Reasoning about Advisors for Seller Selection in E-Marketplaces via POMDPs

Frans A. Oliehoek[1], Ashwini A. Gokhale[2], and Jie Zhang[3]

[1] Maastricht University, Maastricht, The Netherlands
[2] MIT, Cambridge, MA, USA
[3] Nanyang Technological University, Singapore

Abstract. In an e-marketplace populated with a large number of sellers, some of which may be dishonest, the selection of good sellers to do business with is crucial but challenging especially when buyers do not have much experience with these sellers. In this paper we introduce the SALE POMDP, a framework for the seller selection problem that allows the decision maker to reason both about the quality of the sellers, as well as the trustworthiness of the advisors. In particular, the framework allows the agent to ask advisors about the trustworthiness of other advisors while still offering the benefit of optimally trading off information gathering and exploitation of knowledge as afforded by a POMDP based approach. Via this model, we present a preliminary investigation on the benefit of reasoning about trustworthiness of advisors. We also demonstrate how this enables incorporation of trust propagation as an integral part of the decision making process.

Keywords: Seller Selection, E-Marketplace, POMDPs.

1 Introduction

We consider the 'seller selection' problem in e-marketplaces, where an agent, the *buyer*, is assigned with the task of purchasing a particular item and needs to decide from which of the agents that offer the item it should order. In order to make this decision, the buyer maintains a belief over the quality levels of the various sellers. Also it can ask peers about their beliefs in order to improve its estimate of the quality levels. Only when the buyer is sufficiently sure that it has identified a seller with sufficient quality, should it go ahead and order the item (so the problem includes the decision of *whether* to place an order).

There have been a number of approaches to maintaining Bayesian beliefs over the quality levels of sellers, by integrating the buyer's own beliefs as well as the beliefs of other buyers (*advisors*) [12, 13]. These approaches tend to focus only on obtaining an accurate estimate of seller quality, but fail to reason about when it is necessary to query advisors in order to make optimal decisions.

An approach was suggested to perform full Bayesian decision making by casting the seller selection problem as a *partially observable Markov decision process (POMDP)*, named *Advisor POMDP* [6]. POMDPs provide a generic framework

for optimal decision making for an agent in a stochastic and partially observable environment [5]. The advantages of a POMDP approach are as follows: 1) rather than trying to achieve the most accurate estimate of sellers, the approach tries to select good sellers and does that optimally; reasoning about sellers is a means, not an end, and 2) a POMDP approach explicitly reasons about information gaining actions in partially observable environments, which allows the agent to optimally trade off the cost of obtaining and benefit of more information.

However, the Advisor POMDP framework assumes that all advisors are equally trustworthy. Following other approaches [12, 13], we acknowledge that it is important to model the trustworthiness of advisors; modeling the trustworthiness of advisors has a big impact on the optimal policy. We also believe that by allowing the buyer to query about (other) advisors we can integrate trust propagation into the decision making process, and thus improve the approach.

We introduce a new model called the *(S)eller & (A)dvisor se(LE)ction POMDP (SALE POMDP)*, which implements these ideas by explicitly incorporating a model of the advisors' trustworthiness in the state description. By asking advisors about both sellers and other advisors, a SALE POMDP-agent can improve its belief and subsequently take an informed decision on whether to place an order and if so from which seller. In this paper, we demonstrate how this belief revision process works, show that taking into account trust of the advisors is important in the seller selection problem and that, under certain circumstances, allowing the agent to ask advisors about other advisors allows it to realize a higher expected utility for its owner.

2 Background

2.1 Reputation Systems

Some sellers in e-marketplaces may be dishonest and not deliver products with the quality levels as they promised or declared. Thus, seller selection in such uncertain environments is important. Reputation systems have been introduced to address this issue and are particularly useful when buyers do not have much direct experience with sellers [4]. Among them, Bayesian approaches [12, 13] have drawn large attention. For example, Teacy et al. [12] proposed the TRAVOS model, which is a trust and reputation model based on the beta probability density function, and integrates a buyer's own beliefs about sellers as well as the beliefs of advisors. However, these approaches do not provide optimal decision making for the buyer on whether and from which seller to place an order, which is exactly what our approach tries to offer.

