

A POMDP Based Approach to Optimally Select Sellers in E-Marketplaces

Paper XXX

ABSTRACT

In e-marketplaces, buying agents choose among multiple sellers, finding the best one to perform a transaction. A seller's quality is mainly determined using buyers' assessments on the seller based on previous transactions. How a buyer utilizes existing information and chooses the right seller, while maximizing its utility derived from the decision process is crucial, since the value derived from a successful transaction with the seller, may be negligible relative to the cost of choosing the right seller. This paper presents SALE POMDP, a framework based on the *Partially Observable Markov Decision Process*, which helps a buyer to optimally select the right seller to perform a transaction. The framework is robust in the sense that it also allows to reason about the quality of other buying agents (advisors) in the market, who provide opinions about the sellers. Experimental results on the ART Testbed demonstrate that SALE POMDP can optimally select the right sellers and advisors (to ask opinions), effectively balancing the trade-off between the cost of obtaining and benefit of more information, than traditional trust models.

Categories and Subject Descriptors

I.2.11 [Distributed Artificial Intelligence]: Intelligent Agents; Multiagent Systems

General Terms

Performance; Design

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 the quality of sellers and determine which sellers to do business with. Buyers maintain beliefs over the quality levels of the sellers, based on their previous transactions, using which they may select the best seller to interact with. 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 their beliefs on the sellers' quality.

A number of trust models (*e.g.* BRS [14], TRAVOS [13], Personalized [15], BLADE [10], 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. However, the above approaches mainly focus on accurately estimating the quality of the seller rather than optimally choosing a right seller to perform transaction with. This is because the above trust models fail to reason *when* it is necessary to query advisors, though they may determine *whom* to query by analyzing the quality levels (trustworthiness) of the advisors. Most approaches simply query all advisors who have interacted with the seller. This will in fact result in negligible utility for the buyer, especially when the querying advisors are large in numbers, since the cost of querying advisors may be greater than the value derived from a successful transaction with the seller.

Regan *et al.* [9] proposed the *Advisor* POMDP, which considers the seller selection problem as a *Partially Observable Markov Decision Process*. POMDPs provide a framework for making optimal decisions in partially observable environments [4]. The main advantage of using POMDPs is that rather than trying to achieve the most accurate estimate of sellers, the approach tries to select good sellers and does that optimally. However, *Advisor* POMDP did not reason about advisors' quality levels and considered all advisors to be trustworthy. Also, in *Advisor* POMDP, each advisor provides opinions about its beliefs on all sellers, when queried, resulting 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.

In this paper, we present SALE ((*S*)eller & (*A*)dvisor *se(LE)ction*) POMDP [6], a framework to deal with the seller selection problem in an effective manner. Here, the buying agent is modeled as a POMDP, with the main goal of optimally selecting the right seller to perform a transaction. In this process, it maintains beliefs over the quality levels of sellers, as well as that of advisors. To query advisors about the quality of the sellers, it firstly determines the trustworthiness of the advisors, thereby limiting its interactions to the honest ones. It also makes provision to query advisors about their beliefs on other advisors, sufficiently using it to improve its own beliefs and make optimal decisions, especially in sophisticated attacking scenarios. It can also account for the buyers' experience in the market by considering the influence of previous responses (history) in the POMDP formulation. SALE POMDP can also be modeled as a *factored* POMDP to resolve the scalability issues, which is a limiting factor in solving most POMDPs [8].

In this paper, we conducted a thorough evaluation of SALE POMDP using the ART testbed to prove the effectiveness of SALE POMDP against traditional trust models in optimally selecting sellers, even in the presence of strategic attacks.

The rest of the paper is organized as follows. Section 2 gives a brief discussion of existing trust models and explains *Advisor* POMDP in detail. The detailed description of the SALE POMDP model is present in Section 3, which also discusses the *factored* nature of SALE POMDP. In Section 4, we conduct an extensive experimentation of SALE POMDP using the ART testbed, and demonstrate the effectiveness of SALE POMDP in optimally selecting the right sellers to perform transaction with. The scalability of the *factored* approach is also analyzed. Finally, Section 5 concludes the current work and proposes future work.

2. BACKGROUND

2.1 Trust Models

Many trust models have been proposed in literature to address the seller selection problem. Reputation System (BRS) [14] models the seller's reputation as the expected value of the beta probability distribution of the trusted advisors' beliefs about the seller. Teacy *et al.* [13] propose TRAVOS which is also based on beta probability distribution. It integrates a buyer's own beliefs about sellers as well as the beliefs of advisors to determine seller's quality. The personalized approach [16], models seller's quality as a weighted score of its public and private reputation which is based on the buyer's own experience and that of the advisors respectively. The BLADE approach of Regan *et al.* [10] allows a buyer to learn other advisors' evaluation functions on different features of the services delivered by sellers, by which it can reinterpret advisors' beliefs instead of filtering them. However, all the above trust models do not provide optimal decision making for the buyer on whether and from which seller to perform transaction, which is exactly what our approach tries to offer.

