A POMDP Based Approach to Optimally Select Sellers in Electronic Marketplaces

Athirai A. Irissappane
Nanyang Technological University, Singapore
athirai001@e.ntu.edu.sg

Frans A. Oliehoek
University of Amsterdam / Maastricht University
The Netherlands
f.a.oliehoek@uva.nl

Jie Zhang
Nanyang Technological University, Singapore
zhangj@ntu.edu.sg

ABSTRACT

Selecting a seller in e-markets is a tedious task that we might want to delegate to an agent. Many approaches to constructing such agents have been proposed, building upon different foundations (decision theory, trust modeling) and making use of different information (direct experience with sellers, reputation of sellers, trustworthiness of other buyers called advisors, etc.). In this paper, we propose the SALE POMDP, a new approach based on the decision-theoretic framework of POMDPs. It enables optimal trade-offs of information gaining and exploiting actions, with the ultimate goal of maximizing buyer satisfaction. A unique feature of the model is that it allows querying advisors about the trustworthiness of other advisors. We represent the model as a factored POMDP, thereby enabling the use of computationally more efficient solution methods. Evaluation on the ART testbed demonstrates that SALE POMDP balances the cost of obtaining and benefit of more information more effectively, leading to more earnings, than traditional trust models. Experiments also show that it is more robust to deceptive advisors than a previous POMDP based approach, and that the factored formulation allows the solution of reasonably large instances of seller selection problems.

Categories and Subject Descriptors

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

General Terms

Design; Performance

Keywords

Seller Selection; E-Marketplace; POMDPs

1. INTRODUCTION

In multi-agent based e-marketplaces, self-interested selling agents can act maliciously by not delivering products with the same quality as promised. It is thus important for buying agents to analyze their quality and determine which sellers to do business with. Buyers maintain beliefs over the quality levels of sellers, based on their previous transactions, which may help them choose good sellers. However, realistically, in most e-marketplaces, buyers often encounter sellers with which they have no previous experience. In such cases, they can query other buyers (called advisors) about the sellers.

A number of trust models (e.g., BRS [1], TRAVOS [2], Personalized [3], BLADE [4], etc.) have been proposed by researchers in the multi-agent community to help buyers assess seller quality and choose transaction partners. These approaches work by combining the buyer's own belief and those of the advisors, to estimate the true quality of the seller [5]. However, the above approaches mainly focus on accurately estimating the quality of sellers rather than optimally choosing a seller to perform transaction; they simply query *all* advisors about the sellers' quality and fail to reason *when* it is necessary to query advisors (about the sellers' quality), though they may determine *whom* to query by analyzing the quality levels (trustworthiness) of advisors. In settings where there are costs associated with queries, such approaches may lead to diminished utility, since the cost of querying advisors may be greater than the value derived from a successful transaction with the seller.

To help overcome this problem, Regan *et al.* [6] propose the *Advisor* POMDP, which considers the seller selection problem as a *Partially Observable Markov Decision Process* (POMDP). POMDPs provide a natural model for sequential decision making under uncertainty [7]. The main advantage that this approach brings to the seller selection problem is that, rather than trying to achieve the most accurate estimate of sellers, it tries to select good sellers optimally with respect to its belief. However, the *Advisor* POMDP does not reason about advisors' quality. Also, in *Advisor* POMDP, each advisor, when queried, provides opinions about all sellers, which may result in a lot of unnecessary information; rather than estimating the quality of all sellers, the only goal should be to select the seller with high quality.

This paper presents the *Seller & Advisor seLEction (SALE) POMDP*, a novel POMDP based framework to deal with the seller selection problem and overcome the above problems by reasoning about advisor trustworthiness and selectively querying for information. Like the *Advisor* POMDP, the agent tasked with selecting the seller is modeled using a POMDP, which allows it to trade-off the expected benefit and cost of more information gaining action, thus aiming to optimize the total utility for its owner (buyer). However, in this paper we make a number of additional contributions: 1) Because the SALE POMDP models the behavior (e.g., quality and/or trustworthiness) of advisors as part of the

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.*

state, it can deal with deceptive or poor quality advisors, provided that the agent has accurate beliefs. 2) In order to provide accurate beliefs about advisors without having to engage in many costly transactions, the SALE POMDP agent can ask advisors about other advisors. 3) Because these so-called 'advisor queries' are modeled as normal POMDP actions, the optimal policy will balance the expected benefit of obtaining more information about the advisors against the cost of obtaining this information. This results in an agent that will selectively query about advisors' quality just enough to identify a trustworthy advisor to ask about the various sellers. In this way, trust propagation becomes an integral part of optimal sequential decision making for the seller selection problem. 4) We show how SALE POMDP can use both reputation information from advisors as well as direct experience with sellers. This is crucial in settings where there is a large number of deceptive advisors: in such cases, the real experience allows us to identify such deceptive behavior. 5) While optimally solving a POMDP is a computationally hard problem, we show that by modeling the SALE POMDP as a *factored* POMDP and using solvers that exploit this structure [8], we can overcome these issues to a great extent. 6) We present an extensive empirical evaluation on the ART testbed [9] to demonstrate the efficacy of the SALE POMDP. In particular, our experiments demonstrate that in single transaction settings, SALE POMDP outperforms other trust models in terms of revenue, and the method is resilient to strategic attacks. In sequential competitive settings, SALE POMDP significantly outperforms other trust models. In addition, an analysis shows that the performance of the SALE POMDP is quite robust to the specification of its parameters, and that the factored formulation allows it to scale to reasonably large seller selection problems without loss in quality.

2. BACKGROUND

Formally, a *Partially Observable Markov Decision Process (POMDP)* [7] is described by a tuple: $\langle \mathcal{S}, \mathcal{A}, T, R, \Omega, O \rangle$, with $\mathcal{S}$, the set of states; $\mathcal{A}$, the set of actions; T, the transition model; R, the reward function; Ω, a finite set of observations and O, the observation model. It is also common to assume an initial state distribution b_0. At each time step, the environment has a state $s \in \mathcal{S}$. The agent takes some action $a \in \mathcal{A}$, which causes a state transition from s to a new state s', using T, the transition model that specifies probabilities $\Pr(s'|s,a)$. The agent also receives observations based on the observation model O, that specifies the probabilities $\Pr(o|a,s')$. For a transition, the agent receives a reward $R(s,a,s')$. Additionally, the *horizon*, h, represents the number of time steps, or *stages*, for which we want to plan. We will assume that h is infinite in this paper.