Those Bayesian approaches also suggest to model the trustworthiness of advisors as some advisors may lie about their experience with sellers. For example, [10] proposes to learn about the advisors by trying to estimate the properties of sellers and using those to estimate the advisor's advice given those properties. This is on a somewhat different time-scale than our approach. In particular, in order to learn about an advisor, the agent should first have many interactions

with a seller (to be certain enough about the sellers properties) at that point, the agent can learn what type of advice the advisor gives for such seller properties. This means, however, that in order to learn about a single advisor multiple transactions of both our agent and the advisor with the same seller are required. In contrast, we hope to be able to learn about advisors by asking other advisors, thereby avoiding the need to engage in costly transactions.

2.2 The Advisor POMDP

Regan et al. [6] introduced the Advisor POMDP, an approach for dealing with the seller section problem based on the POMDP framework. Formally, an Advisor POMDP consists of the following elements:

- There are I advisors that can be queried about the reputation of all J sellers.
- $\mathcal{S}$—a set of possible states of the environment. A state $s = \langle \boldsymbol{q}, sat \rangle$, where $\boldsymbol{q} \in [0,1]^J$ is a vector indicating the quality q_j of each seller and $sat \in \{-1, 0, +1\}$ indicates whether the result of a purchase is 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 transfering 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 $s = \langle \boldsymbol{q}, 0 \rangle$ changes stochastically to $s' = \langle \boldsymbol{q}, -1 \rangle$ or $s = \langle \boldsymbol{q}, +1 \rangle$ with probabilities depending on q_j.
- R—a reward function specifying $R(s,a,s')$. For ask actions, a small cost is paid independent of the state. 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 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 how certain the advisor is.
- 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. We will assume that h is infinite in this paper.

When the agent interacts with the environment, it can maintain a so-called *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 (as we will do in the remainder of the paper), this update can be written as follows:

$$b'(s') = \frac{\Pr(s', o|b, a)}{\Pr(o|b, a)} = \frac{1}{\Pr(o|b, a)} \Pr(o|a, s') \sum_{s} \Pr(s'|s, a) b(s). \tag{1}$$

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*:

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

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

Finding an optimal policy π^* is intractable in general (PSPACE complete [8]), however, in recent years substantial advances have been made in the approximate solution of POMDPs (e.g., [7, 11]).

3 Reasoning about which Advisors to Trust

The Advisor POMDP presents a coherent and principled framework to making decisions in the seller selection problem. However, there are some limitations to this model, as we now discuss.

A severe limitation is that the Advisor POMDP puts equal trust in all the advisors.[4] In a real system it is absolutely not a priori clear that all advisors can be trusted and we hypothesize that this may have a big impact on how one should act (i.e., what the optimal policy is). In fact, there is a large field of research on trust propagation that deals with the question of how one should adapt the trust in peers [3, 2]. A disadvantage of current approaches, however, is that they deal with the problem of most accurately estimating the trust levels, rather than integrating this type of reasoning with the decision process of selecting a seller. As a result, it is not clear how one would actually optimally apply such approaches in the context of seller selection. Here we try to overcome this problem by presenting a new model that incorporates these ideas from trust propagation within a POMDP formulation.

Also, in the Advisor POMDP, each advisor gives its ratings about all the sellers. However, instead of estimating the quality of all sellers, the only goal should be to select the seller with high quality. As such, the observation in the advisor POMDP may contain a lot of unnecessary information, leading to unnecessary communication. Therefore we will consider an approach in which our agent has to indicate about which seller (or other advisor) it wants to ask.

These ideas lead us to the formulation of a new model called the *(S)eller & (A)dvisor se(LE)ction POMDP (SALE POMDP)*, which we will formally introduce in the next section. Section 3.2 will present an example instantiation of the framework that we use in our experimental evaluation.