2.2 Advisor POMDP

Regan *et al.* [9] introduced the Advisor POMDP, an approach for dealing with the seller selection problem based on the POMDP framework. Partially observable Markov decision processes (POMDPs) provide a natural model for sequential decision making under uncertainty. A POMDP agent works towards the main goal of maximizing the expected discounted future reward [4] (buyers' utility in this case). Formally, an Advisor POMDP consists of the following elements: If there are I advisors that can be queried about the reputation of all J sellers then,

- \mathcal{S} —a set of possible states of the environment. A state $s = \langle \vec{q}, sat \rangle$, where $\vec{q} \in [0, 1]^J$ is a vector indicating the quality q_j of each seller and $sat \in \{-1, 0, +1\}$ indicates the result of a purchase: satisfactory (+1), unsatisfactory (-1) or whether no purchase took place yet (0).
- \mathcal{A} —a set of actions. There is one action ask_i for each advisor i , and one buy_j action for each seller j .
- T —a transition function that specifies $\Pr(s'|s, a)$, the probability of transferring to a state s' given that action a was taken in state s . For ask_i actions, the state does not change. For buy_j action a state $s = \langle \vec{q}, 0 \rangle$

changes stochastically to $s' = \langle \vec{q}, -1 \rangle$ or $s = \langle \vec{q}, +1 \rangle$ with probabilities depending on q_j .

- R —a reward function specifying $R(s, a, s')$. For ask actions, a small cost is paid. For transitions to a satisfied state (i.e., from $sat = 0$ to $sat = +1$) a reward is received, while transitions to an unsatisfied state yield a large penalty. Once the state has changed to satisfied or unsatisfied, no further rewards are given.
- Ω —a set of observations o . In the advisor POMDP, the advisors respond with a tuple $o = \langle rep_j, cf_j \rangle_{j=1}^J$ that expresses the knowledge of that advisor about all sellers. Here rep_j is the reputation according to the advisor and cf_j is a measure of its certainty.
- O —the observation function that specifies $\Pr(o|a, s)$. Since the semantics of the certainty factors are not formalized, there is some freedom in its specification.
- b^0 —the initial state distribution.
- h —the *horizon* of the problem. That is the number of time steps, or *stages*, for which we want to plan. h can also be infinite.

However, *Advisor* POMDP does not consider advisor's quality to be a part of the state space, making it vulnerable to unfair rating attacks [3], where advisors may provide incorrect opinions about sellers. Also observations in *Advisor* POMDP contain beliefs about all sellers, resulting in an unnecessary exchange of information during the *ask* actions.

3. THE SALE POMDP MODEL

3.1 Basic SALE POMDP

In this section, we introduce the SALE POMDP model for dealing with the seller selection problem by optimally selecting the right seller to perform transaction (rather than accurately determining the quality of the sellers). Unlike *Advisor* POMDP, SALE POMDP assumes that advisors also have quality levels (*trustworthiness*) and considers it as part of the state space. It also allows agents to query about the quality of other advisors in the system. Each SALE POMDP agent can be described in terms of states, actions, observations and rewards such that:

- \mathcal{S} —a set of possible states of the environment. A state contains the quality levels¹ of each seller, advisor and the status of the current transaction. 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 indicates the status of the current transaction, which can be *not_started*, *satisfactory*, *unsatisfactory*, *gave_up* or *finished*. 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 decision process begins when $sat = not_started$. The end of the decision process is modeled using sets of terminal states (such as *satisfactory*, *unsatisfactory*, *gave_up*) i.e. buy_j actions may result in a successful ($sat = satisfactory$)

¹we assume discrete quality levels, in order to use standard POMDP solvers.

or unsuccessful ($sat = unsatisfactory$) transaction and do_not_buy action will result in $sat = gave_up$. Any transitions from the terminal states will result in $sat = finished$.

- \mathcal{A} —a set of actions which can be:
 - $seller_query_{ij}$, ask advisor i about seller j ,
 - $advisor_query_{ii'}$, ask advisor i about advisor i' ,
 - buy_j , buy from seller j .
 - do_not_buy , decide not to buy from any seller.
- T —a transition function that specifies $\Pr(s'|s, a)$, the probability of transferring to a state s' given that action a was taken in state s . We assume that when taking a query action, the state does not change:

$$\forall_{i,j} \quad \Pr(s'|s, seller_query_{ij}) = \delta_{ss'} \quad (1)$$

$$\forall_{i,i'} \quad \Pr(s'|s, advisor_query_{ii'}) = \delta_{ss'} \quad (2)$$

$\delta_{ss'}$ is the Kronecker delta *i.e.* 1 if and only if $s = s'$. When taking a buy_j action, the state will always transition to a terminal state. The transition probabilities to terminal states give a definition of the quality levels. In general, chances of transition to *satisfied* should be higher when buying from higher quality sellers. Together, the specifications of these transitions imply the assumption that quality and trust-levels are stationary for the duration of the decision process.

- R —a reward function specifying $R(s, a, s')$: a small cost is associated with ask action $R(s, seller_query_{ij}) = R(s, advisor_query_{ii'}) = R_{ask}$; a reward is associated with a good purchase $R(s, buy_j, s' = \langle \vec{q}, \vec{u}, sat = +1 \rangle) = R_{sat}$; a penalty is levied when the outcome of a transaction is *unsatisfied* $R(s, buy_j, s' = \langle \vec{q}, \vec{u}, sat = -1 \rangle) = R_{unsat}$ and for taking the do_not_buy action when in fact there is a seller of high enough quality, otherwise there is a small reward for this action. Once the transaction state is *satisfied*, *unsatisfied* or *gaveup*, no further rewards are given.
- Ω —a set of observations. When a *query* action is performed the agent will receive an observation from the set of discriminated quality levels. That is, after a $seller_query_{ij}$ action, the agent receives an observation $o \in \{good, bad\}$ corresponding to the quality of seller j , while after an $advisor_query_{ii'}$ it will get an observation $o \in \{trustworthy, untrustworthy\}$ corresponding to the quality of advisor i' . When the agent transitions to a terminal state, it receives the observation *ended*.
- O —the observation function that specifies $\Pr(o|a, s)$. As in the Advisor POMDP, there is no a priori correct way to specify the observation probabilities. The probabilities picked 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.
- b^0 —the initial state distribution. It is dependent on the subjective beliefs of the agent, when the need for purchasing an item 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. That is, once the buy_j or do_not_buy 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 and 2) actual experiences with sellers: even though in the decision making task we model a transition to a terminal state with a deterministic *ended* observation, the actual order 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 task²

