The Advisor-POMDP: A Principled Approach to Trust through Reputation in Electronic Markets

Kevin Regan
University of Waterloo
School of Computer Science
kmregan@cs.uwaterloo.ca

Robin Cohen
University of Waterloo
School of Computer Science
rcohen@uwaterloo.ca

Pascal Poupart
University of Waterloo
School of Computer Science
ppoupart@cs.uwaterloo.ca

Abstract

This paper examines approaches to representing uncertainty in reputation systems for electronic markets with the aim of constructing a decision theoretic framework for collecting information about selling agents and making purchase decisions in the context of a social reputation system. A selection of approaches to representing reputation using Dempster-Shafter Theory and Bayesian probability are surveyed and a model for collecting and using reputation is developed using a Partially Observable Markov Decision Process.

1. Introduction

Trust is a desirable property of any market, because it reduces the friction by which we do business. A good example of ease which trust provides is a business agreement using a handshake instead of legal contracts. The success of trust in streamlining transactions in traditional markets motivates the search for a comparable measure of trust in emerging electronic markets which can be populated by agents that automate the transactions between buyers and sellers. We examine trust from the perspective of an agent required to decide who to do business with, based in part on trust. A model of reputation can aid in these trust decisions by providing a reputation for each agent, essentially modeling how trustworthy an agent was in the past. In any market without perfect security, trust will be an important factor in any purchase decisions made by a buyer.

The aim of this paper is to construct a principled framework for buyers to choose the best seller based on some measure of reputation in a market consisting of autonomous agents. We proceed to this framework by first examining three distinct classes of reputation systems. Upon choosing the reputation system most appropriate for our multi-agent environment, we take a close look at how reputation can be modeled so as to capture the most relevant aspects of the reputation system. We then present a framework based on Markov Decision Processes and give a brief discussion of some of the challenges to the efficient functioning of it.

As a whole, our research explores the topic of trust in communities populated by intelligent agents, specifically in electronic marketplaces. In particular, we provide a basis for designing effective buying agents that will be trustworthy for their users by making informed purchase decisions. Our research also proposes an environment in which trust between buying agents is fostered, including the sharing of information in order to make the most effective purchases.

2. Reputation Systems

We can define three classes of reputation systems based on the source of the information used to construct the reputation and who has access to the reputation.

2.1. Global Reputation Systems

One approach to modeling the reputation of sellers is to establish a central service which is responsible for collecting feedback from buyers, constructing a single reputation for each seller in the market and making this global reputation available to all buyers. Examples of this approach are reputation systems used by online auction sites such as eBay[1] and amazon.com auctions[2] and theoretical models such as CONFESS [8]. While the presence of a global reputation allows buyers to learn about sellers they have not yet interacted with, there are some drawbacks to this approach.

The central service that tracks and publishes seller reputations must be trusted by all agents in the marketplace. There is no simple way to evaluate the truthfulness of feedback given by buyers. Since the central service is not directly involved in any transaction, it can not easily verify

[1] www.ebay.com
[2] auctions.amazon.com

the quality of the goods being shipped or received. Furthermore, the presence of a central service carries the common disadvantages inherent in centralized architectures, such as having a single point of failure, and not scaling well as the number of agents increase.

2.2. Personal Reputation Systems

Another approach to modeling seller reputation is to allow each buying agent to individually collect feedback from past purchases to develop a personal model of a selling agent's reputation w hich is constructed using only the transactions in which the individual buying agent has been involved. An example of this approach is the reputation model developed by Cohen and Tran [29] in which a buying agent uses only its past purchases from sellers to learn to avoid dishonest sellers.

The advantage to only using transactions the buying agent has been involved with is the agent is certain regarding the outcome of those transactions. However, with this approach the buying agent is limited to modeling only those sellers from whom the buying agent has purchased in the past. There are many situations in which the set of potential selling agents for a good may be comprised of agents with whom a buying agent has no direct experience.

2.3. Social Reputation Systems

A natural extension of the personal reputation model is one in which a buying agent can choose to query other buying agents for information about sellers for which the original buying agent has no information. We describe the other buying agents in this context as advisors. There are many examples in the literature of reputation systems that allow agents to share reputation [12, 24, 33, 32]. However, not all systems use the same representation of reputation.

A social reputation system allows for a decentralized approach whose strengths and weaknesses lie between the extremes of the personal and public reputation system. The main advantage is that the responsibility for collecting feedback and constructing a reputation model rests with the individual buying agent. While a buying agent may not have access to a global seller reputation that takes into account *all* past buyer interactions, the buying agent has the freedom to solicit as much or as little information as it needs from others until it has constructed a reasonable model of a seller's reputation.

Using the social reputation model as a foundation we will now examine possible representations of the reputation that will be the basis of our buying agents' decisions.

3. Reputation Representation

The model of reputation will be constructed from a buying agent's positive and negative past experiences with the aim of predicting how satisfied the buying agent will be with the results of future interactions with a selling agent. The model of reputation needs to capture two important and distinct notions of uncertainty about how past interactions will dictate future interactions. We classify these two classes of uncertainty in a similar fashion to Sentz and Ferson [25] as:

Stochastic Uncertainty - uncertainty which results from the randomness of a system.
Also known as: irreducible, aleatory, or objective uncertainty as well as variability.

Epistemic Uncertainty - uncertainty which results from a lack of knowledge about the randomness of a system.
Also known as: reducible or subjective uncertainty as well as ignorance.