[4] By using a different observation function it would be possible to have observations from different advisors result in different beliefs, thereby modeling different levels of trust. These levels, however, would be assumed known to the agent.

3.1 The SALE POMDP Model

Like the Advisor POMDP, the SALE POMDP is a sub-class of POMDP problems. On the one hand the SALE POMDP is more complex than the Advisor POMDP: we assume that the advisors also have a quality, or *trustworthiness*, and that this is part of the state space. Moreover we introduce extra actions as we allow the agent to ask about the quality of other advisors. On the other hand, we make the simplifying assumption of having discrete sets of quality levels, which allows us to use standard POMDP solvers.

Since the SALE POMDP is a POMDP, it can be described in terms of states, actions, observations and rewards.

States. Like in the Advisor POMDP, a state contains the quality levels of all sellers, however, it also contains the quality, or *trustworthiness*, of each advisor. 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 $s = \langle \boldsymbol{q}, \boldsymbol{u}, sat \rangle$, where $\boldsymbol{q} \in \mathcal{Q}^J$ is a vector indicating the quality of each seller, $\boldsymbol{u} \in \mathcal{U}^I$ a vector indicating the quality of each advisor, and $sat \in \{-1, 0, +1\}$ as before. We also write q_j for the j-th element of $\boldsymbol{q}$ and u_i for the i-th element of $\boldsymbol{u}$. After a buy action is taken, the decision process ends. This is modeled using sets of *terminal states*. That is, a terminal state is a state where $sat = +1$ or $sat = -1$. We will think of these sets of states as single states called *satisfied* and *unsatisfied*.

Actions. The model knows the following types of actions:

- $seller\_query_{ij}$ — 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.

Transitions. As in the Advisor POMDP, we assume that when taking a query action, the state does not change:

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

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

where $\delta_{ss'}$ is the Kronecker delta that is 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 transitioning to 'satisfied' should be higher when buying from higher quality sellers j.

Together, the specifications of these transitions imply the assumption that quality and trust-levels are stationary for the duration of the decision process.

Rewards. The SALE POMDP specifies the following rewards: A small cost associated with ask actions $R(s, seller\_query_{ij}) = R(s, advisor\_query_{ii'}) = R_{ask}$, a reward associated with a good purchase $R(s, buy_j, s' = \langle \boldsymbol{q}, \boldsymbol{u}, sat = +1 \rangle) = R_{sat}$, and a penalty associated with dissatisfaction $R(s, buy_j, s' = \langle \boldsymbol{q}, \boldsymbol{u}, sat = -1 \rangle) = R_{unsat}$. There is a penalty associated with taking the do_not_buy action when in

fact there is a seller of high enough quality (we use $-R_{sat}$), otherwise the reward for this action is 0.

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 $o \in \mathcal{Q}$ corresponding to the quality of seller j, while after an $advisor\_query_{ii'}$ the agent will get an observation $o \in \mathcal{U}$ corresponding to the quality of the advisor i'. When the agent transitions to a terminal state, it receives the observation 'ended'. As such $\mathcal{O} = \mathcal{Q} \cup \mathcal{U} \cup \{ended\}$.

As in the Advisor POMDP, there is no a priori correct way to specify the observation probabilities. In fact, 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.

Initial State Distribution. The initial state distribution is dependent on the subjective beliefs of the agent (or its owner) when the need for purchasing an item arises. In the case that nothing is known, it makes sense to start with a uniform belief over the quality levels, but a different initial belief could have resulted from previous interactions.