- h —the *horizon* of the problem. That is 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 SALE POMDP agent interacts with the environment, it maintains a *belief* 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:

$$b'(s') = \frac{\Pr(s', o|b, a)}{\Pr(o|b, a)} = \frac{1}{\Pr(o|b, a)} \Pr(o|a, s') \sum_s \Pr(s'|s, a)b(s) \quad (3)$$

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

These beliefs are the basis for decision making: a policy π maps beliefs to actions $\pi(b) = a$. The goal of solving the POMDP is to find an optimal policy that maximizes the expected discounted cumulative reward, also called *value*:

$$V(\pi) = \mathbb{E} \left[\sum_{t=0}^{h-1} \gamma^t R(s, a, s') \mid \pi, b^0 \right], \quad (4)$$

with $0 \leq \gamma < 1$ the discount factor.

Finding an optimal policy π^* is intractable in general (PSPACE complete [7]), however, in recent years substantial advances have been made in approximate solutions (e.g., [5, 12]). Since the SALE POMDP really is a *factored* POMDP, we try and use solvers such as symbolic Perseus [8] that exploit this property, as discussed in the next section.

3.2 Factored Representation

POMDPs with very large state spaces are impractical to be solved by the classic solution algorithms. However, 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) in the form of decision-trees or graphs are often used to represent these effects compactly referred to as a *factored* representation [11]. Solvers such as Symbolic Perseus [8] exploit this *factored* structure to resolve scalability issues in POMDPs.

To illustrate the *factored* nature of the SALE POMDP, we will consider a simple case of the seller selection problem

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

with 1 seller and 2 advisors such that $\mathcal{Q} = \{L, H\}$, representing *L*(ow) and *H*(igh) seller quality and $\mathcal{U} = \{T, U, A\}$, representing *T*(rustworthy), *U*(ntrustworthy) and *A*(dversarial) advisor quality levels, allowing to accommodate attacking behaviors: 1) *random*, where advisors provide random opinions; 2) *adversarial*, where advisors provide complimentary opinions. Each state $s \in \mathcal{S}$ is a compilation of the quality of each seller, quality of each advisor and transaction outcome, i.e. $s = \langle \langle q_1 \rangle, \langle u_1, u_2 \rangle, sat \rangle$. To factor the SALE POMDP model, we decompose the transition and observation functions as follows.

Buy Action The DBN and CPT for the buy_1 action (buy from s_1) are shown in Fig 1. The probability of transferring to a 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 shown in Eqn. 5. $\Pr(sat' | q'_1, sat, buy_1)$ signifies that the outcome of a transaction sat' at a time step depends on the previous status of the transaction sat , the true quality of the seller s'_1 and the buy_1 action. The CPT for sat' variable is framed assuming that buying from a high quality seller will lead to *satisfied* with 80% probability: $\Pr(sat = -1 | \langle sat = 0, q_j = L \rangle, buy_j) = 0.8$. Similarly, we will assume that a low quality seller will lead to *unsatisfied* with probability 0.8. For the *do_not_buy* action, the outcome sat' is not dependant on the seller quality q'_1 and $\Pr(sat' | sat, do_not_buy) = 1$.

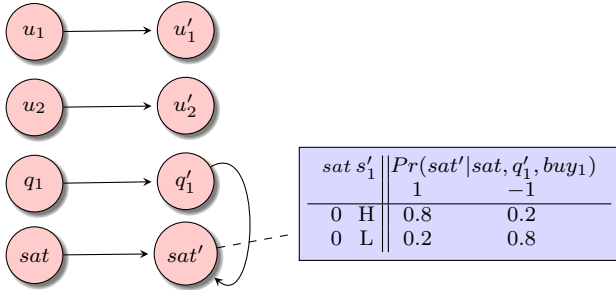


Figure 1: DBN for buy_1

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

Seller Query Fig 2 shows the DBN and CPT for the action $seller_query_{11}(sq_{11})$ (query a_1 about s_1). Eqn 6 shows that observing the quality of a seller s_1 through o' is conditionally dependant on the current quality of the seller q'_1 (after the $seller_query_{11}$ action) and the quality of the advisor u'_1 , who provides opinion. 2 possible observations are considered: *good* (i.e., the advisor says that the seller is high quality) denoted by g and *bad* (the seller is said to be low quality) denoted by b . The observation probabilities for the $seller_query_{ij}$ action are such that asking a trustworthy (T) advisor gives more accurate observations, while untrustworthy ones may provide a random (U) opinion or they can be adversarial (A) in nature. We will obtain a similar figure for $seller_query_{21}$ action.

$$\begin{aligned} \Pr(o' | s, sq_{11}) &= \Pr(o' | u'_1, u'_2, q'_1, sat', sq_{11}) \\ &= \Pr(o' | u'_1, q'_1, sq_{11}) \end{aligned} \quad (6)$$

Advisor Query Fig 3 shows the DBN and CPT for the action $advisor_query_{12}(aq_{12})$ (query a_1 about a_2). Eqn 7

