

Reasoning under Uncertainty in Social Reputation Systems: The Advisor-POMDP

Kevin Regan

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-Shafer 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 the ease which trust provides is a business agreement using a handshake instead of legal contracts. The success of trust in making transactions more efficient in traditional markets motivates the search for a comparable measure of trust in emerging electronic markets which can be populated by agents who automate the transactions between buyers and sellers. We will examine trust from the perspective of the agent required to decide who to do business with based in part on trust. Reputation systems can aid in these trust decisions by providing a reputation for each agent essentially modeling how trustworthy they were in the past.

The aim of this paper is to construct a robust framework for buyers to choose the best seller based on some measure of reputation in a market consisting of autonomous agents. We will 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 will take a close look at how reputation can be modeled so as to capture the most relevant aspects of the reputation system. We will then present a framework based in Bayesian decision theory and give a brief discussion of some of the challenges to the efficient functioning of it.

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 eBay¹ and amazon.com Auctions² and theoretical models such as the CONFESS ?. 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 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 have 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 construct a personal model of a selling agent's reputation which is constructed using only the transactions in which the individual buying agent has been involved with. An example of this approach is the reputation model developed by Cohen and Tran (?) in which a buying agent uses only their past purchases from sellers to learn to avoid dishonest sellers.

The advantage to only using transactions the buying agent has been involved with is no uncertainty regarding the outcome of those transactions. However, with this approach the buying agent is limited to modeling only those sellers who the buying agent has purchased from 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. We will 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 ?????, 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

¹www.ebay.com

²auctions.amazon.com

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 ? as:

Stochastic Uncertainty - uncertainty which results from the randomness of a system.

Described elsewhere as: irreducible, aleatory, or objective uncertainty as well as variability

Epistemic Uncertainty - uncertainty which results from a lack of knowledge about a system

Described elsewhere 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 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 collect from other buyers would not take into account the fact that some of the other buyers may have lied in the past and 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, to systematic models which rely on Dempster-Shafer Theory ?? 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 ? which uses the following complicated expression to update the single reputation value of a seller:

$$R_{t+1} = \frac{1}{\Theta} \sum_1^t \left(1 - \frac{1}{1 + e^{-\frac{(R-D)}{\sigma}}} \right) \cdot R_{i+1}^{other} \cdot \left(W_{i+1} - \frac{R_{t+1}}{D} \right) \quad (1)$$

A full understanding of the preceding expression is not necessary; the aim is simply to demonstrate the complexity of some ad-hoc reputation models. While complex, expression ?? does allow 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, there is 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 an lower and upper bound on the likelihood of a hypothesis represented by A .

The reputation system developed by Yu and Singh ? should help make our discussion of DST concrete and illustrate how DST can be used to model reputation. They define $\{T, -T\}$ 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. $m(\{-T\})$ 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. $m(\{T, -T\})$ 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 ? 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 ? to allow for reputations with different weights to be combined.

The reputation model developed by Yu and Singh ? 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. This is a Bayesian approach, since we are representing our *belief* about the future using probability. At first glance it does not capture the *epistemic* uncertainty since if we had seen 800 heads and 200 tails, the probability 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 we can capture both classes of uncertainty.

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 buying 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:

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

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 ?? shows a beta distribution given by Jøsang and Ismail ? for process in which x has been observed 7 times and \bar{x} has been observed once which gives us $f(p|8, 2)$.

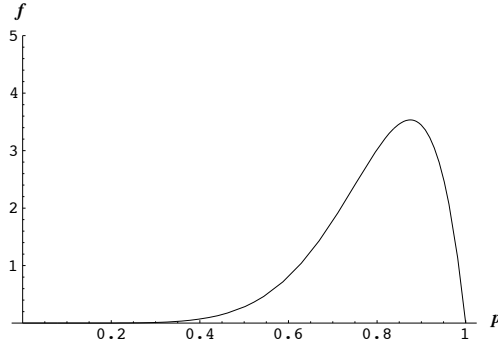


Figure 1: Beta function $f(p|8, 2)$

Jøsang and Ismail develop the Beta Reputation System ? 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 Y³. 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:

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

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

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 ?. Like the DST model presented

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

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. [?] develop a similar model based on the beta distribution, but do not include mention of methods for combining and weighing information from other buying agents. Barber and Kim [?] use a Bayesian belief 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.

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 [?] and others [??] 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.

4 Decision Framework

4.1 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. 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 then assign utilities to possible events and use these utilities to decide on the best possible action. A natural way to model the decision making process given uncertainty about possible utility is a Markov Decision Process.