That is, once the buy action is taken, the resulting belief can be used as the basis for an initial belief for a new seller selection instantiation. 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 task.[5]

3.2 Example

Suppose that there are $J = 2$ sellers for the item in concern, each of which can have $|\mathcal{Q}| = 2$ quality levels. In this example we use $\mathcal{Q} = \{L, H\}$ for L(ow) and H(igh) quality. Then we have $2^J = 4$ possible 'quality states' $\boldsymbol{q}$:

$$\boldsymbol{q} \in \{\langle L, L\rangle, \langle L, H\rangle, \langle H, L\rangle, \langle H, H\rangle\}. \tag{5}$$

Also suppose that there are $I = 3$ advisors, each of which is T(rustworty) or U(ntrustworty). That is $\mathcal{U} = \{T, U\}$. This leads to 2^I 'trust states' $\boldsymbol{u}$:

$$\boldsymbol{u} \in \{\langle T, T, T\rangle, \langle T, T, U\rangle, \ldots, \langle U, U, U\rangle\}. \tag{6}$$

As such, an example of a fully specified state is $s = \langle\langle L, H\rangle, \langle T, U, U\rangle, 0\rangle$.

The transition function for the *query* actions is specified as explained: the underlying state does not change. There is some freedom in specifying the transition

[5] 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 is untrustworthy.

u_i	q_j	good	bad
T	H	0.9	0.1
T	L	0.1	0.9
U	H	0.5	0.5
U	L	0.5	0.5

(a) $\Pr(o|seller\_query_{ij}, q_j, u_i)$ for the possible observations 'good' and 'bad'.

u_i	$u_{i'}$	good	bad
T	T	0.9	0.1
T	U	0.1	0.9
U	T	0.5	0.5
U	U	0.5	0.5

(b) $\Pr(o|advisor\_query_{ii'}, u_i, u_{i'})$ for the possible observations 'good' and 'bad'.

Table 1: Observation probabilities.

probabilities for reaching *satisfied* and *unsatisfied*, since this encodes the definition of the quality levels. In this example, 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.

In our example, we label the two possible observations 'good' (i.e., the advisor says that the seller/other advisor is high quality) and 'bad' (the seller/other advisor is said to be low quality). As mentioned above, the observation probabilities when transferring to a terminal state are fixed: the agent will observe *ended* with probability 1. The observation probabilities for the $seller\_query_{ij}$ action (ask advisor i about seller j) should be such that asking a trustworthy advisor i gives more accurate observations. One possible way to specify these probabilities is shown in Table 1a. Similarly, Table 1b shows example observation probabilities for the $advisor\_query_{ii'}$ action.

4 Experiments

In this section we report upon a first empirical investigation of the SALE POMDP model. In particular, we demonstrate how the belief update (1) leads to correlation of particular states which forms the basis of improved decision making. We also show that in our example setting it is important to explicitly take into account the advisor's trustworthiness and that asking advisors about other advisors can be beneficial in certain settings.

In order to perform the empirical evaluation we utilize SARSOP [7], a state-of-the-art POMDP solver, which reads in problems in a standardized POMDP description format. SARSOP does not exploit the factored structure of our problem, therefore, in specifying our models, we substituted all the terminals by two separate states $satisfied$, $unsatisfied$, reducing the number of states. Furthermore, the models we used specified two quality and trust levels as in the example of Section 3.2. Also, unless noted otherwise, the transition and observation models used the same parameters as described in that section. For the rewards, we used $R_{ask} = -1, R_{sat} = 50, R_{unsat} = -100$. Also, we penalized taking the *do_not_buy* action from states where there was a high quality seller with -50.

4.1 Illustration of Belief Update: Correlation between States

Here we provide some intuition behind the SALE POMDP model by illustrating the process of belief updating. The basic idea is that the belief updates should 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 trustworthy $u_i = U$. This is clearly demonstrated in a number of transitions in Figure 1b, which shows the policy found for a one seller ($J = 1$) one advisor ($I = 1$) SALE POMDP. Similarly, observing T_j after $advisor\_query_{ii'}$ should put more weight on states where $u_{i'} = T$ and $u_i = T$, and decrease weight on states where $u_{i'} = U$ and $u_i = T$.