To function within our social reputation system, we must be able to perform some specific operations on the reputation of a seller held by a buyer. Given a set of reputations collected from other buyers we need to be able to combine these reputations. This combination needs to respect the differing levels of trust that one buyer may have in another. For example, if reputations were represented by a single number, then a simple average over all the reputations collected from other buyers would not take into account the fact that some of the other buyers may have less experience or are less trustworthy.

Work has been done to represent reputation in many different ways. We will now survey some of this work, moving from fairly simple ad-hoc reputation models [33], to systematic models [2, 7, 15, 31, 32] which rely on Dempster-Shafer Theory [4, 26] and Bayesian probability.

3.1. Ad-hoc Reputation Models

There are many models of reputation in the literature that allow the reputation of a seller to be represented by a single value. Most of the work on these models involves deriving equations for the update of this reputation value such that it exhibits some desired behavior. An often cited example of such a reputation model is the Sporos reputation mechanism [33] which uses the following expression to update the single reputation value of a seller:

$$R_{t+1} = \frac{1}{\Theta} \sum_{1}^{t} \Phi\left(R_i\right) \cdot R_{i+1}^{other} \cdot \left(W_{i+1} - \frac{R_{t+1}}{D}\right) \quad (1)$$

$$\Phi(R) = \left(1 - \frac{1}{1 + e^{\frac{-(R-D)}{\sigma}}}\right)$$

A full understanding of the preceding expression is not necessary; the aim is simply to draw the reader's attention to some aspects of the model. Expression 1 allows for the weighted combination of reputation information for a seller given by other buyers. In the expression the rating W_i for a seller given by another agent i is weighted by reputation of that other agent denoted by R_{i+1}^{other}. The major drawback of such ad-hoc reputation models that represent reputation using only a single value is that they do not contain any measure of the *epistemic* uncertainty. In the context of our social reputation system, the use of equation 1 leaves no clear way to determine when enough other buyers have been consulted to make an informed decision about which seller to purchase from.

3.2. Dempster-Shafer Theory

Dempster-Shafer Theory (DST) is a mathematical theory of evidence which rests on a generalization of probability theory in which probabilities are assigned to sets instead of mutually exclusive atomic events. We can interpret the elements of the sets as possible hypotheses about events. DST does not force the sum of the probability of the atomic elements to sum to one, so the *epistemic* uncertainty due to, for instance, the lack of evidence against a hypothesis is easily expressed. The likelihood of a particular hypothesis given a set of evidence can be reasoned about using the following three functions:

Basic Probability Assignment - The basic probability assignment, denoted bpa or m, defines a mapping of all possible subsets of the set of our atomic elements to a number between 0 and 1.

Belief function - The belief function, denoted $bel(A)$ for a set A, is defined as the sum of all the basic probability assignments over all proper subsets of A.

Plausibility function - The plausibility function, denoted $pl(A)$ for a set A, is defined as the sum of all the basic probability assignments over all the sets B that intersect the set A.

The basic probability assignment for a given set A can be thought of as expressing the proportion of evidence that supports the claim that some element X belongs to the set A, but to no particular subset of A. The belief and plausibility functions essentially represent a lower and upper bound on the likelihood of a hypothesis represented by A.

The reputation system developed by Yu and Singh [32] should help make our discussion of DST concrete and illustrate how DST can be used to model reputation. They define $\{T, \neg T\}$ to be their set of hypotheses. In their model the bpa $m(\{T\})$ represents the evidence for a good seller reputation and can be calculated by taking the proportion of all past experiences in which the buying agent's satisfaction with a purchase was above some threshold. $m(\{\neg T\})$ represents the evidence for a bad seller reputation, and can be calculated by taking the proportion of all past experiences in which the buying agent's satisfaction with a purchase was below another threshold. $m(\{T, \neg T\})$ measures the *epistemic* uncertainty or lack of evidence and is found by simply taking the proportion of past experiences that fall between the two thresholds.

In his original work on the subject, Shafer [26] developed a method for combining beliefs about the same set of elements that are based on distinct bodies of evidence. This allows for reputation information collected from other buyers in the market to be combined to form a new reputation. To this basic approach to combining reputation, the authors Yu and Singh add a method for taking into account how trustworthy other agents are by adapting Littleston and Warmuth's weighted majority algorithm [10] to allow for reputations with different weights to be combined.

The reputation model developed by Yu and Singh [32] provides a representation for reputation in our social reputation system that takes into account both *stochastic* and *epistemic* uncertainty while allowing for reputation to be updated through weighted combinations of the reputation collected from other buying agents. However, an even richer representation of the *epistemic* uncertainty can be obtained with Bayesian interpretations of tradition probability theory.

3.3. Bayesian Approaches

We can represent the *stochastic* uncertainty inherent in a process using basic probability. Given a coin that has yielded 8 heads and 2 tails after 10 flips, it is natural to say we believe the next flip will be heads with 0.8 probability. At first glance it does not capture the *epistemic* uncertainty since if we had seen 800 heads and 200 tails, the probability $p = 0.8$ of heads does not capture our increasing certainty about our knowledge of the underlying process. However, by using a probability density function which represents a second-order probability assigning a probability to each value of p we can capture both classes of uncertainty. We begin with a prior distribution over all the values of the probability p of heads and update this distribution with each coin flip we observe. This approach can be referred to as Bayesian since we represent our beliefs about the outcome of the coin flip using a distribution which is updated as we gather evidence.