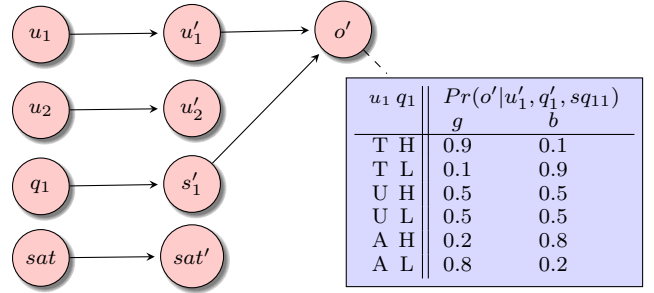


Figure 2: DBN for $seller_query_{11}$

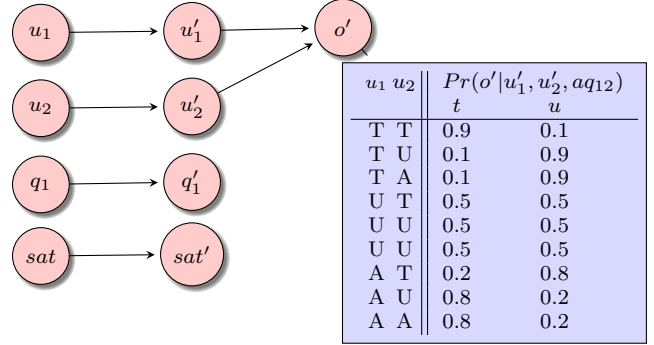


Figure 3: DBN for $advisor_query_{12}$

show that observing the quality of advisor a_2 using the variable o' (we use the same observation variable for simplicity) is conditionally dependant on the current quality of the advisor u'_2 (after the $advisor_query_{12}$ action) and the quality of the advisor a_1 (u'_1), who provides opinion. We will obtain a similar figure for $advisor_query_{21}$ action.

$$\begin{aligned} \Pr(o' | s, aq_{12}) &= \Pr(o' | u'_1, u'_2, s'_1, sat', aq_{12}) \\ &= \Pr(o' | u'_1, u'_2, aq_{12}) \end{aligned} \quad (7)$$

The conditional probability tables in the above figures can be further compressed by exploiting context-specific independence which refers to the fact that some random variables are independent of each other in some contexts (i.e., for some specific values of other random variables). In general, context-specific independence allows one to compactly represent probability tables using decision trees (DTs) or algebraic decision diagrams (ADDs) [1]. We will show one such ADD for the CPT in Fig 2.

Fig. 4 illustrates the ADD representations of the CPT for the observation variable o' in Fig 2. The value of o' for some trust assignment for variables u'_1, q'_1 can be found by following the corresponding branch in the ADD. Note that when u'_1 is *adversarial* or *trustworthy* in nature and seller quality q'_1 is *low*, o' will result in the exact compliment of the values obtained when traversed through the path when q'_1 is *high*. This can be compactly represented, as shown using the blue dashed lines in Fig. 4. Also, when the advisor behaves in a random manner u'_1 is *untrustworthy*, o' is independent of q'_1 and results in a value of 0.5 whether q'_1 is of *high* or *low* quality. Hence, by using the notion of context-specific independence, the ADD in Fig. 4 can encode with

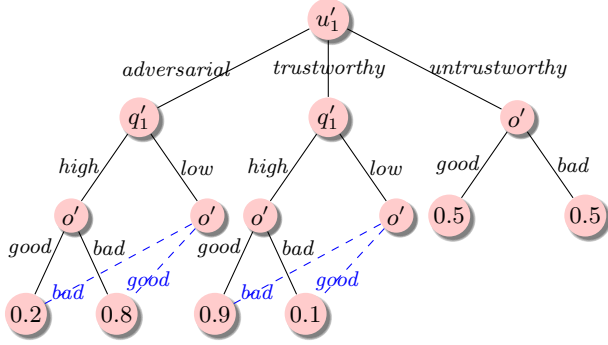


Figure 4: ADD for *seller_query11*

only 6 leaves the 12 values of the CPT in Fig 2. Thus we show how SALE POMDP exhibits conditional-independence and context-specific independence, which can be exploited by solvers such as Symbolic Perseus to find the optimal policy, especially for POMDPs with a large number of states.

3.3 Belief Update in SALE POMDP

The belief updates are performed such that they correlate the state factors in meaningful ways. For instance, observing *good* after *seller_query_{ij}* should give more weights to states where the seller is *high* quality $q_j = H$ and the advisor is *trustworthy* $u_i = T$, and less weights to states where the seller is low quality $q_j = L$ and the advisor is *untrustworthy* or *adversarial* $u_i = U/A$. Similarly, observing *T* after *advisor_query_{ii'}* should put more weight on states where $u'_i = T$ and $u_i = T$, and decrease weight on states where $u'_i = U/A$ and $u_i = T$.

Fig. 5 shows the partial SALE POMDP policy for the 1 seller, 2 advisors case. The transition and observation probabilities are same as those in Sec. 3.2. As mentioned in Sec. 3.2, each state $s = \langle \langle q_1 \rangle, \langle u_1, u_2 \rangle, sat \rangle$. The beliefs corresponding to each action (represented by nodes) are shown using tables associated with the nodes. The beliefs in Fig. 5 are presented in their factored form. AQ_{12} represents *advisor_query₁₂* and SQ_{11} represents *seller_query₁₁*. We have not shown the belief update for the variable *sat* in the tables for simplicity.

