following parameters are considered to design candidate trust models: 1) For BRS, the *quantile* parameter

Table 1: Selection of the Most Relevant Features

	Features	**Pearson (C1)**	**Kendall (C2)**	**Spearman (C2)**	**Backward (C3)**	**Stepwise (C4)**
1	Variance of the Percentage of Ratings for each Seller					$*$
2	Average Number of Ratings Provided by each Buyer for each Seller	$*$	$*$	$*$	$*$	$*$
3	Ratio of Number of Buyers versus Number of Sellers	$*$	$*$	$*$		
4	Skewness of Rating Period	$*$	$*$	$*$		
5	Variance of Percentage of Ratings Provided by each Buyer	$*$	$*$	$*$	$*$	$*$
6	Skewness of Number of Ratings Provided by each Buyer				$*$	$*$
7	Percentage of Satisfactory Sellers	$*$	$*$	$*$	$*$	$*$
8	Number of Buyers	$*$	$*$	$*$	$*$	
9	Average Number of Ratings for each Seller	$*$	$*$	$*$	$*$	$*$
10	Variance of Number of Ratings provided by each Buyer				$*$	$*$
11	Total Number of Ratings	$*$	$*$	$*$	$*$	$*$
12	Variance of Number of Ratings for each Seller	$*$	$*$	$*$	$*$	$*$
13	Skewness of Number of Ratings for each Seller	$*$	$*$	$*$		$*$
14	Average Number of Transactions in each Day				$*$	$*$
15	Total Percentage of Sellers Rated by Buyers	$*$	$*$	$*$	$*$	$*$
16	Time Period the Marketplace Operates	$*$	$*$	$*$		
17	Maximum Percentage of Ratings for Sellers	$*$	$*$	$*$	$*$	
18	Total Percentage of Buyers who are Active in the Marketplace	$*$			$*$	$*$

$q \in \{0.05, 0.1, 0.3, 0.5\}$ used to filter dishonest buyers; 2) In iCLUB, *minimum number of ratings* required to form a cluster in DBSCAN $minPts \in [1, 6]$, *maximum neighbor distance* in DBSCAN $\theta \in [0.3, 0.7]$ and *threshold* to choose the local or global component $\epsilon \in [3, 6]$; 3) For TRAVOS, number of *bins* to determine the acceptable error level in buyers' ratings $bin \in \{2, 3, 5, 8, 10\}$; 4) for Personalized, *error level* $\epsilon \in \{0.3, 0.5, 0.7\}$ and *confidence level* $\gamma \in \{0.3, 0.5, 0.7\}$; 5) For Referral Networks, *number of neighbors* in $\{2, 4, 6\}$ and *depth limit* of referral networks in $\{4, 6, 8\}$; 6) In Prob-Cog, *incompetency tolerance threshold* to filter out dishonest buyers $\mu \in \{0.1, 0.2, \ldots, 0.9\}$. In the end, we obtain 45 candidate trust models in total.

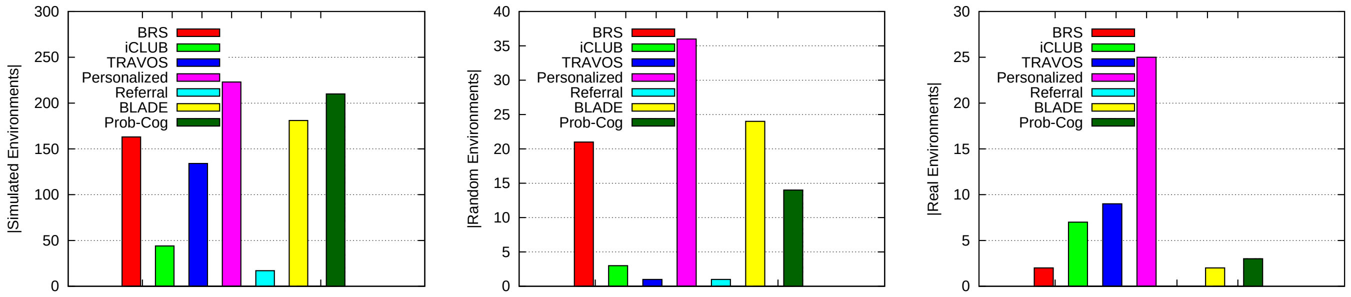
Best Environment-Model Pairs For each simulated environment, we find out the best candidate trust model by adopting the most commonly used evaluation metric, the mean absolute error (MAE) measured as the average difference between predicted trustworthiness of sellers and actual honesty of the sellers. We first calculate MAE of each candidate trust model for the simulated environments and select the one with the lowest MAE value. In the end, we obtain 972 best environment-model pairs. Fig. 2(a) illustrates the number of simulated environments where each candidate trust model achieves the best performance, which are 163, 44, 134, 223, 17, 181 and 210 for BRS, iCLUB, TRAVOS, Personalized, Referral, BLADE and Prob-Cog, respectively.