The beta probability density function allows us to represent the probability distribution over the outcome of binary events such as heads/tails, or in our market setting, the transactions in which a buyer is satisfied/unsatisfied.

Beta Distribution - The beta distribution is a family of

probability density functions indexed by the parameters α and β and can be expressed using the gamma function as follows:

$$f(p|\alpha,\beta) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} p^{\alpha-1}(1-p)^{\beta-1} \quad (2)$$

which yields the following simple expression for the expectation:

$$E(p) = \frac{\alpha}{\alpha+\beta} \quad (3)$$

A nice property of the beta distribution is the ease at which a distribution can be calculated that incorporates a prior distribution and new observations. If we have r observations of the outcome x and s observations of the outcome $\bar{x}$, we can express the beta distribution in terms of these observations by setting $\alpha = r + 1$ and $\beta = s + 1$. Figure 1 shows a beta distribution given by Jøsang and Ismail [7] for a process in which x has been observed 7 times and $\bar{x}$ has been observed once which gives us $f(p|8, 2)$.

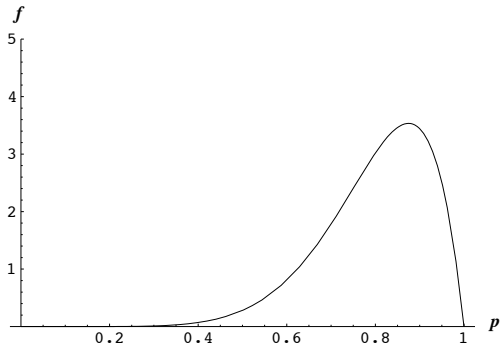

Figure 1. Beta function $f(p|8, 2)$

Jøsang and Ismail develop the Beta Reputation System [7] in which the binary process is a series of transactions in which a buyer is either satisfied or unsatisfied. The observations r are interpreted as positive feedback and the observations s as negative feedback. The expectation $E(p)$ models the *stochastic* uncertainty while the distribution over all possible values of p models the *epistemic* uncertainty.

Combining the feedback r_1, r_2 and s_1, s_2 from two different buying agents (1 and 2) in this model is as simple as constructing a new distribution with $r = r_1 + r_2$ and $s = s_1 + s_2$. To take into account the trust one agent may have for another when combining feedback, the authors develop a more sophisticated model that allows buying agents to model the reputation of other buying agents. This is best illustrated with an example in which we have a buying agent X who uses the feedback provided by buying agent Y about a selling agent Z. The buying agent X models the trustworthiness of the other buying agent Y by keeping track of feedback r_Y^X and s_Y^X from past interactions with Y[3]. The buying agent Y provides the feedback r_Z^Y and s_Z^Y about the selling agent Z and our buying agent X can weigh this feedback with what it knows about Y to construct the feedback $r_Z^{X:Y}$ and $s_Z^{X:Y}$ about Z as follows:

$$r_Z^{X:Y} = \frac{2\, r_Y^X r_Z^Y}{\left(s_Y^X + 2\right)\left(r_Z^Y + s_Z^Y + 2\right) + 2\, r_Y^X} \quad (4)$$

$$s_Z^{X:Y} = \frac{2\, r_Y^X s_Z^Y}{\left(s_Y^X + 2\right)\left(r_Z^Y + s_Z^Y + 2\right) + 2\, r_Y^X} \quad (5)$$

This feedback is then incorporated into the density function to arrive at what Jøsang and Ismail call the discounted reputation function by X through Y [7]. Like the DST model presented by Yu and Singh, the Beta Reputation System provides methods for combining weighted reputation information from other agents. Each model captures both the *stochastic* and *epistemic* uncertainty; however, the Bayesian approach used by the Beta Reputation System allows for a richer representation of the *epistemic* uncertainty since a distribution is maintained over each possible value of the probability modeling *stochastic* uncertainty. The Yu and Singh model, in comparison, uses a single scalar value to represent the *epistemic* uncertainty.

The Beta Reputation System is not the only work using Bayesian methods to model reputation. Mui et al. [15] develop a similar model based on the beta distribution, but do not mention methods for combining and weighing information from other buying agents. Barber and Kim [2] use a Bayesian network to combine reputation information gathered from other buying agents where each connection represents the conditional dependence of a selling agent's reputation on the reputation contributed by each buying agent. Wang and Vassileva [31] use a Bayesian network to represent how the buyer/client's observations of a seller/server's different capabilities can influence a single trust rating for the seller.

The Beta Reputation System illustrates how Bayesian methods can be used to construct a rich model of reputation. Unfortunately, in the context of a social reputation system, the work of Jøsang and Ismail [7] and others [2, 15, 32] does not address how reputation information is collected from other buyers, and how purchase decisions are eventually made. The next section will lay out a decision theoretic framework that makes use of Bayesian methods to develop policies about when to ask other buyers and when to make a purchase.

[3] the superscript denotes who is holding the feedback, while the subscript denotes who it is about

4. Decision Framework

4.1. Definition of Reputation

Reputation is a rich concept with many dimensions, but for our purposes we restrict the definition of the reputation of a seller in a similar way to Carter and Ghorbani [3] as a measure of how well the seller fulfills the role of providing quality goods at a reasonable price. The extent to which a selling agent fulfills this role is defined by how the buyer's expectations are met and these expectations are captured by a utility function. As in Tran and Cohen [29] the buyer has an expected utility which it demands of a transaction with a seller. After a good is delivered the buyer compares this expected utility with the utility which was actually realized. If the realized utility surpasses the expected utility, then a buyer is said to be satisfied and if the realized utility falls below the expected (demanded) utility, the buyer is said to be unsatisfied. A seller who consistently fulfills its role and satisfies the buyer will attain a high reputation and engender trust.