4.2 The Impact of Taking into Account Trust

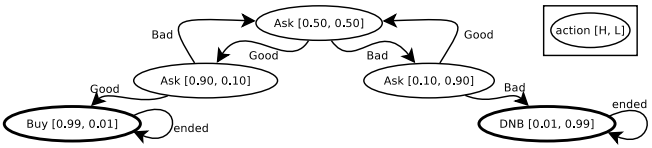
Here we compare (a simplified version of) the Advisor POMDP with one seller ($J = 1$) and one advisor ($I = 1$) to the SALE POMDP model. For both models we use the same discretization of quality levels ($\mathcal{Q} = \{L, H\}$), so the only difference is that the SALE POMDP includes the trustworthiness of the advisor as a state variable and that the observations are dependent on this factor. The observation model is as shown in Table 1.

Figure 1 shows the policies found for the two models. It clearly shows that the policies are qualitatively different. In particular, while in the Advisor POMDP it is possible to return to the initial belief after observing an equal number of 'good' and 'bad' observations. In contrast, in the SALE POMDP this leads to a belief where the advisor is thought to be untrustworthy. As such, the agent is able to reason about the trustworthiness of the advisor by repeated interactions.

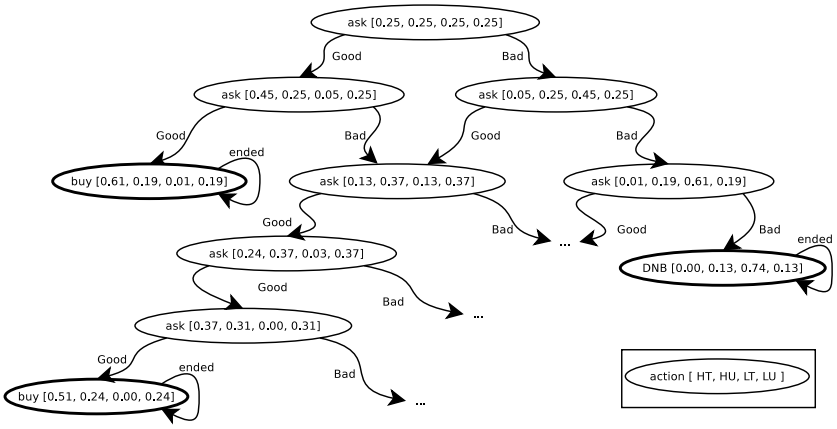
Since, the Advisor POMDP corresponds to the setting in which there is a single trustworthy advisor, it achieves higher value (mean value of 1000 evalutions is 5.36) than the SALE POMDP (-8.56). However, the policy found for the Advisor POMDP with only one untrustworthy advisor (i.e., with 'advice accuracy' 0.5) is much lower (-19.88). Interestingly, the mean of 5.36 and -19.88 (-7.26) corresponds to the setting where when an advisor type is chosen with 50% probability and then revealed to the agent. We see that the SALE POMDP achieves value fairly close to this 'oracle' upper bound.

4.3 Multiple Advisors: Trust Propagation

We also hypothesize that allowing the agent to query advisors about other advisors, thereby integrating a form of trust propagation in the seller selection decision procedure, can allow for further improvements. In order to test this hypothesis, we consider the SALE POMDP framework with three advisors and compare it to a baseline model: the same SALE POMDP model but without the $advisor\_query_{ii'}$ actions, which we will call the NoAQ model. The top row of Table 2 lists the results of this comparison. In contrast to our expectation,



(a) Advisor POMDP policy.



(b) (Partial) SALE POMDP policy.

Fig. 1: Comparison of policies found. Thick nodes indicate where *buy* and *do_not_buy* actions are taken. Nodes also show the belief over (non-terminal) states.

we see that the NoAQ model actually performs better. However, since the set of policies for the regular SALE POMDP model is a strict superset of those for the NoAQ model, we know that the former should be able to achieve at least the same value. The fact that this does not happen can be attributed to the additional computational complexity (induced by the additional actions).