Feature Selection We consider 18 potential features to analyze the characteristics of the simulated environments, as listed in Table 1. We use some general statistical metrics to describe the features. For example, skewness describes the asymmetry from the normal distribution. A satisfactory seller refers to the one who receives more positive ratings than negative ones from buyers. An active buyer refers to the one who provides at least 1 rating to any seller.

To select the most relevant features, we adopt the five correlation and regression analysis techniques mentioned earlier. The results of the analysis of the 18 features on how they are correlated to the performance (MAE) of the framework is shown in Table 1. Here, '*' denotes that the feature has a significant correlation to the performance of the framework. In Table 1, columns C1, C2, C3 and C4 represent the combination of the features flagged with '*'. C5 represents a combination of all the features. To verify the effectiveness of the 5 feature combinations, we randomly generate a large number of known environments and compare the results. We obtain an average MAE of 0.44, 0.36, 0.25, 0.33, 0.32 for the combinations C1, C2, C3, C4 and C5 respectively. C3 has the lowest mean MAE, and is used for comparing the simulated and unknown environments hereafter.

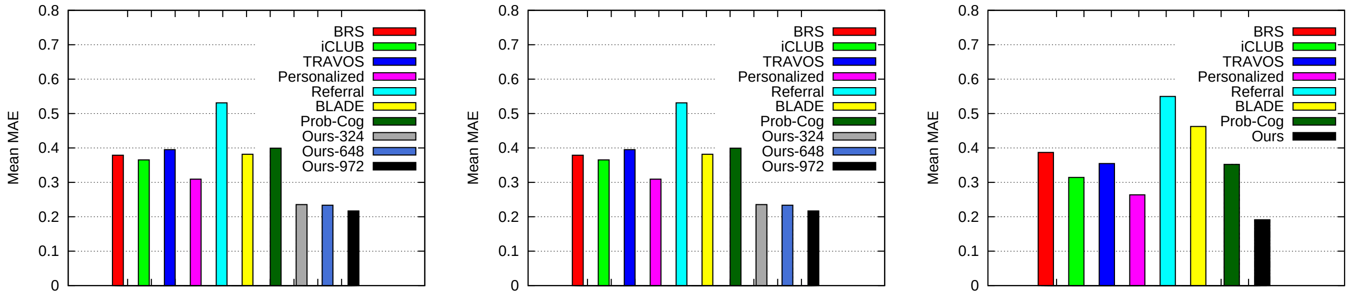
Unknown Environments for Testing The framework is evaluated using 6 categories of unknown environments (where ground truth about seller honesty or unfair ratings is in fact known) in both normal and extreme scenarios.

- *Unknown Random Environments* are generated using parameter values different from simulated environments: 1) number of sellers from $\{33, 66, 99\}$; 2) total number of ratings from $\{333, 666, 999\}$; 3) ratio of number of unfair ratings versus fair ratings from $\{0.1, 1, 10\}$; 4) time period of attacks from $\{50, 100\}$. We randomly choose 100 environments for testing data from 972 generated environments.
- *Unknown Real Environments* are real environments with simulated unfair rating attacks. Real data is obtained from IMDB.com where users rate movies directed by different directors. We remove outlying ratings, then select only directors whose movies are very highly rated, and simulate 3 types of unfair rating attacks, namely RepBad, RepSelf and RepTrap [Yang *et al.*, 2008], to bad-mouth targeted directors. We also employ a combination of these attacks. Finally, we generate 48 such real environments. The aim here is to correctly model the trustworthiness of directors.
- *Large Environments* where the number of sellers is larger than 50, number of ratings larger than 100, number of buyers larger than 80. We generate 160 such environments.
- *Extremely Sparse Environments* where buyers do not provide sufficient ratings. Specifically, each buyer gives an average of 0.1 ratings to sellers. We generate 36 such environments where the number of sellers is 10, total number of ratings 100, and overall attacking rate in $\{0.25, 1, 4\}$.
- *Environments with Dynamic Seller Behavior* where sellers change their behavior dynamically. In this category, the num-