When the POMDP agent interacts with the environment, it maintains a *belief* $b \in \mathcal{B}$, i.e., a probability distribution over states via Bayes' rule. That is, when $b(s)$ specifies the probability of s (for all s), we can derive b' an updated belief after taking some action a and receiving an observation o. Assuming discrete sets of states and observations, this update can be written as follows:

$$b'(s') = \frac{\Pr(s', o|b, a)}{\Pr(o|b, a)} = \frac{\Pr(o|a, s')}{\Pr(o|b, a)} \sum_s Pr(s'|s, a)b(s) \quad (1)$$

Here, $\Pr(o|b,a)$ is a normalization factor.

A POMDP policy $\pi : \mathcal{B} \rightarrow \mathcal{A}$, maps a belief $b \in \mathcal{B}$ to a prescribed action $a \in \mathcal{A}$. A policy π is associated with a value function $V_\pi(b)$ that specifies the expected total reward of executing policy π starting from b:

$$V_\pi(b) = \mathbb{E}\Big[\sum_{t=0}^{h-1} \gamma^t R(s,a,s') \mid \pi, b\Big] \quad (2)$$

The solution to a POMDP is an optimal policy that maximizes the expected total reward. Finding an optimal policy π^* is considered to be intractable in general (PSPACE complete [10]), however, in recent years substantial advances have been made in approximate solutions (e.g., [11], [12], [13]).

3. THE SALE POMDP MODEL

In this section, we will introduce the SALE POMDP model, showing how the selection problem is mapped to a POMDP.

3.1 Basic SALE POMDP

The main aim of the SALE POMDP model is to optimally select a seller with sufficient quality to buy from. This stands in contrast to methods that focus on accurately determining the quality of sellers. The SALE POMDP framework assumes that both sellers and advisors have quality levels and considers them as part of the state space. Briefly, the model works by improving its beliefs over the quality levels of sellers and advisors by querying advisors about the quality of sellers/other advisors in the system, until it is sure that it has identified a seller with sufficient quality.

We will mainly focus on how the SALE POMDP models a buying agent in a single transaction scenario, when the buyer is in fact a newcomer to the market, which is the case with most real world e-marketplaces. Discussions on how the model can be extended to a multiple transaction scenario is also presented in Sec. 3.4.

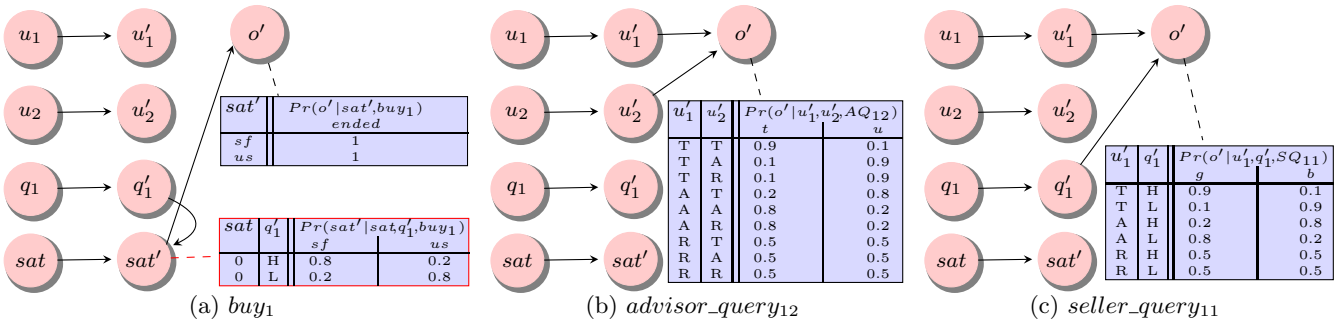
Given I advisors that can be queried about the quality of J sellers, each SALE POMDP agent can be described in terms of states, actions, observations and rewards as follows.

States. A state contains the quality levels[1] of each seller, advisor and the status of the transaction with the seller. Let $\mathcal{Q}$ be the discrete set of seller quality levels and $\mathcal{U}$ be the set of advisor quality levels. Then, a state is a tuple $s = \langle \vec{q}, \vec{u}, sat \rangle$, where $\vec{q} \in \mathcal{Q}^J$ is a vector indicating the quality of each seller, $\vec{u} \in \mathcal{U}^I$ a vector indicating the quality of each advisor and *sat* is a variable that indicates the status of the transaction (with values $not\_started(ns)$, $satisfactory(sf)$, $unsatisfactory(us)$, $gave\_up(gu)$, $finished(f)$). We also write q_j for the j-th element of $\vec{q}$ and u_i for the i-th element of $\vec{u}$. The end of the decision process (with initial value $sat = not\_started$) is modeled using sets of terminal states ($satisfactory$, $unsatisfactory$, $gave\_up$). Any transitions from the terminal states will result in $sat = finished$.

Actions. The model knows the following types of actions: 1) $seller\_query_{ij}(SQ_{ij})$, ask advisor i about quality of seller j; 2) $advisor\_query_{ii'}$ $(AQ_{ii'})$, ask advisor i about quality of advisor i'; 3) buy_j, buy from seller j; 3) do_not_buy (DNB), decide not to buy from any seller in the market.

Transitions. We assume that when taking a query action,

[1]We assume discrete quality levels, in order to use standard POMDP solvers. Extensions to continuous quality levels lead to continuous POMDPs [14].