Many other duties can be incorporated into the seller's role, such as delivering a good on time, but for the purposes of exploring how to gather and use reputation information in a principled manner, we limit our selves to this narrow definition of seller reputation.

4.2. The Advisor-POMDP

In our social reputation system a buyer will ask other buyers (which we denote advisors) in order to accumulate information about a seller's reputation before making a decision about which seller to purchase from. Generally, we seek some principled way to integrate the reputation information provided by each advisor with what the buyer already knows, while allowing for the possibility that the advisors may vary in accuracy and experience.

We can use the a Bayesian interpretation of probability to represent the uncertainty about what information an advisor may provide and the satisfaction a buying agent will experience after a purchase. We assign utilities to possible outcomes of actions and and use these utilities to determine the best possible action. A natural way to model the decision making process given uncertainty about possible utility is a Markov Decision Process.

A Markov Decision Process (MDP) is defined by the tuple $\langle \mathcal{S}, \mathcal{A}, T, R \rangle$ where $\mathcal{S}$ is the set of states and $\mathcal{A}$ is the set actions that can be taken from each state. Each action will probabilistically move the agent into another state as determined by a transition function T. A reward is associated with each state using the reward function R, and a policy can be constructed which dictates which action to take from each state so as to maximize the expected reward. Puterman [21] gives a good introduction to methods for computing optimal policies for a given Markov Decision Process.

We can define an MDP for our social reputation system in which the states which represent the *stochastic* uncertainty about each seller and the actions a buyer can take are to *ask* an advisor or *buy* from a seller. Specifically the state will hold a reputation for each seller represented as the probability that seller will satisfy the buyer by fulfilling its role. However, our buyer has only partial knowledge of the underlying *stochastic* process that the reputation is modeling since it only has information about a subset of a seller's past interactions. We can model this *epistemic* uncertainty by extending the MDP to a Partially Observable Markov Decision Process (POMDP), described by Kaelbling et al. [9], which places a belief distribution over the possible states and uses observations to adjust this belief. Instead of knowing exactly what the current state is, the agent will have a belief about the current state, represented by the function $b : State \rightarrow [0, 1]$ which assigns a probability to every possible state. In our reputation system the state represents the actual reputations of the sellers (given all interactions), and is only partially observable through the information given by advisors.

The way in which the elements of the POMDP interact is illustrated in Figure 2. A directed arrow in the figure indicated that the probability of the reward, state or observation being pointed to is influenced by the state or action that is the source of the arrow.

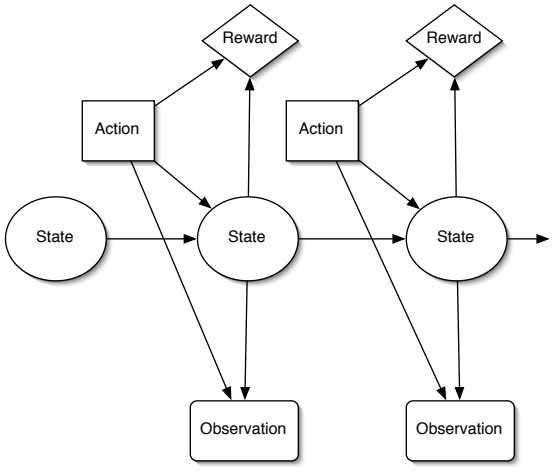

Figure 2. An influence diagram for the POMDP

Our Advisor-POMDP is described by the tuple $\langle \mathcal{S}, \mathcal{A}, T, R, \Omega, O \rangle$ for which each element is defined as follows:

$\mathcal{S}$ - State

A state $\langle \vec{r}, sat \rangle$ in our POMDP is composed of a vector $\vec{r}$ of real values in the range [0,1] representing the reputation of each seller and a scalar value $sat \in \{-1, 0, 1\}$ representing the satisfaction resulting from a purchase. Before a purchase the sat component of our state modeling buyer satisfaction will be equal to zero. For convenience, we will refer to any state in which sat is zero as an advice state. After a purchase the sat component is set to -1 or 1 and we refer to such a state as a purchase state. A sat value of 1 will signify that the buyer is satisfied with a purchase, while a sat value of -1 means the buyer is unsatisfied.

$\mathcal{A}$ - Actions

A buying agent can choose from two sets of possible actions. It can either choose to *ask* an advisor for information about a selling agent or it can choose to *buy* from a selling agent.

T - State-Transition function

In an advice state we can interpret the $\vec{r}$ component of the state as representing the actual reputations of sellers and an *ask* action will not change this, nor will it change the sat value. Asking an advisor for information leaves us in the same state, but changes our belief b about how likely we are to be in each possible state, since we cannot directly observe the current state. Formally, given any advice state q, the transition function, for any *ask* action, will map back to q. The *buy* action will transition from a advice state to a purchase state where the sat component of the state represents the outcome of the purchase.

R - Reward

There is no reward associated with advice states, however we associate a small cost for ask actions. A purchase state with a sat value of 1 will yield a large positive reward, while a sat value of -1 will yield a large negative reward.

Ω - Observations

The observations in our POMDP are composed of the information received by our buying agent in response to asking advisors. The advisor will respond with $\langle rep_i, cf_i \rangle$ where rep is the reputation and the certainty factor cf is a measure of the epistemic uncertainty for each seller i.