Initially (before action AQ_{12}), the beliefs for advisors u_1 and u_2 are 0.5 *trustworthy* and 0.5 *adversarial* (we consider the attacking behavior to be *adversarial*) and those for seller q_1 is 0.5 *high* and 0.5 *low*. When the agent receives an observation $o' = T$ after taking action AQ_{12} (traversing through the left child of the root of the tree), the beliefs for u_2 is updated to $\{T = 0.55; A = 0.45\}$. Similarly the beliefs for u_1 is updated (after multiple AQ_{21} actions), resulting in $\{T = 0.61; A = 0.38\}$ for both u_1 and u_2 . Advisors u_1 is determined to be *trustworthy* and thereby queried about the seller (action SQ_{11}). When the agent receives an observation $o' = G(\text{good})$, the beliefs for the seller are updated and thereby *BUY* (buy_1) action is taken. Similarly, if the agent receives an observation $o' = B(\text{bad})$, the *DNB* (*do_not_buy*) action is taken. The beliefs obtained when the observation of AQ_{12} action, $o' = U$ can be seen by traversing through the right child of the root.

4. EVALUATION

Detailed experiments are conducted to: 1) determine the

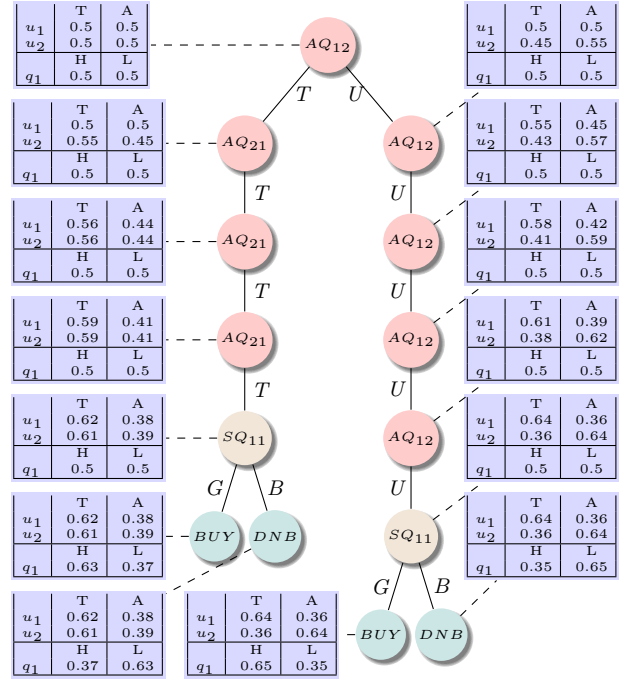


Figure 5: (Partial) SALE POMDP Policy

efficiency of SALE POMDP in optimally choosing the right seller, in comparison with the other trust models and 2) analyze the scalability of SALE POMDP.

4.1 Comparison with Other Trust Models

We evaluate the SALE POMDP model using the Agent Reputation and Trust (ART) testbed [2]. In the ART testbed, agents function as painting appraisers with varying levels of expertise in different artistic eras. Clients request appraisals for paintings from different eras; if an appraising agent does not have the expertise to complete the appraisal, it can request opinions from other appraiser agents (at a specified cost). Appraisers receive more clients, and thus more profit, for producing more accurate appraisals. The appraiser send the received opinions along with the weights (trustworthiness) of opinion providers to the simulation engine, which calculates the final value of the painting.

To model the seller selection problem in a more effective way, we modify the ART testbed such that, appraisers receive more clients, and thus more profit for correctly identifying the quality of a painting (high/low), rather than accurately determining the value of the painting. Also, appraisers (irrespective of their level of expertise) can request opinions from other appraisers, about the appraisal and reputation of other appraisers. In the modified version, the appraisers are allowed to calculate the final appraisal of the painting, making the ART testbed decentralized. Modifying ART testbed in the above manner, makes it ideal to depict the seller selection problem, where the winning appraiser is assessed based on his bank balance in the long run.

The specification of the ART testbed used for experimentation are: 1) client fee of 100; 2) certainty cost 1; 3) opinion cost 10; 4) reputation cost 1 and 5) Old Client Share Influence 0.1. Using the above settings, we compare the performance of SALE POMDP with state-of-the-art trust models *Advisor* POMDP, TRAVOS, Personalized and BLADE, each