(a) buy_1 (b) $advisor\_query_{12}$ (c) $seller\_query_{11}$

Figure 1: DBN and CPT for the SALE POMDP transition and observation functions for each action type. Variables without a CPT shown are 'static': they preserve the previous value with probability 1.0.

the state does not change:

$$\forall_{i,j} \quad \Pr(s'|s, SQ_{ij}) = \delta_{ss'} \tag{3}$$

$$\forall_{i,i'} \quad \Pr(s'|s, AQ_{ii'}) = \delta_{ss'} \tag{4}$$

$\delta_{ss'}$ is the Kronecker delta, i.e., 1 if and only if $s = s'$. When taking buy_j and DNB actions, the state will always transition to a terminal state, i.e., buy_j actions may result in a successful ($sat = satisfactory$) or unsuccessful ($sat = unsatisfactory$) transaction and DNB will result in $sat = gave\_up$. The transition probabilities to terminal states give the *definition* of quality levels. Generally, chances of transition to *satisfactory* should be higher when buying from 'high quality' sellers. Note, however, the framework allows much richer interpretations of the quality levels: essentially each level corresponds to a potential model of sellers.

Rewards. There is small cost for the ask actions $R(s, SQ_{ij})$ and $R(s, AQ_{ii'})$. A reward is associated with a successful transaction $R(s, buy_j, s' = \langle \vec{q}, \vec{u}, sat = sf \rangle) = R_{sat}$, otherwise a penalty is levied $R(s, buy_j, s' = \langle \vec{q}, \vec{u}, sat = us \rangle) = R_{unsat}$. There is a penalty ($R_{unsat}$) for taking the DNB action, when in fact there is a seller of high quality, otherwise there is a reward for this action (R_{sat}). Once the terminal states are reached, no further rewards are given.

Observations. When a *query* action is performed, the agent will receive an observation based on the set of discriminated quality levels. After SQ_{ij} action, the agent receives an observation $o \in \{good, bad\}$, corresponding to the quality of seller j. After $AQ_{ii'}$ action, it gets an observation $o \in \{trustworthy, untrustworthy\}$, corresponding to the quality of advisor i'. On transition to a terminal state, it receives the observation *ended*.

Observation Function. It specifies $\Pr(o|a, s')$. As in the *Advisor* POMDP, there is no a priori correct way to specify the observation probabilities. Similar to the transition probabilities for the buy action, the probabilities for the observation function *define* the meaning of different trust levels. In general, the idea is that trustworthy advisors will give more accurate and consistent answers than untrustworthy ones, but again, much richer models of advisors are possible.

Initial State Distribution. It is dependent on the subjective beliefs of the agent, when the need for purchasing from sellers arises. For simplicity, one may start with a uniform belief over the quality levels, but a different initial belief can also be obtained as a result of previous interactions. This will be further discussed in Sec. 3.4.

3.2 Factored Representation

POMDPs with very large state spaces are impractical to be solved by the classic solution algorithms. However, often, such state spaces can be described using a set of state variables, and the effects of actions in terms of their effects on these variables. Dynamic Bayesian Networks (DBNs) with Conditional Probability Tables (CPTs) are often used to represent these effects compactly. The resulting representations are referred to as *factored* representations, and solvers such as symbolic Perseus [8] can exploit such *factored* nature of POMDPs to resolve scalability issues that arise due to large state spaces. The SALE POMDP can in fact be represented as a *factored* POMDP, which thereby, allows it to scale to larger seller selection problems.

To illustrate the factored nature of SALE POMDP, we will consider a simple case of seller selection problem with 1 seller (Sel_1) and 2 advisors (Adv_1, Adv_2) such that $\mathcal{Q} = \{H, L\}$, representing H(igh) and L(ow) seller quality levels and $\mathcal{U} = \{T, A, R\}$, representing advisor quality levels: T(rustworthy), always providing true opinions; A(dversarial), often untrustworthy providing complimentary opinions and R(andom), being trustworthy or untrustworthy randomly.

Buy Action. Fig. 1(a) illustrates the transition probabilities for the sat variable (CPT in red) and the observation probabilities for the variable o. The probability of transferring to state s', given that action buy_1 was taken in state s, can be factored into a product of smaller conditional distributions with respect to its parent variables as in Eqn. 5.

$$\begin{aligned}\Pr(s'|s, buy_1) = \Pr(u'_1|u_1, buy_1) \times \Pr(u'_2|u_2, buy_1) \times \\ \Pr(q'_1|q_1, buy_1) \times \Pr(sat'|sat, q'_1, buy_1)\end{aligned} \tag{5}$$

The transition probabilities are framed such that buying from a high quality seller will lead to *satisfactory* with 80% probability: $\Pr(sat' = sf|\langle sat = ns, q'_1 = H \rangle, buy_1) = 0.8$. Similarly, $\Pr(sat' = us|\langle sat = ns, q'_1 = L \rangle, buy_1) = 0.8$, i.e., buying from a low quality seller leads to $unsatisfactory$ with 80% probability. Since the buy_1 action always results in a terminal state ($satisfactory$, $unsatisfactory$), the observation *ended* is received with probability 1.0. Also, any further transitions from the terminal state will lead to $sat' = finished$: $\Pr(sat' = f|\langle sat = sf, q'_1 = H \rangle, buy_1) = 1.0$.

For the DNB action, sat' is not dependent on seller quality q'_1 and leads to $gave\_up$ with probability 1.0: $\Pr(sat' = gu|\langle sat = ns \rangle, DNB) = 1.0$. Since the terminal state $gave\_up$ is reached, the observation *ended* is received with probability 1.0 and any transitions hereon results in $sat' = finished$.

Advisor Query. Fig. 1(b) shows the observation probabilities for the variable o, on taking action AQ_{12} (query Adv_1 about Adv_2). Since query actions do not alter the state, all CPTs for state factors are static. The figure clearly illustrates that the observation probability for this action only depends on the trust levels of advisors: $\Pr(o'|u_1', u_2', AQ_{12})$. It also shows the CPT containing the actual probabilities. In this case, only two observations have non-zero probability: 1) *trustworthy* (t), where Adv_1 says Adv_2 is trustworthy; and 2) *untrustworthy* (u), where Adv_1 says Adv_2 is untrustworthy. The observation probabilities encode that asking a trustworthy (T) advisor gives more accurate observations than untrustworthy advisors, exhibiting *adversarial* (A) or *random* (R) behavior.

Seller Query. Similarly, Fig. 1(c) shows the probabilities for the action SQ_{11} (query Adv_1 about Sel_1). Also in this case, the transitions are static and observation probabilities depend on a subset of state factors: $\Pr(o'|u_1', q_1', SQ_{11})$. Two observations are possible after a seller query: 1) *good* (g), where Adv_1 says Sel_1 is high quality; and 2) *bad* (b), where Adv_1 says Sel_1 is low quality. Again, the probabilities are such that asking a trustworthy advisor gives more accurate observations.