The fact that the regular model does not outperform NoAQ also means that the latter is able to sufficiently figure out which advisors are trustworthy using only *seller_query* actions (as also discussed in Sect. 4.2). Therefore we form a new hypothesis that asking about advisors is beneficial when *seller_query* actions do not provide much information about the trustworthiness of an advisor. This is confirmed by the other test results shown in the table that show what happens if the accuracy with which trustworthy advisors report about sellers (i.e., the '0.9' from Table 1a) diminishes.

The mentioned additional computational complexity of the SALE POMDP model is further demonstrated in the bottom part of the table. It shows that allowing for additional solution time over 100s is improving the quality of the policy further, while for NoAQ the further improvement is marginal.

	regular			NoAQ		
	mean	LB	UB	mean	LB	UB
Varying accuracy of advisors w.r.t. seller quality (1000s)						
accuracy						
0.9	−3.24	−3.67	−2.81	−2.87	−3.30	−2.44
0.8	−5.31	−5.74	−4.87	−6.39	−6.79	−5.99
0.7	−8.08	−8.43	−7.73	−9.77	−10.10	−9.44
0.6	−13.20	−13.50	−12.90	−14.82	−15.09	−14.56
Impact of solution time (0.7 accuracy)						
time(s)						
100	−8.50	−8.85	−8.15	−9.80	−10.13	−9.47
500	−8.63	−8.99	−8.27	−9.78	−10.11	−9.46
1000	−8.08	−8.43	−7.73	−9.77	−10.10	−9.44

Table 2: Results for the SALE POMDP with multiple advisors. Shown are mean values from 1000 evaluations, together with 95% confidence bounds.

5 Discussion & Future Work

Here we discuss the limitations of the SALE POMDP model and point to avenues for future work.

5.1 Overcoming Restricting Modeling Assumptions

A restrictive assumption is that we assume independent responses from the same advisor. That is a response of an advisor does not depend on a earlier response from that same advisor. While this is a common assumption (e.g. [10]), it might not be the most realistic. This restriction can be overcome by modeling a previous response as part of the state. The state factor for an advisor i could for instance have value ⟨trustworthy, not queried yet⟩ or ⟨untrustworthy, seller-3-high-quality⟩. The current model can also be a good model for groups of advisors, when we ask a different person within that particular group (e.g., IP-range).

Another issue is that we are currently assuming a simplistic form of untrustworthy: basically (more) random. However, in real-life, untrustworthy advisors may give very biased answers. In such cases it is much more difficult to identify untrustworthy advisors, but this would also be a problem for a human decision maker. Still, the result of the final belief update (transferring to a unsatisfied state), will correct for the wrong belief that that advisor was trustworthy. Moreover, other approaches to dealing with deceptive advisors based on Bayesian updating (such as [10]) can be neatly integrated in our approach.

5.2 Scaling Up

Solving POMDPs is intractable in general, but in recent years, huge advances have been made in approximate solutions: good policies have been found for

problems with thousands of states. Still, our empirical evaluation indicates that computational costs are a limiting factor in the use of our framework. However, there are a number of different forms of structure in the proposed POMDP formulation that can potentially be exploited for faster solutions. For instance, the SALE POMDP really is a *factored* POMDP, so we may try and use solvers that exploit this property such as symbolic Perseus [9]. Moreover, there is other special structure that we may be able to exploit. For instance, we do not care about which seller we end up ordering from, but only about whether that seller has sufficient quality. This means that there are symmetries between sets of states (e.g., a $\langle H, L, T\rangle$ has the same value as $\langle L, H, T\rangle$) which may be exploited [1]. Moreover, there may be little difference in value between (beliefs assigning high probability to) states that have one high-quality seller and states that have multiple high quality sellers. In future work we hope to exploit these types of structure for improved computational efficiency.