(a) (b) (c)

Figure 2: The Number of Times each Trust Model is Selected as the most Suitable Model for: (a) Simulated Environments; (b) Unknown Random Environments; and (c) Unknown Real Environments



(a) (b) (c)

Figure 3: Mean MAE of our Framework and other Trust Models for: (a) Unknown Random Environments; (b) Unknown Real Environments; and (c) Large Environments

ber of sellers is 10 and total number of ratings is 50.

- *Environments with Many Attacks* are intensive attacking scenarios where attack rate is larger than 10. For such environments, we use real data from IMDB and simulate RepBad, RepSelf, RepTrap attacks and their combination.

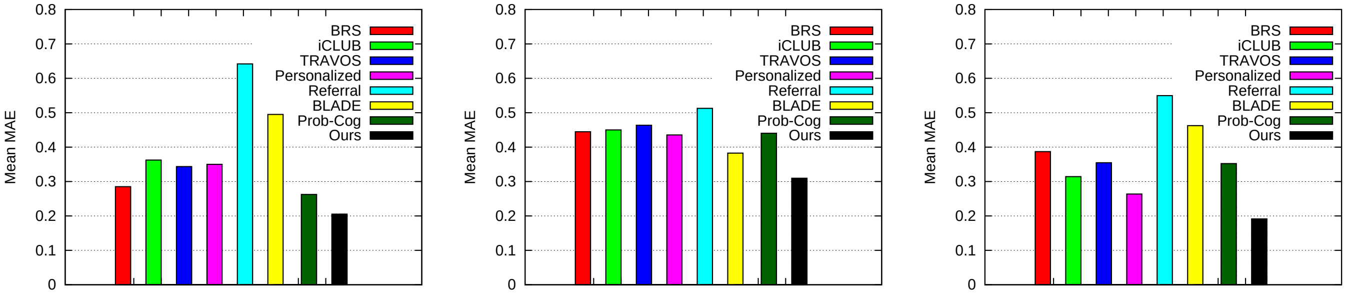
Experimental Results The aim of our framework is to choose the most suitable trust models and parameter values for given unknown environments. Fig. 2(b-c) shows the number of the unknown random environments and unknown real environments respectively for which each trust model is chosen as the most suitable one. The numbers are 21, 3, 1, 36, 1, 24 and 14 for BRS, iCLUB, TRAVOS, Personalized, Referral, BLADE and Prob-Cog respectively in the 100 unknown random environments, and 2, 7, 9, 25, 0, 2 and 3 for these models in the 48 unknown real environments. It indicates that our framework is able to choose different models from a candidate set for various unknown environments.

Table 2 presents the accuracy of our framework in choosing the most suitable trust models (with the most suitable parameters) in unknown environments. A correct selection indicates that the trust model chosen is the same as the best model identified by evaluating all candidate trust models in a given unknown environment. In Table 2, ϵ is a tolerance value, indicating that the difference between the MAE of the chosen trust model and that of the truly most suitable model is within ϵ. From Table 2, we can see that the accuracy of our framework increases as the number of simulated environments (SE) increases, and is the best when there are 972 simulated environments. Even with only 324 simulated environments, the performance of our framework is still acceptable, selecting the most suitable models for 81.0% and 81.3% of unknown random and real environments, respectively. The performance of our framework with tolerance $\epsilon = 0.05$ shows considerable improvement in correctly selecting the most suitable trust models (an average increase of 7.9%) and greater improvement in correctly selecting the most suitable models and parameters (15.2% on average). Thus, it shows that our framework can choose candidate models whose performance is very close to the ideal case.