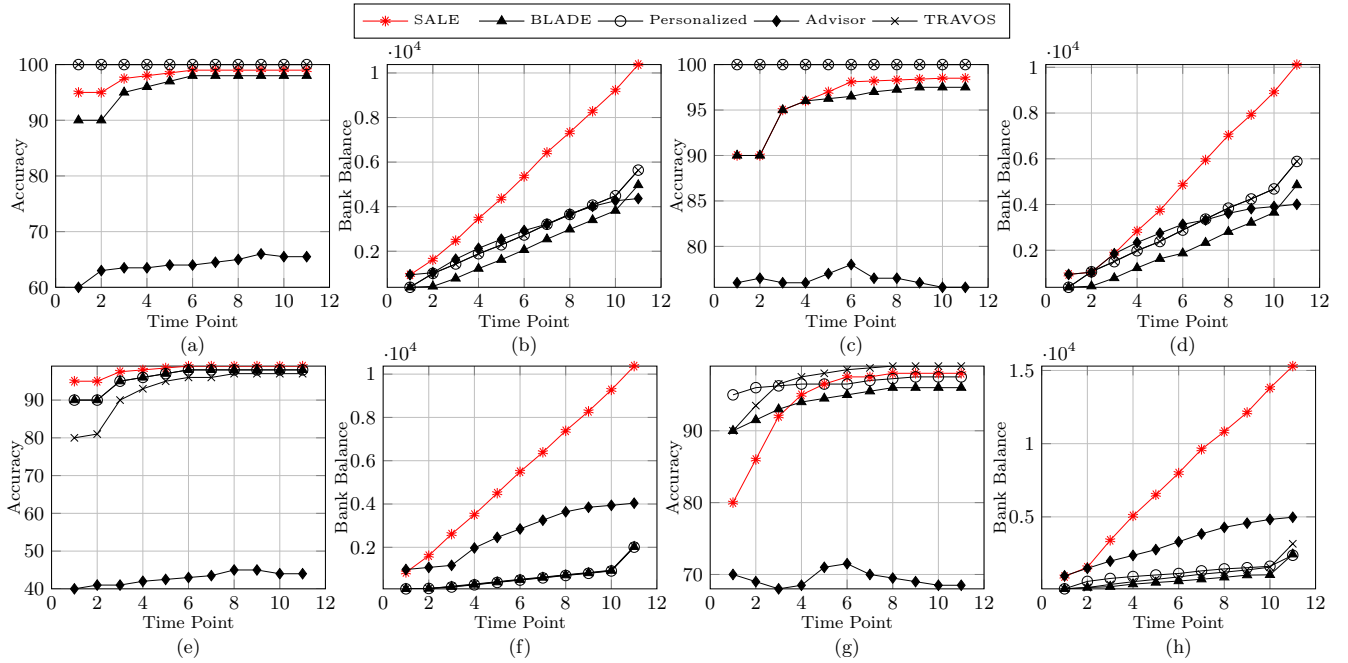


Figure 6: Sequential Setting

of which is modeled as an appraiser in the ART testbed. The trust models are assumed to behave honestly all the time, providing accurate opinions. Opinions refer to the quality of paintings (high/low) and reputation $[0 - 1]$ of other appraisers. Two attacking behaviors are considered: 1) random, where the appraisers provide opinion about the painting/other appraisers in a random manner; 2) adversarial, where the advisors provide the exact compliment of the true quality of the painting/other appraisers. The attackers are appraisers following the BRS model.

For SALE POMDP, the rewards are: $R(s, seller_query_{ij}) = 10$, $R(s, advisor_query_{ii}) = 1$, $R_{sat} = 100$, $R_{unsat} = -100$ and the penalty for taking the *do_not_buy* action when in fact there is a seller of high enough quality is -100 , otherwise the reward is 100. In the experiments, we use Symbolic Perseus [8], a state-of-the-art POMDP solver, which reads problems in a standardized POMDP description format. Symbolic Perseus exploits the *factored* nature of SALE POMDP and helps to achieve better scalability.

The evaluation metrics used are: 1) *accuracy* which is the percentage of appraisals, for which the quality of painting has been correctly identified by the appraiser (trust model); 2) *bank balance*, representing the profit gained by identifying the correct quality of paintings (in an optimal way).

We first conduct experiments in a sequential setting where appraisers are engaged in multiple transactions. The ART testbed simulation is run for 10 timesteps and the average number of clients per appraiser is 10. In SALE POMDP, each painting appraisal problem is modeled as a separate POMDP and the modified belief state resulting from advice in a previous problem is retained and is used as the initial belief for solving the next painting appraisal problem i ;e (finding the quality of the next painting).

Fig 6 shows the results in the sequential setting. From Fig 6, we can find that accuracy and bank balance of most trust models increase with time, depicting that experience from previous transactions can significantly affect the per-

formance of the trust models. Fig.6 (a-d) show the case when trustworthy appraisers form the majority in the market (2 untrustworthy appraisers are introduced) and Fig.6 (e-h) show that case when attackers form the majority (5 attackers are introduced).

In Fig.6 (a-b), the attackers are *adversarial* in nature. From Fig.6 (a), we find that accuracy of SALE POMDP is 95.0% in the beginning. This is because SALE POMDP, at times cannot accurately determine the quality of paintings, in the presence of attackers who provide inaccurate opinions and when the appraiser has no prior experience in the market. However, its accuracy increases to 99.0%, at the end of simulation, as it uses previous transaction information to identify *adversarial* appraisers and refrain from asking opinions. TRAVOS and Personalized obtain a better accuracy of 100% throughout the simulation. This is because the majority of the advisors (other appraisers) are trustworthy. TRAVOS and Personalized use the majority rule to determine the quality of the painting, when the appraiser has no experience in the market and they also incorporate the usage of history information, which helps to determine the quality of painting in the subsequent timesteps correctly. BLADE obtains an initial accuracy of 90.0%, as it cannot appropriately determine the quality of painting when the appraiser is a new comer, since it mainly relies on previous experience with advisors to reinterpret their opinions. However, its accuracy increases with time due to the increase in number of transactions. *Advisor* POMDP considers all appraisers to be trustworthy and hence its accuracy depends on the probability of choosing a trustworthy appraiser to ask opinion, thereby obtaining a low accuracy of 65.5%.