O - Observation function

The observation function expresses the likelihood of receiving an observation given the current state and the action that led to this state. We can interpret the observation function as the likelihood of an advisor giving a set of seller reputations, given the actual seller reputations. For instance, an honest, knowledgeable advisor would be likely to paint a seller as reputable, given that the seller is in fact actually reputable.

We can specify the observation function to account for how each individual advisor will change our belief about the actual reputations of sellers, and we can specify the transition function that maps *buy* actions in reputation states to purchase states to capture how our knowledge about a seller's reputation will dictate the outcome of a purchase from a particular seller. Given both the transition and observation functions, the POMDP provides a principled framework for integration of this information to find the best action given the agents current belief.

Given a POMDP specified by $\langle \mathcal{S}, \mathcal{A}, T, R, \Omega, O \rangle$, we would like to find the best course of action that maximizes expected rewards. We define a *policy* π to be a mapping from belief b (a probability distribution over states) to actions a. Intuitively, a policy encodes a strategy specifying which action should be executed at each time step given the current belief state.[4] We can measure how good a policy π is by defining a value function $V^{\pi}(b_0)$, which indicates the expected total rewards earned by following π from any initial belief state b_0 (e.g. $V^{\pi}(b_0) = \sum_{t=0}^{\infty} E_{\pi}[R(b_t)]$ where $R(b_t) = \sum_s b_t(s)R(s)$ and $E_{\pi}[\cdot]$ denotes an expectation b_t with respect to π). A policy π^* is optimal when its value function V^{π^*} is at least as high as any other policy π for all belief states (e.g. $V^{\pi^*}(b) \geq V^{\pi}(b)\ \forall \pi, b$).[5] We will explain in Section 4.4 how to find an optimal policy for a given POMDP.

In the context of reputation modeling, the POMDP framework provides a principled approach for optimizing the exploration/exploitation tradeoffs that arise when having to decide between buying from a seller or asking advisors for more information about some seller's reputation. Intuitively, asking an advisor to share its experience about a seller provides information that reduces the uncertainty in a seller's reputation. In turn, this reduced uncertainty, will allow the buyer to make a more informed decision when selecting a seller in the future. We can quantify the *value* of the information gained by the amount of utility that we expect to gain when using this information in our buying decision. In general, it will only make sense to ask an advisor to share its experience about a seller, when the expected value of the information gained is higher than the cost of consulting the advisor. By definition, optimal policies for the advisor-POMDP earn the highest possible expected to-

[4] We assume a *stationary* policy that uses the same mapping from belief states to actions at each time step. This assumption is reasonable since there always exists an optimal policy within the class of stationary policies [11].

[5] There always exists a *dominating* policy, which has a value function at least as high as any other policy for all belief states [11].

tal reward, and therefore pick at any point in time the best action (e.g., the best advisor to consult or the best seller to buy from).

The next section illustrates with an example how a policy might dictate the actions of our buyer.

4.3. Example

To help ground our discussion, we take a closer look at a scenario involving two sellers s_1, s_2 and four advisors a_1, a_2, a_3, a_4. Given that we are representing two sellers, the $\vec{r}$ component of state space representing the seller reputations for our Advisor-POMDP will be two dimensional. It may be helpful to use some images to conceptualize what is happening with the belief over the possible states. If the $\vec{r}$ component were composed of a single seller reputation r and we ignored the sat component, then the possible advice states would be all the possible values of $r \in [0, 1]$. A belief representing high *epistemic* uncertainty would assign a small probability close to zero to every state and could be graphed as follows:

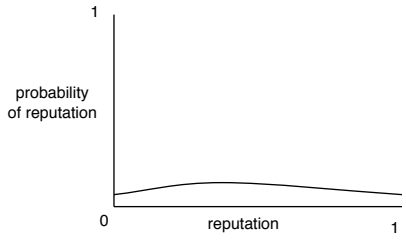

Figure 3. Belief - high epistemic uncertainty

A belief representing low *epistemic* uncertainty would assign low probabilities most states and a high probability to a small number of states. A graph of one such belief is as follows:

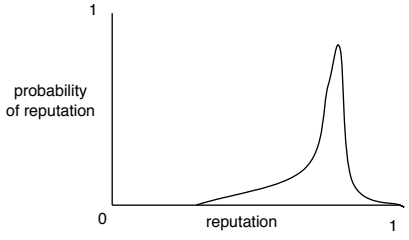

Figure 4. Belief - low epistemic uncertainty

The state space for our example with two seller reputations can be conceptualized as a two dimensional plane and the belief over these reputations as a surface of this plane. When the belief expresses high *epistemic* uncertainty, the surface would be relatively flat. As the *epistemic* uncertainty decreases and the agent becomes more certain about what state it is in, the surface will develop valleys over states that are not likely and peaks over states which are likely. To simplify our example let us discretize this state space into three possible reputations. Let us assume that a seller can either have a low, medium or high reputation, represented by the values 0.1, 0.5 and 0.9 respectively. Now let us assume that the buyer begins with little information about the reputations of sellers. Table 1 gives a belief for each possible advice state with the sat component omitted given that it will be zero.