3.3 Belief Update in SALE POMDP

In the SALE POMDP, belief updates are performed such that they correlate the state factors in meaningful ways. Here, we briefly illustrate this process. A more detailed illustration on the process of belief updating in SALE POMDP can be found in [15].

Fig. 2 shows the partial SALE POMDP policy for our running example (seller selection problem with 1 seller and 2 advisors). Each state $s = \langle \langle q_1 \rangle, \langle u_1, u_2 \rangle, sat \rangle$, where $q_1 \in \{H, L\}$, u_1 and $u_2 \in \{T, A\}$ (so we only consider adversarial advisors here). The transition and observation probabilities have the same values as mentioned in Sec. 3.2. The beliefs prior to taking the actions (represented by nodes in Fig. 2) are shown using tables associated with the nodes. The state variable *sat* is not shown for simplicity.

Initially (before action AQ_{12}), we assume uniform quality levels for Sel_1 (0.5 *high* and 0.5 *low*) and advisors Adv_1 and Adv_1 (0.5 *trustworthy* and 0.5 *adversarial*). The belief state corresponding to the initial state distribution is shown in the table associated with the first AQ_{12} action in Fig. 2. On receiving observation $o' = t$ (*trustworthy*) after action AQ_{12}, (traversing through the left child of the root in Fig. 2), beliefs are updated (using Eqn. 1) such that states with $u_1 = T$ and $u_2 = T$ are given more weights than states with $u_1 = A$ and $u_2 = T$. Similarly, beliefs are updated for the AQ_{21} actions, determining both Adv_1 and Adv_2 to be trustworthy ($HTT = 0.3$). Here, states where $q_1 = L$ have same values as those with $q_1 = H$, hence not presented in the tables.

Adv_1 is then queried about Sel_1 (action SQ_{11}). When the agent receives an observation $o' = g$ (*good*), the beliefs for the seller are updated such that more weights are given to the states where seller is *high* quality and advisor is *trustworthy*, and less weights to states where seller is *low* quality. We can see from Fig. 2 that ($HTT = 0.55$ and $LTT = 0.06$) at this point, resulting in the BUY (buy_1) action. Similarly, if the agent receives an observation $o' = B$ (*bad*), the DNB action is taken. The beliefs obtained when the observation of AQ_{12} action, $o' = u$ (*untrustworthy*) can be seen by traversing through the right child of the root.

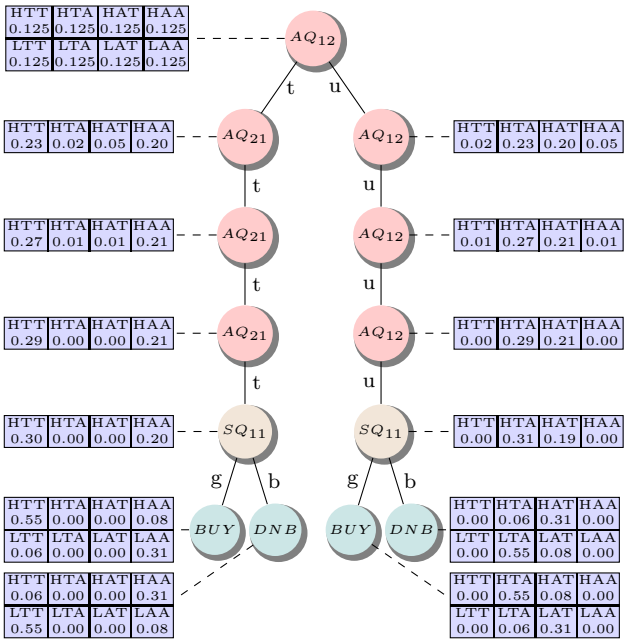


Figure 2: (Partial) SALE POMDP policy

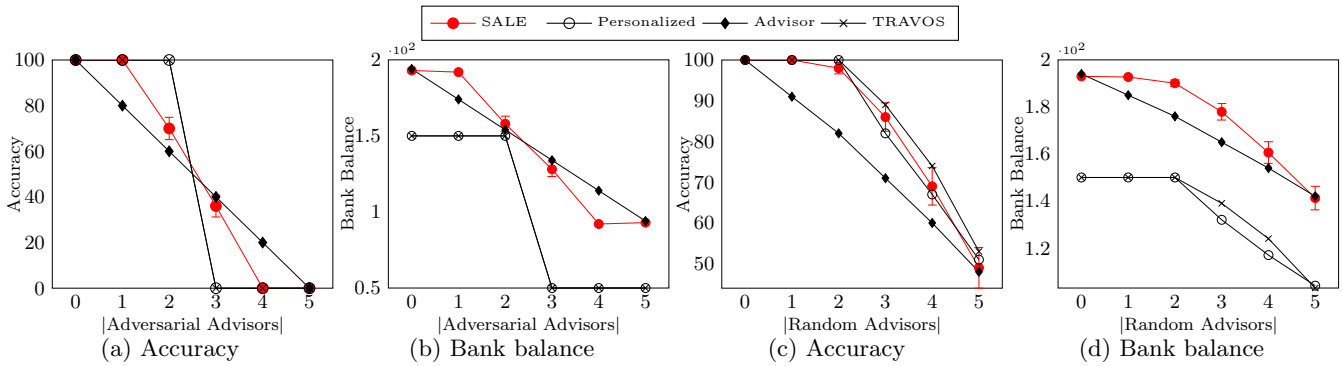
3.4 Multiple Transactions

While the basic SALE POMDP models just a single transaction, it is also possible to apply the model in a sequential setting, where the buyer may engage in multiple transactions. That is, once the buy_j or DBN action is performed, the resulting belief can be used as the basis for an initial belief for a new seller selection instantiation. In a bit more detail, there are two sources of previous experience: 1) previous seller selection tasks: the modified belief state resulting from advice in a previous problem can be retained; 2) actual experiences with sellers: even though in the decision process, we model a transition to a terminal state with a deterministic *ended* observation, the actual transaction will result in the owner of the agent being satisfied or not and this information can be used to update the final belief of the agent's previous seller selection task, giving a new initial belief for a new task[2]. This type of sequential tasks can be implemented in two ways. First it is possible to change the POMDP formulation, such that already during planning we anticipate multiple transactions. While this is the most principled solution, it will make the model more complex and we do not expect that reasoning over future transactions will bring a significant improvement. Instead, we solve the single-transaction SALE POMDP models sequentially, updating the beliefs in between as explained.