Fig.6 (b) shows the bank balance of the appraisers corresponding to Fig.6 (a). We find that SALE POMDP obtains the highest bank balance. This is because SALE POMDP uses fewer opinion transactions (request opinion about paintings from other appraisers), the costs of which is 10. SALE POMDP mainly relies on reputation transactions (request

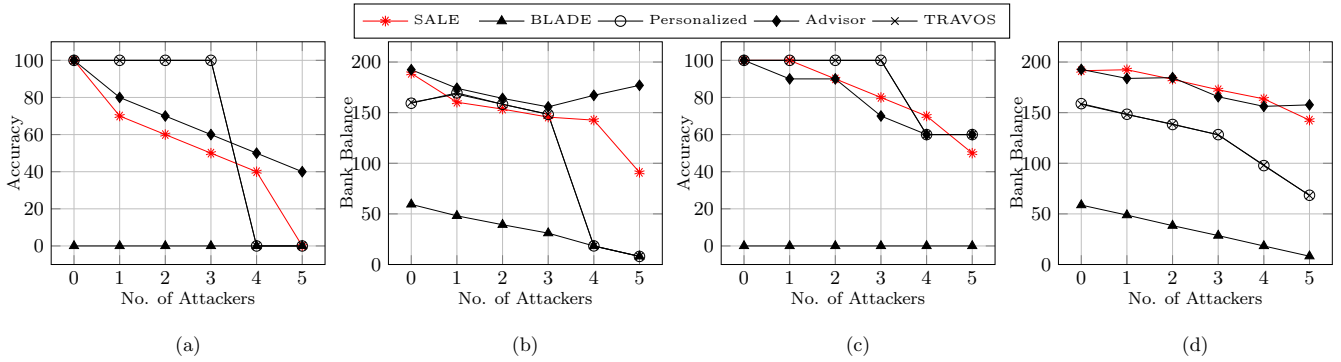


Figure 7: Single Transaction Setting

reputation about appraisers from other appraisers) with a cost of 1, to determine a trustworthy appraiser and then request opinions about the paintings. TRAVOS, Personalized and BLADE, on the other hand simply query all appraisers about the quality of the paintings resulting in a huge cost, when compared to the profit gained by accurately determining the quality of the painting. Also, TRAVOS and Personalized do not query about the reputation of other appraisers to determine their trustworthiness, which is a unique feature offered by the SALE POMDP model. Though *Advisor* POMDP uses fewer opinion transactions, its bank balance is lower than SALE POMDP due to its lower accuracy.

In Fig.6 (c-d), the attackers are *random* in nature. Fig.6 (c) shows that accuracy of SALE POMDP increases from 90.0% to 98.5%. However, from Fig.6 (c), we can find that accuracy of SALE POMDP increases at a slower rate than in Fig.6 (a). This is because the attackers are *random* in nature, whereby they may also provide truthful opinions at times. SALE POMDP assumes consistent advisor behavior and hence requires more number of transactions to learn about the trustworthiness of the *random* attackers than when they behave in an *adversarial* manner, always providing incorrect opinions. TRAVOS and Personalized obtain an accuracy of 100% thought the simulation and accuracy of BLADE increases from 90.0% to 97.5%. The accuracy of *Advisor* POMDP increases, then decreases. Such behavior is hugely attributed to the random nature of the advisor who provide opinions. Again SALE POMDP obtains a higher bank balance than others as shown in Fig.6 (d).

In Fig.6 (e-f), the attackers are *adversarial* in nature and form the majority. Fig.6 (e) shows that SALE POMDP follows a similar trend as in Fig.6 (a). But, we find that TRAVOS and Personalized obtain an initial accuracy of 90.0% and 80.0% respectively when compared to 100% in Fig.6 (a). This is because when the appraiser is newcomer, both approaches rely on the majority rule and since the majority appraisers are attackers in this case, their accuracy is less. However, with experience their accuracy increases. Again, Fig.6 (f) shows that SALE POMDP obtains a higher bank balance than others.

Fig.6 (g), (h) show the case when attackers are *random* in nature and form the majority. The initial accuracy of SALE POMDP is 80.0% (as compared to 90.0% in Fig.6 (c)), due to the increase in the number of attackers. Also TRAVOS outperforms Personalized as it is able to correctly identify the trustworthiness of the appraisers by comparing their opinions on paintings with similar value. Personalized considers all previous opinions from appraisers, thereby in-

creasing the chance to incorrectly model the trustworthiness of the attackers due to their random behavior. The accuracy of the *Advisor* POMDP is also lower than the case where the trustworthy appraisers form the majority.

We also conduct experiments to verify the performance of SALE POMDP in a single transaction setting. Here all appraisers are new comers to the market and have no prior experience (which is the case with most real world e-market places). The simulation is run only for a single time step and average number of clients per agent is limited to 1. Fig. 7 shows the results of the single transaction setting. Fig. 7(a-b) show the result when the attacking behavior is *adversarial* in nature. The accuracy of SALE POMDP decreases with the increase in the number of attackers, as the probability of obtaining an incorrect opinion increases with the increase in the number of attackers. TRAVOS and Personalized obtain a greater accuracy of 100% for the cases when the majority of the advisors are honest in nature (No. of Attackers < 4). In the scenario with equal number of trustworthy and adversarial appraisers SALE POMDP performs better than TRAVOS and Personalized. The *Advisor* POMDP considers all advisors to be trustworthy and relies on the opinion from the single appraiser whom it asks opinion. Its performance is hugely dependant on the honesty of the appraiser it chooses to ask opinion, thereby making its accuracy as the probability of selecting the honest appraiser as shown in Fig. 7 (a). Fig. 7(b) shows the bank balance of the trust models which correspond to the accuracy in Fig. 7(a). We can see that SALE POMDP obtains a better bank balance than TRAVOS and Personalized even in a single transaction setting. The *Advisor* POMDP achieves the best bank balance in all cases because, it performs fewer opinion transactions. However, its performance is hugely reliable on the single advisor it chooses to ask opinions, which cannot be successful at all times. Also, its performance decreases in multiple transaction settings as shown in Fig. 6. Fig. 7(c), (d) show the results in a scenario when the attacking behavior is *random*. In general all trust models perform better than the *adversarial* scenario, as the *random* attackers can provide correct opinions in some cases. Again from Fig. 7(d), we find that SALE POMDP obtains a better bank balance than others.