Table 1. Belief $b(\langle r_{s_1}, r_{s_2}\rangle)$ for state representing reputations of seller s_1 and s_2

$b(\langle .1, .1\rangle) = 1/9$	$b(\langle .1, .5\rangle) = 1/9$	$b(\langle .1, .9\rangle) = 1/9$
$b(\langle .5, .1\rangle) = 1/9$	$b(\langle .5, .5\rangle) = 1/9$	$b(\langle .5, .9\rangle) = 1/9$
$b(\langle .9, .1\rangle) = 1/9$	$b(\langle .9, .5\rangle) = 1/9$	$b(\langle .9, .9\rangle) = 1/9$

We adopt a simple observation function which defines advisor observations that are similar to a given state as likely and observations that are very different from a given state as unlikely. The observation function incorporates the *epistemic* uncertainty of the observation o by interpreting cf as the number of transactions the reputation is stemming from. We can formally define the observation function $O(o, s) = Pr(o|s)$, capturing the probability of an observation o given a state s as follows:

$$\begin{aligned} P(o = \langle\langle rep_{s_1}, cf_{s_1}\rangle, \langle rep_{s_1}, cf_{s_1}\rangle\rangle | s = \langle r_{s_1}, r_{s_2}\rangle) \\ = (r_{s_1})^{rep_{s_1} cf_{s_1}} (1 - r_{s_1})^{(1-rep_{s_1}) cf_{s_1}} \times \\ \times (r_{s_2})^{rep_{s_2} cf_{s_1}} (1 - r_{s_2})^{(1-rep_{s_2}) cf_{s_2}} \end{aligned} \quad (6)$$

Given the policy generated for the POMDP we will now step through an example of the kinds of actions that would be chosen based on the current belief state, the observation each action would generate and how the observation will influence the next belief state.

Action: *ask* a_1 leads to observation: $\langle\langle 0.5, 12\rangle, \langle 0.9, 20\rangle\rangle$

The policy weighs the small cost of asking an advisor, with the decrease in uncertainty about seller reputations. Given the relatively flat belief state, the buyer will have a higher expected satisfaction after asking advisors, since the eventual result will be a more informed purchase decision. For the purposes of this example, suppose that a_1 has been determined to be the best advisor to ask.

Once the *ask* action is taken the buyer observes the reputation rep_i and certainty factor cf_i given by the advisor for each seller i. Using the observation function, the buyer's new belief b' for each state is updated using Bayes' Rule as follows:

$$\begin{aligned} b'(s) &= P(s|o) \\ &= \frac{P(s)P(o|s)}{P(o)} \\ &= k \cdot b(s)O(s,o) \end{aligned} \quad (7)$$

where $k = \dfrac{1}{P(o)}$ is a constant normalization factor

The result of the belief update is shown in following table.

$b(\langle .1, .1\rangle) = .00 \quad b(\langle .1, .5\rangle) = .05 \quad b(\langle .1, .9\rangle) = .06$
$b(\langle .5, .1\rangle) = .01 \quad b(\langle .5, .5\rangle) = .36 \quad b(\langle .5, .9\rangle) = .42$
$b(\langle .9, .1\rangle) = .00 \quad b(\langle .9, .5\rangle) = .05 \quad b(\langle .9, .9\rangle) = .06$

Action: *ask* a_3 leads to observation: $\langle\langle 0.5, 4\rangle, \langle 0.8, 6\rangle\rangle$

Given the updated belief about actual state, the policy would once again dictate that our buyer should ask an advisor and specify the best advisor to ask. The advisor has a similar impression of the reputation of each seller, but is less certain about the reputations it holds. The buyer incorporates this observation updating its belief to the following.

$b(\langle .1, .1\rangle) = .00 \quad b(\langle .1, .5\rangle) = .00 \quad b(\langle .1, .9\rangle) = .01$
$b(\langle .5, .1\rangle) = .00 \quad b(\langle .5, .5\rangle) = .18 \quad b(\langle .5, .9\rangle) = .79$
$b(\langle .9, .1\rangle) = .00 \quad b(\langle .9, .5\rangle) = .00 \quad b(\langle .9, .9\rangle) = .01$

Action: *buy* s_2

At this point, there is enough of a peak in the belief space that the best action is to select a seller. The transition function captures what the outcome of a purchase will be given the actual state. Given the belief about the likelihood of each state, the policy chooses to buy from the seller that will maximize the expected reward based on buyer satisfaction.

In our case, the expected reward for buying from s_2 will be far higher than that of s_1 and the policy chooses to buy from s_2. The buyer will then complete the transaction, purchasing from s_2 and evaluating the result of the purchase, comparing the expected outcome with the actual outcome, to gauge satisfaction.

This example demonstrates how a policy can lead a buyer to gather information from advisors. It also shows how to use this information with a simple observation function to update its belief about the true reputation of the sellers and when the *epistemic* uncertainty is low enough, buy from the seller that will lead to the highest expected satisfaction.

The Advisor-POMDP can use a state with an $\vec{r}$ component representing seller reputations that take on a small set discrete integral values between 0 and 1 as in our example. However, to more accurately model the reputation of sellers a real value for the reputation can be used. The next section discusses some of the challenges of calculating a policy given the way in which we have defined our POMDP.