4. EVALUATION

We perform detailed experiments to demonstrate the effectiveness of the SALE POMDP. In particular, we compare its performance to other trust models in a single transaction scenario, as well as a multiple transaction setting. Additionally, since the specification of the behavior of advisors

[2]In fact this can be an important mechanism to deal with advisors that are consistent but deceptive and settings in which the majority of advisors are untrustworthy.



(a) Accuracy (b) Bank balance (c) Accuracy (d) Bank balance

Figure 3: Single transaction setting: (a-b) adversarial scenario; (c-d) random scenario

(via specification of the transition and observation model) is done by the designer, we analyze the robustness of the SALE POMDP to changes in these choices. Finally, we conduct experiments to demonstrate the scalability achieved using the factored formulation of the SALE POMDP.

Experimental Setup. We evaluate the SALE POMDP model using a modified version of the Agent Reputation and Trust (ART) testbed [9]. In this testbed, a set of agents interact over the course of multiple rounds. Each agent at every round gets a (varying) number of clients that ask it to appraise a painting. That is, the agent should report to the client whether the painting is of high or low quality. Additionally, each of the agents acts as an advisor for the other agents. If an appraising agent does not have the expertise to complete the appraisal, it can request opinions from these advisors about the quality of the painting by paying a specified cost. The advisor providing opinion, obtains information about the ground truth of the requested painting from the simulation engine[3], using which it may decide whether to provide an incorrect opinion (by altering the true quality of the painting obtained from the simulation engine) or not, before sending it to the requester. Appraisers can also request reputation values about other appraisers at a specified cost. Again, advisors may alter the reputation values (which are calculated by their own model) before sending it to the requesting agent. The predicted value of the appraisal is determined by the appraiser using advisor's opinions and then submitted to the testbed simulation engine. Appraisers correctly identifying the quality of a painting (high/low), receive more clients and thus more profit, increasing his bank balance in the long run. The client share for each appraiser (initially evenly distributed among appraisers) is calculated based on their average appraisal error in identifying the quality of the painting, such that the appraiser with the least average appraisal error achieves the highest client share [9].

The ART testbed can be easily mapped to the seller selection problem as shown in Table 1. In the remainder of this section, we will use the terms *buyer*, *advisor* and *seller* to denote appraiser, opinion provider and painting respectively.

The specifications of the ART testbed are: 1) client hiring fee 100; 2) opinion transaction cost 10; 3) reputation transaction cost 1; 4) certainty assessment cost 1 and 5) old client share influence 0.1 (to update bank balance). For the SALE

[3]While not realistic, this allows us to focus on the quality of the appraiser without obfuscating the results by additional effects caused by advisors not knowing the ground truth.

Table 1: Seller selection problem using ART testbed

ART testbed	Sellers Selection Problem (SSP)
appraiser	buyer
painting quality	seller quality
appraisal	finding the seller's quality
opinion provider	advisor
opinion transaction	query about seller quality
reputation transaction	query about advisor quality
certainty	advisor's certainty on seller quality
client fee	satisfactory transaction reward
timesteps	no. of rounds of simulation
avg. no. of paintings per agent	avg. no. of SSPs per buyer
appraise as high quality	buy
appraise as low quality	do not buy

POMDP, rewards $R(s, SQ_{ij})$ = opinion cost, $R(s, AQ_{ii'})$ = reputation cost and $R_{sat} = -R_{unsat}$ = client fee. In the experiments, we use symbolic Perseus [8], which exploits the *factored* nature of SALE POMDP as the POMDP solver.

Using the above settings, we compare the performance of SALE POMDP with state-of-the-art trust models BLADE, Personalized, *Advisor* POMDP and TRAVOS, each of which is modeled as a trustworthy buyer in the ART testbed. Also, untrustworthy advisors exhibiting *adversarial*/*random* behavior (as described in Sec. 3.2) are introduced. The evaluation metrics used are: 1) *accuracy* which is the percentage of sellers, whose quality has been correctly identified by the buyer; and 2) *bank balance*, the accumulated reward.

Single Transaction Setting. We first conduct experiments to verify the performance of the SALE POMDP in a single transaction setting. We assume that buyers have no prior experience in the market (which is the case with most real world e-marketplaces), e.g., we initialize the SALE POMDP with a uniform belief that assigns each advisor a 50% probability of being trustworthy. The ART testbed simulation is run for a single round and each agent only receives a single painting to appraise. For each trust model, we perform simulations in which it interacts with 5 advisors (some of which are untrustworthy). The results are shown in Fig. 3, whose x-axes represents the number of untrustworthy advisors (so there are '5-x' trustworthy advisors). Error bars indicate the standard error over 100 iterations.

Fig. 3(a-b) show the results when untrustworthy advisors act in an *adversarial* manner. In Fig. 3(a), accuracy of SALE POMDP decreases with increase in number of *adversarial* advisors, due to the increase in the probability of obtaining incorrect opinion (for a uniform initial belief, the resulting SALE POMDP policy can be interpreted as a smart way of performing a majority vote on