4.2 Scalability of SALE POMDP

In SALE POMDP, the state of the environment within a POMDP is defined using a set of state variables. The state space of such a factored representation is exponential in the number of variables, due to which Algebraic Deci-

sion Diagram (ADDs) were introduced. In this section, we will discuss the scalability of SALE POMDP, utilizing the Symbolic Perseus. Table 1 shows the time taken to generate the policy for the given number of sellers and advisors. The number of state variables in each case is also shown. For the case with 2 advisors and 3 sellers, the time taken is 18.89s and 22.48s when the attackers are *adversarial* and *random* respectively. We can see that, to solve for the case with 4 sellers and 8 advisors, the time taken is nearly 3 hours, when the attackers show *random* behavior. Though this may be a considerable running time, we would like to further investigate methods which scale to larger domains.

Table 1: Time to Generate Policy

Attack Type	Sellers	Advisors	S	Time (sec)
Adversarial	2	3	2^6	00018.89
	2	5	2^7	00060.20
	3	6	2^{10}	00232.14
	4	8	2^{13}	08540.27
Random	2	3	2^6	00022.48
	2	5	2^7	00058.82
	3	6	2^{10}	00419.12
	4	8	2^{13}	12306.51

5. CONCLUSION AND FUTURE WORK

The paper suggests a novel method for dealing with the seller selection problem using POMDPs. Traditional trust models focus on accurately determining the quality of sellers rather than finding an optimal way to select sellers, in order to perform transactions. SALE POMDP optimally selects the right sellers by asking advisors opinions about the seller. The model also has the provision to ask advisors about the trustworthiness of other advisors in the market, which is a novel feature the approach offers. The *factored* nature of SALE POMDP can also be exploited to resolve the scalability issues in solving POMDPs with large number of states. Experimental results using the ART testbed verify that SALE POMDP outperforms state-of-the-art trust models in determining the quality of the seller, especially in various attacking scenarios. The scalability of the model is also analyzed, using the Symbolic Perseus solver.

Using techniques such as genetic algorithms to determine the optimal values for the observation and transition probabilities used in the SALE POMDP model will be considered as future work. Also methods to improve the robustness of the model, such as more detailed representation of advisor trustworthiness (different state variables to represent its trustworthiness in providing opinions about the seller and trustworthiness in providing opinions about other advisors, etc) by which SALE POMDP can defend more strategic attackers, will be studied. Though the *factored* nature of SALE POMDP has been analyzed to considerably reduce the complexity, we will also investigate methods to make the approach more scalable.

6. REFERENCES

- [1] C. Boutilier, N. Friedman, M. Goldszmidt, and D. Koller. Context-specific independence in bayesian networks. In *Proceedings of the Twelfth international conference on Uncertainty in artificial intelligence*, pages 115–123. Morgan Kaufmann Publishers Inc., 1996.
- [2] K. K. Fullam, T. B. Klos, G. Muller, J. Sabater, A. Schlosser, Z. Topol, K. S. Barber, J. S. Rosenschein, L. Vercoeur, and M. Voss. A specification of the agent reputation and trust (art) testbed: experimentation and competition for trust in agent societies. In *Proceedings of the fourth international joint conference on Autonomous agents and multiagent systems*, pages 512–518. ACM, 2005.
- [3] A. A. Irissappane, S. Jiang, and J. Zhang. Towards a comprehensive testbed to evaluate the robustness of reputation systems against unfair rating attack. In *UMAP Workshops*, volume 12, 2012.
- [4] L. P. Kaelbling, M. L. Littman, and A. R. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial intelligence*, 101(1):99–134, 1998.
- [5] H. Kurniawati, D. Hsu, and W. S. Lee. Sarsop: Efficient point-based pomdp planning by approximating optimally reachable belief spaces. In *Robotics: Science and Systems*, pages 65–72, 2008.
- [6] F. A. Oliehoek, A. A. Gokhale, and J. Zhang. Reasoning about advisors for seller selection in e-marketplaces via pomdps. In *Proceedings of the International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS) Workshop on Trust in Agent Societies*, pages 67–78, 2012.
- [7] C. H. Papadimitriou and J. N. Tsitsiklis. The complexity of markov decision processes. *Mathematics of operations research*, 12(3):441–450, 1987.
- [8] P. Poupart. *Exploiting structure to efficiently solve large scale partially observable Markov decision processes*. PhD thesis, University of Toronto, 2005.
- [9] K. Regan, R. Cohen, and P. Poupart. The advisor-pomdp: A principled approach to trust through reputation in electronic markets. In *Proceedings of the Third Annual Conference on Privacy, Security and Trust (PST)*, pages 121–130. Citeseer, 2001.
- [10] 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.
- [11] G. Shani. Task-based decomposition of factored pomdps. 2013.
- [12] D. Silver and J. Veness. Monte-carlo planning in large pomdps. In *Advances in Neural Information Processing Systems*, pages 2164–2172, 2010.
- [13] W. Teacy, J. Patel, N. 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.
- [14] A. Whitby, A. Jøsang, and J. Indulska. Filtering out unfair ratings in bayesian reputation systems. In *Proceedings of the International Joint Conference on Autonomous Agents Systems (AAMAS) Workshop on Trust in Agent Societies (TRUST)*, 2004.
- [15] J. Zhang. *Promoting Honesty in E-Marketplaces: Combining Trust Modeling and Incentive Mechanism Design*. PhD thesis, University of Waterloo, 2009.
- [16] 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.