4.4. Calculating Policies

Partially Observable Markov Decision Processes provide a principled and expressive framework for reputation modeling; however calculating optimal policies for POMDPs is not a trivial exercise. There are two broad classes of algorithms for finding optimal POMDP policies. The first class, known as *value iteration*, uses dynamic programming to indirectly build an optimal policy by incrementally computing the optimal value function. The idea is to compute the optimal value function for each time step starting from the end and going backwards in time. More precisely, the optimal value function $V^n(b)$ for n steps-to-go is the one that yields the highest possible rewards for the n remaining steps. We can recursively compute the n-step optimal value function V^n from the $n-1$-step optimal value function V^{n-1} using Bellman's equation:

$$V^n(b) = \max_a R^a(b) + \sum_o \Pr(o|b,a) V^{n-1}(b') \quad (8)$$

where

$$R^a(b) = \sum b(s) R^a(s)$$
$$Pr(o|b,a) = \sum_s b(s) \Pr(s'|s,a) \Pr(o|s')$$

The value b' is the updated belief state after executing a and observing o.

Note that by remembering the action a that maximizes the right-hand-side of Bellman's equation, we indirectly obtain a policy. While Bellman's equation gives us a dynamic programming algorithm that can be used in theory to find optimal policies, in practice, the continuous belief space is problematic since it is not possible to apply Bellman's equation to an infinite number of belief points. Fortunately, as pointed out by Smallwood and Sondik [27], optimal value functions are *piecewise-linear and convex*, which can be exploited to apply Bellman's equation a finite number of times (once per linear piece of the value function). Alternatively,

approximate solutions can be efficiently computed by *point-based value iteration* algorithms [17, 28].

The second class of algorithms, often referred as *policy search* techniques, seek to directly optimize a policy. Popular approaches include policy iteration [6, 19] and gradient descent [14, 16], which incrementally improve a policy by searching for possible modifications that could improve the value of the policy.

Most of the value iteration and policy search techniques are designed to find policies for POMDPs with discrete states. Similar to the example in Section 4.3, we can often simplify the space of reputations to a few possible representative reputations, essentially discretizing the state space. While this sidesteps the continuous nature of the state space, the number of resulting discrete states can grow exponentially with the number of sellers. Nevertheless, techniques such as VDCBPI [20] or Perseus+ADD [18] can often exploit problem-specific structural properties to work with compressed yet lossless state representations, which may be polynomial (instead of exponential) with respect to the number of sellers.

If we do not restrict the space of reputations to a few representatives, several model free approaches have been proposed that can deal with continuous state spaces [14, 16, 1]. These approaches optimize policies by stochastic simulations, essentially circumventing the continuous nature of the state space. However, it is unclear how much simulation will be required in practice. Alternatively, the Advisor-POMDP can also be viewed as a special case of *Bayesian Reinforcement Learning* [13], for which the design of efficient algorithms is an active area of research [5, 23].

The investigation and implementation of efficient algorithms for the Advisor-POMDP is an important direction of future research.

5. Conclusion

This work examines the problem of reasoning under the uncertainty present in social reputation systems for electronic markets with buying and selling agents. A brief survey of other reputation models was presented and the degree to which they satisfy the requirements of a social reputation system was analyzed.

The main contribution of this paper is the Advisor-POMDP, a decision theoretic framework in which a buyer can ask other advisors to accumulate information about a seller's reputation and eventually make an informed purchase. This framework captures both the *stochastic* and *epistemic* uncertainty that is inherent in the problem posed.

The Advisor-POMDP allows buyers to make decisions about which sellers to make purchases from using information about a seller's reputation. Since the past fulfillment of the seller's role of providing quality goods is taken into account, sellers are discouraged from neglecting this role and our reputation system serves as a partial mechanism for ensuring good market behavior and further engendering trust on behalf of all agents involved.

In this paper, we have argued that a POMDP approach is effective for buying agents to determine whether to do business with a selling agent. In particular, we discuss how buying agents can elect to ask other buying agents for advice, in order to improve their decision making ability regarding appropriate business partners. This research therefore provides a basis for incorporating a social reputation system within the electronic marketplace, fostering trust between agents and resulting in buying agents that will themselves be trustworthy for their users.

6. Future Work

The Advisor-POMDP defined here is preliminary and the bulk of future work will center around developing methods for extracting usable policies using reinforcement learning methods while taking care to limit the amount of sampling necessary. Some subset of the approaches listed in Section 4.4 need to be adapted to our specific POMDP instance and an analysis done to gauge the complexity of finding policies given the large state space. There is some hope that we may be able to exploit structure that is specific to the Advisor-POMDP to limit the potential policies that must be evaluated. Once a reasonable approach to finding policies is implemented, an empirical analysis of the Advisor-POMDP will be undertaken comparing the policies generated to simpler heuristic approaches.

Another topic for future work is to develop more precise strategies for modeling the advisors in the marketplace and to use these models to adjust the belief values that are calculated following observations from these advisors. Regan and Cohen [22] identify two issues that must be overcome in a social reputation system namely, deception and subjectivity. Each advisor can deceive the buyer when offering information. Furthermore each advisor may be using standards for determining whether a purchase is satisfactory that are specific to the advisor. The deceptiveness of an advisor can be modeled using the observation function by making observations that match a given state less likely. To model the higher standards of a particular advisor, an observation function can be used in which the observations from that advisor that are slightly lower than the actual state are most likely.

In order to address the issue of possibly deceptive advisors, it is in fact ideal to develop some kind of mechanism within the marketplace [30] in order for buyers to be more inclined to share correct information with other buyers. Another avenue for future work is to articulate and integrate some kind of mechanism to foster more trustworthy social

networks of buying agents.

References

[1] D. Aberdeen and J. Baxter. Scaling internal-state policy-gradient methods for POMDPs. In *Proceedings of the Nineteenth International Conference on Machine Learning*, pages 3–10, Sydney, Australia, 2002.