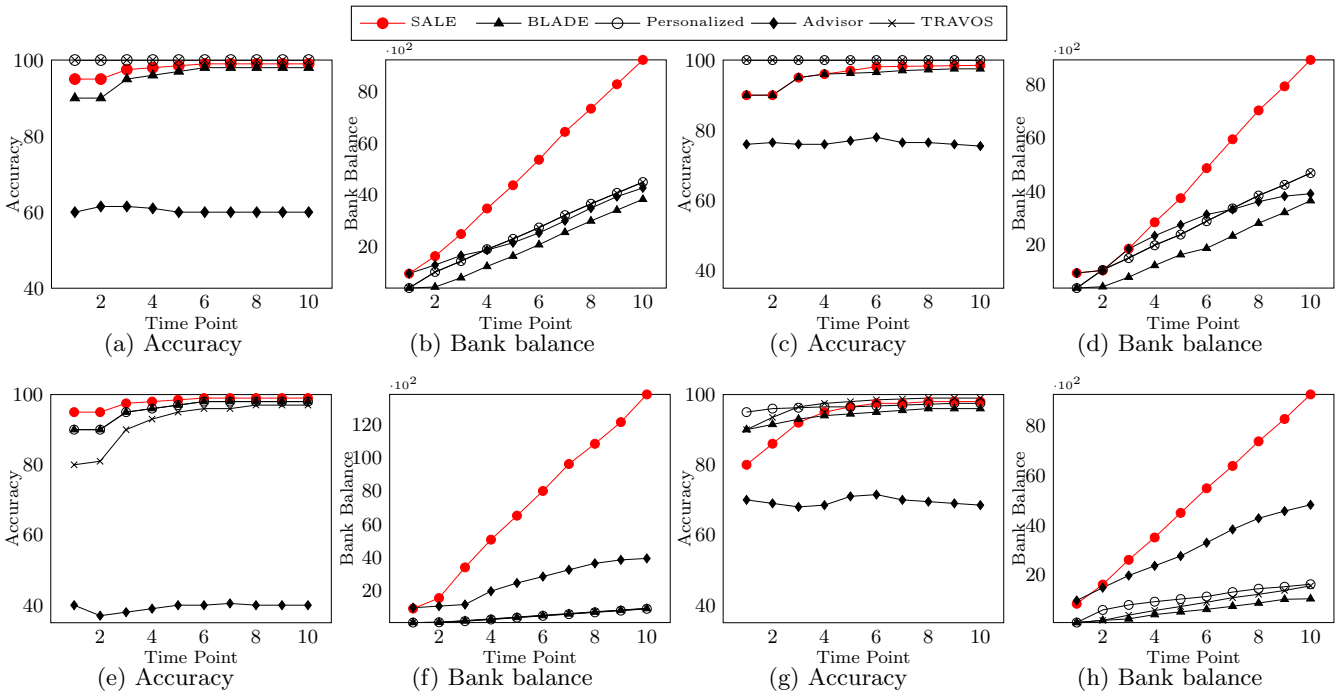


Figure 4: Sequential setting: Majority advisors are trustworthy: (a-b) adversarial scenario (c-d) random scenario; Majority advisors are untrustworthy: (e-f) adversarial scenario (g-h) random scenario

advisor trustworthiness). TRAVOS and Personalized obtain 100% accuracy when majority of advisors are trustworthy, as they use the majority rule to determine seller's quality, when the buyer has no experience in the market. This is also the reason why their performance drastically decreases when majority (> 2) of the advisors are *adversarial*. The *Advisor* POMDP shows better performance when majority of the advisors are *adversarial*, as it considers all advisors equally trustworthy and randomly chooses an advisor to ask opinion (better than relying on the majority in this case). Of course, since SALE POMDP generalizes the *Advisor* POMDP, given an initial belief that reflects that all the advisors are *trustworthy*, it will also find this policy. BLADE mainly relies on previous experience with advisors and obtains 0% accuracy for all cases (not shown in Fig. 3).

Fig. 3(b) shows the bank balance of the trust models after the first transaction as a newcomer (corresponding to their accuracy in Fig. 3(a)). We see that the SALE POMDP obtains a better bank balance than TRAVOS and Personalized. This is because SALE POMDP uses fewer opinion transactions (query advisors about the sellers' quality). The SALE POMDP mainly relies on reputation transactions[4] (query advisors about other advisors' quality) to firstly determine a trustworthy advisor and then queries about the sellers. TRAVOS and Personalized, on the other hand simply query all advisors about the seller's quality, resulting in high opinion costs, reducing their balance. Though the *Advisor* POMDP achieves higher bank balance (due to fewer opinion transactions and better accuracy), its performance is hugely reliable on the single advisor it randomly chooses to ask opinions, which cannot be successful at all times. Also, its performance is lower in the sequential setting (Fig. 4).

Fig. 3(c-d) show the results when untrustworthy advisors exhibit *random* behavior. Generally, all trust models perform better than the *adversarial* scenario, as *random* advisors may provide correct opinions at times. TRAVOS outperforms Personalized as it is able to correctly identify the trustworthiness of advisors by comparing their opinions only on similar sellers. Personalized considers advisors' previous opinions on all sellers, thereby increasing the chance to incorrectly model untrustworthy advisors with *random* behavior. Again, Fig. 3(d) shows that SALE POMDP clearly outperforms the other models in terms of bank balance.

Sequential Setting. We also conduct experiments in a sequential setting, where appraisers are engaged in multiple transactions. The ART testbed simulation is run for 10 rounds and average number of paintings per agent (per round) is 10. Here, we run the different trust models in competition: for a given buyer (say SALE POMDP), all other buyers (BLADE, Personalized, *Advisor* POMDP and TRAVOS) are trustworthy advisors, always providing correct opinions about sellers/other advisors. Untrustworthy advisors (*adversarial*, *random*) are also introduced and compete in determining the correct seller quality (using the BRS model). For the SALE POMDP, each seller selection (painting appraisal) problem (in a given round) is modeled as a separate POMDP. The actual result of each problem (the client being satisfied or not) is used to update the final belief of the SALE POMDP's previous seller selection problem, giving a new initial belief for the next problem (finding the next painting's quality). Fig. 4(a-d) show the results in the sequential setting when trustworthy advisors form the majority (2 untrustworthy advisors are introduced). Fig. 4(e-h) show the results when majority advisors are untrustworthy.

From Fig. 4, we find that accuracy and bank balance of most trust models increase with time, depicting that experience from previous transactions can significantly affect their

[4]The influence of query costs is shown in Fig. 6(a).