Table 2: Accuracy of Choosing Most Suitable Models (with Parameters) for Unknown Environments

Unknown Random Environments	324 SE	648 SE	972 SE
Correct Models	81.0%	84.0%	92.0%
Correct Models with ϵ	87.0%	89.0%	95.0%
Correct Models and Paras	72.0%	76.0%	82.0%
Correct Models and Paras with ϵ	85.0%	86.0%	94.0%
Unknown Real Environments	324 SE	648 SE	972 SE
Correct Models	81.3%	83.3%	83.3%
Correct Models with ϵ	89.6%	95.8%	95.8%
Correct Models and Paras	72.9%	75.0%	77.1%
Correct Models and Paras with ϵ	89.6%	95.8%	95.8%

Fig. 3(a-b) shows the mean MAE of our framework in comparison with the other trust models in unknown random and unknown real environments, respectively. For other trust models in an unknown environment, we use their best parameter values. For all the 3 experimental settings with 324, 648 and 972 simulated environments, our framework obtains the smallest mean MAE values, indicating that our framework is able to choose better trust models to evaluate seller trustworthiness than always applying a single model.



(a) (b) (c)

Figure 4: Mean MAE of our Framework and other Trust Models for: (a) Extremely Sparse Environments; (b) Environments with Dynamic Seller Behavior; and (c) Environments with Many Attacks

Fig. 3(c) and Fig. 4 show the mean MAE of trust models in the 4 extreme scenarios (i.e. large environments, extremely sparse environments, environments with dynamic seller behavior and environments with many attacks). Our framework outperforms all the other trust models in the testing environments. Table 3 presents the probability of choosing trust models in each of these 4 extreme cases.

Table 3: Probability of Choosing the Trust Models in the Four Extreme E-Market Scenarios

Trust Models	Large	Sparse	Dynamic	Many
BRS	0.6%	16.7%	7.4%	6.9%
iCLUB	26.9%	13.9%	0.0%	10.3%
TRAVOS	23.8%	0.0%	4.6%	13.8%
Personalized	38.2%	2.8%	27.8%	58.7%
Referral	0.6%	2.8%	18.5%	0.0%
BLADE	9.3%	11.1%	31.5%	0.0%
Prob-Cog	0.6%	52.7%	10.2%	10.3%

More specifically, Fig. 3(c) shows that iCLUB, TRAVOS and Personalized obtain smaller mean MAE than other trust models in large environments. The reason is that these three trust models are able to distinguish dishonest and honest advisors when they get sufficient rating sources. In Table 3, we can see that our framework selects iCLUB, TRAVOS and Personalized with the highest probabilities as 26.9%, 23.8% and 38.2% for large environments, respectively.

As shown in Fig. 4(a), we find that in sparse environments, BRS and Prob-Cog perform better than other trust models. This is because BRS adopts the “majority-rule” to consider the opinions from other advisors, and Prob-Cog extends the incompetence tolerance threshold to incorporate a larger number of advisors’ ratings. Both models obtain a comparatively low mean MAE value because they are less restrictive in accepting opinions from advisors compared to other trust models. In Table 3, our framework selects BRS and Prob-Cog with the highest probabilities as 16.7% and 52.7% for sparse environments, respectively.

Fig. 4(b) shows that Personalized and BLADE outperform others trust models in the environments where sellers change their behavior dynamically. To explain, Personalized considers advisors’ latest ratings within a certain time window which alleviates the influence of seller dynamic behaviors. BLADE re-interprets advisors’ ratings based on learning thereby takes into account the changing behavior of sellers and buyers. In Table 3, our framework selects Personalized and BLADE with the highest probabilities as 27.8% and 31.5% for these environments, respectively.

Fig. 4(c) shows that Personalized and TRAVOS perform well in the environments with many attacks. The characteristics of attacks play a major role in judging the performance of the trust models. In these extreme environments, the attackers (dishonest advisors) first give honest ratings to non-target sellers to promote themselves, and then provide unfair ratings to bad-mouth target sellers. The performance of Personalized and TRAVOS is better because they both model advisor trustworthiness more accurately by comparing buyers’ own opinions and advisors’ ratings on commonly rated sellers. Also, in the environments, we select only buyers with sufficient personal experience (ratings) which Personalized and TRAVOS take advantage of. In Table 3, our framework selects Personalized and TRAVOS with probabilities as 58.7% and 13.8% for the environments with many attacks, respectively.

In summary, from Table 3, Fig. 3(c) and Fig. 4, the results indicate that our framework is able to select suitable trust models for extreme scenarios and obtain more accurately seller trustworthiness than any individual trust model.