5.3 Learning Accurate Models

For our proof-of-concept experiments, we specified all the parameters in an ad-hoc fashion. However, it is important to know that all those numbers can be estimated in a sensible way. Moreover, for a decentralized peer-to-peer type of system to work, it is required that each peer (each agent) can adapt its model by learning because 1) it needs to be able to adapt to changes over time, and 2) it needs to adapt to the preferences of its owner [10].

Learning POMDP models is a very difficult problem, but there are some special properties that could be exploited. First, the observation model might not need to be learned since it in fact encodes the definition of our different trust-levels.[6] Second, the state of the world in this problem does not change (or perhaps only very slowly), which may lead to easier learning. Finally, we would like to point out that even if the model is not completely accurate, it is still possible to get out very high quality policies: The general form of "buy when certain enough about one seller" does not change. For instance, by setting the reward function conservatively, we know that the agent might perform too many queries in case that the model was right, but it also builds in a robustness against the uncertainty in model parameters.

6 Conclusion

In this paper, we proposed a novel *(S)eller & (A)dvisor se(LE)ction POMDP model (SALE POMDP)* to address the problem of seller selection in e-commerce settings. Our model provides a principled approach for buyers to optimally select sellers trading off the cost and benefit of seeking more information from advisors. We provided examples to demonstrate step by step how our model works, and experiments to demonstrate its effectiveness. Encouraged by this promising first

[6] An interesting other question is how we may *optimize* these definitions.

step, we also discussed several important next steps to improve our model and increase its applicability.

Acknowledgments

This work was supported, in part by AFOSR MURI project #FA9550-09-1-0538, in part by NWO CATCH project #640.005.003, and in part by Singapore Ministry of Education under a grant to the Singapore-MIT International Design Center.

References

1. Doshi, F., Roy, N.: The permutable POMDP: fast solutions to POMDPs for preference elicitation. In: Proc. of the Int. Joint Conference on Autonomous Agents and Multi Agent Systems. pp. 493–500 (2008)
2. Gorner, J., Zhang, J., Cohen, R.: Improving the use of advisor networks for multi-agent trust modelling. In: Proc. of the 9th Annual Conference on Privacy, Security and Trust (PST) (2011)
3. Hang, C.W., Wang, Y., Singh, M.P.: Operators for propagating trust and their evaluation in social networks. In: Proc. of the Int. Joint Conference on Autonomous Agents and Multi Agent Systems (2009)
4. Jøsang, A., Ismail, R., Boyd, C.: A survey of trust and reputation systems for online service provision. Decision Support Systems 43(2), 618–644 (2007)
5. 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)
6. Kevin Regan, R.C., Poupart, P.: The Advisor-POMDP: A principled approach to trust through reputation in electronic markets. In: In Proc. of the Third Annual Conference on Privacy, Security and Trust. pp. 121–130 (2005)
7. Kurniawati, H., Hsu, D., Lee, W.S.: SARSOP: Efficient point-based POMDP planning by approximating optimally reachable belief spaces. In: Proc. Robotics: Science & Systems (2008)
8. Papadimitriou, C.H., Tsitsiklis, J.N.: The complexity of Markov decision processes. Mathematics of Operations Research 12(3), 441–451 (1987)
9. 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)
10. Regan, K., Poupart, P., Cohen, R.: Bayesian reputation modeling in e-marketplaces sensitive to subjectivity, deception and change. In: Proc. of the National Conference on Artificial Intelligence. pp. 1206–1212. AAAI Press (2006)
11. Silver, D., Veness, J.: Monte-carlo planning in large POMDPs. In: Advances in Neural Information Processing Systems 23. pp. 2164–2172 (2010)
12. Teacy, W.T.L., Patel, J., Jennings, N.R., Luck, M.: Coping with inaccurate reputation sources: Experimental analysis of a probabilistic trust model. In: Proc. of the Int. Joint Conference on Autonomous Agents and Multi Agent Systems (2005)
13. Zhang, J., Cohen, R.: Evaluating the trustworthiness of advice about selling agents in e-marketplaces: A personalized approach. Electronic Commerce Research and Applications 7(3), 330–340 (2008)