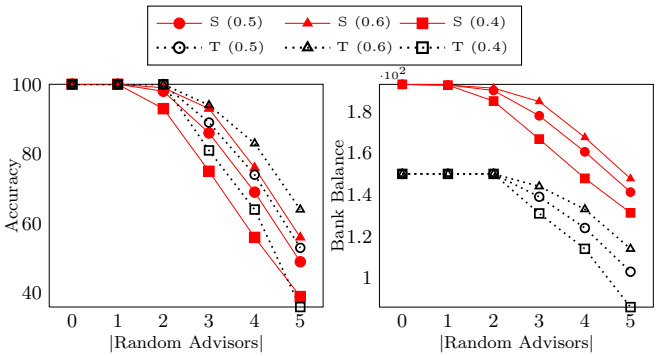
performance. In Fig. 4(a-b), untrustworthy advisors are *adversarial* in nature. Fig. 4(a) shows the SALE POMDP with initial accuracy 95.0%, mainly because of the presence of untrustworthy advisors leading to inaccurate opinions, especially when the buyer has no prior experience in the market. However, its accuracy increases to 99.0%, at the end of simulation, using previous transaction information to identify *adversarial* advisors and refrain from asking opinions. TRAVOS and Personalized obtain a better accuracy of 100% as majority of the advisors are trustworthy. BLADE obtains an initial accuracy of 90.0% and reaches 98.0% by the end of simulation. The *Advisor* POMDP learns about different paintings (sellers), which is in fact not useful in this setting, since every buyer is assigned a different painting to appraise each round. Also, it does not learn about advisor's behavior and considers all advisors to be trustworthy. Therefore, its accuracy depends on the probability of choosing a trustworthy advisor to ask opinion (nearly 60%).

Fig. 4(b) shows the bank balance of trust models corresponding to their accuracy in Fig. 4(a). It clearly shows that the SALE POMDP significantly outperforms its competitors. Though *Advisor* POMDP obtains a lower accuracy, it uses fewer opinion transactions than BLADE, Personalized and TRAVOS, hence obtaining a substantial bank balance.

In Fig. 4(c-d), untrustworthy advisors exhibit *random* behavior. Fig. 4(c) shows that accuracy of the SALE POMDP increases from 90.0% to 98.5%, increasing at a slower rate than in Fig. 4(a). This is because advisors who behave in a *random* manner can give the correct advise (by chance). SALE POMDP, thereby requires more number of transactions to discriminate them from trustworthy advisors on an average. However, this is not the case when untrustworthy advisors are *adversarial* in nature. In fact, in the sequential setting, the performance of SALE POMDP is better when advisors exhibit *adversarial* than *random* behavior. Because, untrustworthy advisors who are *adversarial* in nature, often provide incorrect advise and once identified, they can be more informative than *random* ones. Also, in Fig. 4(c), TRAVOS and Personalized obtain an accuracy of 100% and accuracy of BLADE increases from 90.0% to 97.5%. Again, the accuracy of *Advisor* POMDP depends on the probability of choosing a trustworthy advisor (around 76%). Fig. 4(d) shows that the SALE POMDP obtains a significantly higher bank balance than other trust models.

In Fig. 4(e-f), untrustworthy advisors are *adversarial* in nature and form the majority (5 untrustworthy advisors are introduced with a total of 9 advisors). Fig. 4(e) shows that SALE POMDP follows a similar trend as in Fig. 4(a). But, we find that TRAVOS and Personalized obtain an initial accuracy of 80.0% and 90.0%, respectively when compared to 100% in Fig. 4(a). This is because when the buyer is a newcomer, both approaches rely on the majority rule and since the majority are untrustworthy in this case, their accuracy is less. However, with experience their accuracy increases.

Fig. 4(f) shows that SALE POMDP obtains the highest bank balance. In Fig. 4(g-h) untrustworthy advisors are *random* and form the majority. The initial accuracy of SALE POMDP is 80.0% (as compared to 90.0% in Fig. 4(c)), due to the increase in the number of untrustworthy advisors. The accuracy of the *Advisor* POMDP is also lower than that in Fig. 4(c) (around 70%). SALE POMDP obtains the best bank balance in Fig. 4(h), clearly showing that it performs significantly better also in this setting.



(a) Accuracy (b) Bank balance

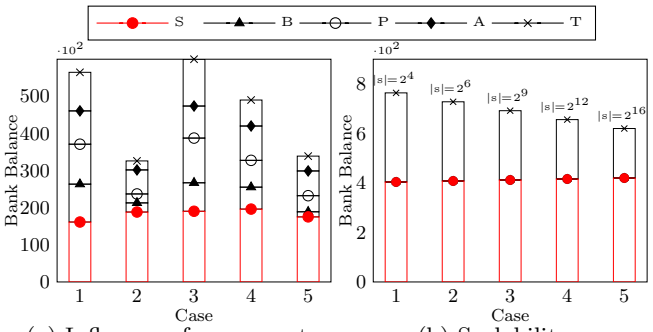
Figure 5: Robustness of SALE POMDP

Robustness. We analyze the robustness of SALE POMDP in the specification of its parameters (e.g., SALE POMDP always models advisors who show *random* behavior to behave with a 50% probability of giving correct opinion). We consider untrustworthy advisors exhibiting various types of *random* behaviors for the experiment, which is conducted in a single transaction setting, where buyers have no prior experience in the market and each agent only receives a single painting to appraise. Three types of advisors showing *random* behavior are considered: 1) advisors who perfectly exhibit random behavior i.e., with a 50% probability of giving correct opinion; 2) advisors inclined to behave honestly i.e., with a 60% probability of giving correct opinion; and 3) advisors inclined to behave in a deceptive manner i.e., with a 40% probability of giving correct opinion. The performance of SALE POMDP against each case of such advisor behaviors is represented using S(SALE) (0.5), S (0.6) and S (0.4) in Fig. 5. We also show the performance of TRAVOS for comparison, denoted by T (0.5), T (0.6) and T (0.4).

Fig. 5(a) shows that even when interacting with advisors that act differently than those assumed in the POMDP model, the performance of SALE POMDP is robust: performance is relatively comparable to TRAVOS (with same amounts of degradation), but in absolute sense it is still much better. Fig. 5(b) also shows that bank balance of S (0.5), S (0.6) and S (0.4) is much higher than TRAVOS.