5 Conclusion

In this paper, we propose a framework to choose suitable trust models for the environments where ground truth about agent behaviors is unknown. Given an unknown environment, we firstly find a closely similar simulated environment (with specified ground truth). Then, the trust models performing the best in the simulated environment are chosen. Experimental results confirm that our framework can accurately select suitable trust models for various unknown environments. Using our framework to choose trust models for unknown environments is better than always applying any single trust model, in terms of the accuracy of evaluating seller trustworthiness. For future work, instead of relying on similarity between environments, we will apply machine learning techniques, such as decision tree, to choose appropriate trust models for unknown environments. We will also continue to evaluate our framework by incorporating more trust models and involving more real datasets with manually detected ground truth [Irissappane *et al.*, 2012].

Acknowledgments

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

References

[Fullam and Barber, 2007] Karen K. Fullam and K. Suzanne Barber. Dynamically learning sources of trust information: experience vs. reputation. In *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2007.

[Hang *et al.*, 2009] Chung-Wei Hang, Yonghong Wang, and Munindar P. Singh. Operators for propagating trust and their evaluation in social networks. In *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2009.

[Huynh, 2009] T.D. Huynh. A personalized framework for trust assessment. In *Proceedings of the ACM Symposium on Applied Computing (SAC)*, 2009.

[Irissappane *et al.*, 2012] Athirai A. Irissappane, Siwei Jiang, and Jie Zhang. Towards a comprehensive testbed to evaluate the robustness of reputation systems against unfair rating attacks. In *Proceedings of the International Conference on User Modeling, Adaptation and Personalization (UMAP) Workshop on Trust, Reputation and User Modeling*, 2012.

[Jøsang *et al.*, 2007] Audun Jøsang, Roslan Ismail, and Colin Boyd. A survey of trust and reputation systems for online service provision. *Decision Support Systems*, 43(2):618–644, 2007.

[Liu *et al.*, 2011] S. Liu, J. Zhang, C. Miao, Y.L. Theng, and A.C. Kot. iCLUB: An integrated clustering-based approach to improve the robustness of reputation systems. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2011.

[Noorian *et al.*, 2011] Z. Noorian, S. Marsh, and M. Fleming. Multi-layer cognitive filtering by behavioral modeling. In *Proceedings of the International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2011.

[Regan *et al.*, 2006] K. Regan, P. Poupart, and R. Cohen. Bayesian reputation modeling in e-marketplaces sensitive to subjectivity, deception and change. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 2006.

[Sabater and Sierra, 2005] Jordi Sabater and Carles Sierra. Review on computational trust and reputation models. *Artificial Intelligence Review*, 24(1):33–60, 2005.

[Sormo *et al.*, 2005] Frode Sormo, Jorg Cassens, and Agnar Aamodt. Explanation in case-based reasoning—perspectives and goals. *Artificial Intelligence Review*, 24(2):109–143, 2005.

[Teacy *et al.*, 2006] W.T.L. Teacy, J. Patel, N.R. Jennings, and M. Luck. TRAVOS: Trust and reputation in the context of inaccurate information sources. *Autonomous Agents and Multi-Agent Systems*, 12(2):183–198, 2006.

[Wang and Singh, 2010] Yonghong Wang and Munindar P. Singh. Evidence-based trust: A mathematical model geared for multiagent systems. *ACM Transactions on Autonomous and Adaptive Systems (TAAS)*, 5(4):1–28, 2010.

[Wettschereck and Aha, 1995] Dietrich Wettschereck and David W Aha. Weighting features. *Case-Based Reasoning Research and Development*, pages 347–358, 1995.

[Whitby *et al.*, 2004] A. Whitby, A. Jøsang, and J. Indulska. Filtering out unfair ratings in bayesian reputation systems. In *Proceedings of the International Joint Conference on Autonomous Agenst Systems (AAMAS) Workshop on Trust in Agent Societies (TRUST)*, 2004.

[Yang *et al.*, 2008] Yafei Yang, Qinyuan Feng, Yan Lindsay Sun, and Yafei Dai. RepTrap: A Novel Attack on Feedback-based Reputation Systems. In *Proceedings of the International Conference on Security and Privacy in Communication Networks (SecureComm)*, 2008.

[Yu and Singh, 2003] B. Yu and M.P. Singh. Detecting deception in reputation management. In *Proceedings of the International Joint Conference on Autonomou Agents and Multiagent Systems (AAMAS)*, 2003.

[Zhang and Cohen, 2007] Jie Zhang and Robin Cohen. A comprehensive approach for sharing semantic web trust ratings. *Computational Intelligence*, 23(3):302–319, 2007.