[2] K. S. Barber and J. Kim. Belief revision process based on trust: Agents evaluating reputation of information sources. In *Proceedings of the workshop on Deception, Fraud, and Trust in Agent Societies held during the Autonomous Agents Conference*, pages 73–82, 2001.

[3] J. Carter and A. A. Ghorbani. Towards a formalization of trust. *Web Intelligence and Agent Systems*, 2(3):167–183, March 2004.

[4] A. P. Dempster. A generalization of bayesian inference (with discussion). *Journal of the Royal Statistical Society*, 30:205–247, 1968.

[5] M. Duff. *Optimal Learning: Computational procedures for Bayes-adaptive Markov decision processes*. PhD thesis, University of Massassachusetts Amherst, 2002.

[6] E. A. Hansen. Solving POMDPs by searching in policy space. In *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, pages 211–219, Madison, Wisconsin, 1998.

[7] A. Jøsang and R. Ismail. The beta reputation system. 15th Bled Electronic Commerce Conference e-Reality: Constructing the e-Economy, June 2002.

[8] R. Jurca and B. Faltings. "confess": Eliciting honest feedback without independent verification authorities. *Sixth International Workshop on Agent Mediated Electronic Commerce (AMEC VI)*, 2004.

[9] L. Kaelbling, M. Littman, and A. Cassandra. Planning and acting in partially observable stochastic domains. *Artificial Intelligence*, 101, 1998.

[10] N. Littlestone and M. K. Warmuth. The weighted majority algorithm. *Information and Computation*, 108(2):212–261, 1994.

[11] W. S. Lovejoy. A survey of algorithmic methods for partially observed Markov decision processes. *Annals of Operations Research*, 28:47–66, 1991.

[12] S. Marti and H. Garcia-Molina. Limited reputation sharing in p2p systems. In *EC '04: Proceedings of the 5th ACM conference on Electronic commerce*, pages 91–101, 2004.

[13] J. Martin. *Bayesian decision problems and Markov chains*. John Wiley & Sons, New York, 1967.

[14] N. Meuleau, K.-E. Kim, L. P. Kaelbling, and A. R. Cassandra. Solving POMDPs by searching the space of finite policies. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pages 417–426, Stockholm, 1999.

[15] L. Mui, M. Mohtashemi, C. Ang, P. Szolovits, and A. Halberstadt. Ratings in distributed systems: A bayesian approach. In *Workshop on Information Technologies and Systems (WITS'2001). AAAI, 2001*, 2001.

[16] A. Y. Ng and M. Jordan. PEGASUS: A policy search method for large MDPs and POMDPs. In *Proceedings of the Sixteenth Conference on Uncertainty in Artificial Intelligence*, pages 406–415, Stanford, CA, 2000.

[17] J. Pineau, G. Gordon, and S. Thrun. Point-based value iteration: an anytime algorithm for pomdps. In *Proceedings of IJCAI-03*, 2003.

[18] P. Poupart. *Exploiting Structure to efficienty solve large scale partially observable Markov decision processes*. PhD thesis, Department of Computer Science, University of Toronto, 2005.

[19] P. Poupart and C. Boutilier. Bounded finite state controllers. In *Advances in Neural Information Processing Systems*, Vancouver, BC, 2003.

[20] P. Poupart and C. Boutilier. Vdcbpi: An approximate scalable algorithm for large POMDPs. In *Advances in Neural Information Processing Systems*, pages 1081–1088, Vancouver, BC, 2004.

[21] M. L. Puterman. *Markov Decision Processes*. Wiley, 1994.

[22] K. Regan and R. Cohen. A model of indirect reputation assessment for adaptive buying agents in electronic markets. In *Proceedings of the Business Agents and Semantic Web (BASeWEB05)*, 2005.

[23] N. F. Richard Dearden and D. Andre. Model based Bayesian exploration. In *Proceedings of the Fifteenth Conference on Uncertainty in Artificial Intelligence*, pages 150–159, Stockholm, 1999.

[24] J. Sabater and C. Sierra. Regret: reputation in gregarious societies. In *AGENTS '01: Proceedings of the fifth international conference on Autonomous agents*, pages 194–195, 2001.

[25] K. Sentz and S. Ferson. Combination of evidence in dempster-shafer theory. Technical report, Sandia National Laboratories, 2003.

[26] G. Shafer. *A mathematical theory of evidence*. Princeton University Press, 1976.

[27] R. D. Smallwood and E. J. Sondik. The optimal control of partially observable Markov processes over a finite horizon. *Operations Research*, 21:1071–1088, 1973.

[28] M. T. J. Spaan and N. Vlassis. A point-based pomdp algorithm for robot planning. 2004.

[29] T. Tran and R. Cohen. Improving user satisfaction in agent-based electronic marketplaces by reputation modelling and adjustable product quality. In *AAMAS04*, New York, USA, July 2004.

[30] H. Varian. Economic mechanism design for computerized agents. In *Proceedings of the First USENIX Workshop on Electronic Commerce*, July 1995.

[31] Y. Wang and J. Vassileva. Bayesian network-based trust model. In *IEEE/WIC International Conference on Web Intelligence (WI2003)*, 2003.

[32] B. Yu and M. Singh. Detecting deception in reputation management. In *AAMAS03*, pages 73–80, 2003.

[33] G. Zacharia, A. Moukas, and P. Maes. Collaborative reputation mechanisms in electronic marketplaces. In *32nd Hawaii International Conference on System Sciences*, 1999.