Influence of Query Costs. Fig. 6(a) shows how the bank balance of trust models (SALE POMDP (S), BLADE (B), Personalized (P), *Advisor* POMDP (A) and TRAVOS (T)) is influenced by different cases of (opinion cost, reputation cost), 1: (1,1); 2: (10,1); 3: (1,10); 4: (5,5); and 5: (10,10), in a sequential setting (ART testbed simulation is run for 10 rounds and average number of paintings per agent is 10), where majority advisors exhibit *random* untrustworthy behavior. We find that SALE POMDP obtains the best bank balance in all cases, demonstrating the use of less query actions (opinion and reputation transactions) in general, when compared to other trust models. Even in the least favorable case, when both opinion and reputation costs are high (10,10), SALE POMDP still accumulates the highest reward. In the most favorable case, i.e., when opinion cost is high and reputation cost is low (10,1) SALE POMDP achieves a higher balance than all other models put together.

Scalability. We also examine whether representing the SALE POMDP in the *factored* form helps to overcome the computational complexity. We use the single transaction



(a) Influence of query costs (b) Scalability

Figure 6: Stacked bars: interval lengths represent values

setting in the ART testbed, where buyers have no prior experience and each agent only receives a single painting to appraise. Fig. 6(b) shows how the performance of SALE POMDP is influenced on scaling to larger problems, by increasing the number of advisors in the market. Case 1 represents the scenario with 2 sellers, 1 trustworthy advisor and 1 untrustworthy advisor exhibiting *random* behavior. Similarly, cases $2-5$ represent the scenario with 2 sellers and $2-5$ trustworthy and untrustworthy advisors each. The state space for each case is also labeled in the corresponding bars of Fig. 6(b). We find that bank balance of SALE POMDP does not degrade as the problems become more complex. This demonstrates that SALE POMDP is able to scale to considerable limits (large seller selection problems, intractable to be solved by non-factored methods such as SARSOP [12]), while still preserving quality. Run times to solve the SALE POMDP vary from 9.6s for the 2 advisor case, to 3170s for the 10 advisor case.

5. CONCLUSION AND FUTURE WORK

The paper suggests a novel method for dealing with the seller selection problem using POMDPs. SALE POMDP optimally selects the right sellers as transaction partners, balancing the trade-off between information gaining and information exploiting actions. In addition to querying advisors about sellers, the model also allows to selectively query advisors about the trustworthiness of other advisors, which is a novel feature the approach offers. We also represent SALE POMDP in its *factored* form to allow solving POMDPs with large number of states. Experiments using the ART testbed verify that SALE POMDP outperforms state-of-the-art trust models in optimally selecting quality sellers. Experiments also demonstrate its robustness in the specification of its parameters. We also show that the factored form helps to scale to reasonably large seller selection problems.

The presented research opens up many directions of future work. While we established that SALE POMDP is robust against the choice of parameters for the transition and observation model, an interesting direction is to automatically optimize these (e.g., using evolutionary optimization techniques). Alternatively, it is possible to improve robustness of the model to different types of attacks, by including more detailed advisor models (e.g., differentiating its trustworthiness in providing opinions about sellers and other advisors). Finally, an important direction of future research is to develop dedicated solution methods that further exploit the structure of SALE POMDP to provide further scalability.

6. ACKNOWLEDGMENTS

We thank Ashwini Gokhale for her contribution to the foundations of this work. This work is supported by MoE AcRF Tier 2 Grant M4020110.020, the Institute for Media Innovation at Nanyang Technological University and in part by NWO Innovational Research Incentives Scheme Veni #639.021.336.

7. REFERENCES

[1] Whitby, A., Jøsang, A., Indulska, J.: Filtering out unfair ratings in bayesian reputation systems. In AAMAS TRUST Workshop, 2004.

[2] Teacy, W.T.L., Patel, J., Jennings, N.R., Luck, M.: Coping with inaccurate reputation sources: Experimental analysis of a probabilistic trust model. In AAMAS, 2005.

[3] Zhang, Jie, Cohen, R.: Evaluating the trustworthiness of advice about selling agents in e-marketplaces. ECRA, 7(3):330–340, 2008.

[4] Regan, K., Poupart, P., Cohen, R.: Bayesian reputation modeling in e-marketplaces sensitive to subjectivity, deception and change. In AAAI, 2006.

[5] Irissappane, Athirai A., Jiang, Siwei, Zhang, Jie: A framework to choose trust models for different e-marketplace environments. In IJCAI, 2013.

[6] Regan, K., Poupart, P.: The Advisor-POMDP: A principled approach to trust through reputation in electronic markets. In PST, 2005.

[7] Kaelbling, L.P., Littman, M.L., Cassandra, A.R.: Planning and acting in partially observable stochastic domains. Artificial Intelligence, 101(1-2):99–134, 1998.

[8] Poupart, P.: Exploiting structure to efficiently solve large scale Partially Observable Markov Decision Processes. Ph.D. thesis, Department of Computer Science, University of Toronto, 2005.

[9] Fullam, K.K., Klos, T.B., Muller, G., Sabater, J., Schlosser, A., Topol, Z., Barber, K.S., Rosenschein, J.S., Vercouter, L., Voss, M.: A specification of the Agent Reputation and Trust (ART) testbed: Experimentation and competition for trust in agent societies. In AAMAS, 2005.

[10] Papadimitriou, C.H., Tsitsiklis, J.N.: The complexity of Markov decision processes. Mathematics of Operations Research, 12(3):441–451, 1987.

[11] Silver, D., Veness, J.: Monte-carlo planning in large POMDPs. Advances in Neural Information Processing Systems, 23:2164–2172, 2010.

[12] Kurniawati, H., Hsu, D., Lee, W.S.: SARSOP: Efficient point-based POMDP planning by approximating optimally reachable belief spaces. Robotics: Science & Systems, 65–72, 2008.

[13] Spaan, Matthijs T.J., Vlassis, Nikos,: Randomized point-based value iteration for POMDPs. JAIR, 24:195–220, 2005.

[14] Porta, Josep M., Vlassis, Nikos, Spaan, Matthijs T.J., Poupart, P.: Point-based value iteration for continuous POMDPs. JMLR, 7:2329–2367, 2006.

[15] Oliehoek, Frans A., Gokhale, Ashwini, Zhang, Jie: Reasoning about advisors for seller selection in e-marketplaces via POMDPs. In AAMAS TRUST Workshop, 2012.