The Markov Decision Process (MDP) is composed of 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. However, our buyer has only partial knowledge of the underlying *stochastic* process since it only has information about a subset of a sellers past interactions. We can model this *epistemic* uncertainty by extending the MDP to a Partially Observable Markov Decision Process (POMDP) which places a belief distribution over the possible states and uses observations to adjust this belief.

We will now construct what we call the Advisor-POMDP (which is partially illustrated in Figure ??). Our Advisor-POMDP is defined by the tuple $\langle S, A, T, R, \Omega, O \rangle$ for which each element is defined as follows:

S - State

The states of our POMDP are a set of real values in the range $[0,1]$ representing the reputations of each seller. Also a part of the state is an implicit real value representing the outcome of a purchase. This implicit *satisfaction* value is 0 except in states that have been reached from the *buy* action. This state can be interpreted as a model of the *stochastic* process from which the outcome of a possible transaction with each of the sellers is drawn. The knowledge represented by the state is from the perspective of all the advisors who have responded with information.

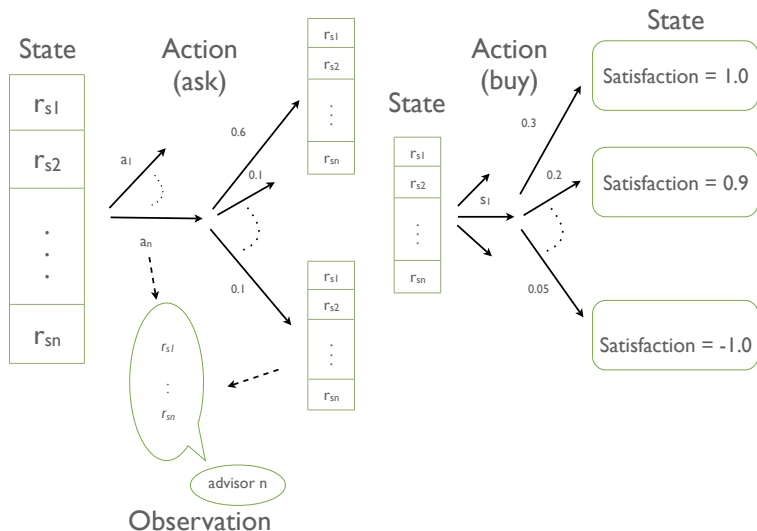


Figure 2: A partial transition diagram of the Advisor-POMDP

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

From each state the *ask* action will transition to a state representing the updated reputation information held by our buying agent after asking a particular advisor. The *buy* action will transition to a state where the *satisfaction* value represents the outcome of the purchase.

R - Reward

For states that have been reached through an *ask* action (where the *satisfaction* value is 0) there is a small negative reward. States that have been reached through the *buy* action have a non-zero *satisfaction* value and have a large reward corresponding to the *satisfaction* value. This reward will be a large negative value for transactions in which the buying agent was unhappy with its purchase and the reward will be a large positive value when the buyer agent is be happy with the purchase.

Ω - Observations

The observations in our POMDP composed of the information received by our buying agent in response to asking advisors. This information take on the form of a set of seller reputation values in the range $[0,1]$.

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 a measure of how the buyer interprets the information given by each advisor.

Each state in our Advisor-POMDP represents everything the buying agent knows about the market. Because information from previous states does not provide any information beyond what is contained in the current state, or influence the buying agent's next action and state, we can say that our Advisor-POMDP obeys the Markov property and we can make use of a wealth of methods for solving POMDPs. The solution to the Advisor-POMDP would yield a policy specifying the action to take given a belief about the state the agent is in. The way in which we have specified our rewards ensures that the policy will attempt to maximize the

buyer’s satisfaction with a purchase while minimizing the number of advisors asked.

In order to find a desirable policy some significant challenges must first be overcome. A buyer who is new to the market cannot be assumed to have knowledge of the dynamics of the market and so the observation and transition functions of the POMDP may not be specified. In the next section we will discuss possible approaches to cope with a lack of knowledge about the observation and transition functions, but first we will outline how a policy can be calculated in the simplest case.

4.2 Some Issues

Assuming we knew both the observation and transition functions of our Advisor-POMDP we could use simple value iteration (as described by Kaelbling et al. [1]) to compute an optimal policy, however, this would take a prohibitive amount of time and space. A single iteration of the value iteration algorithm can have space complexity in the order of $|A| \cdot |V'|^{|O|}$ where $|V|$ is the number of policy trees generated in the previous step of the algorithm. This can be improved on by using algorithms based on Point-Based Value Iteration [2] which approximate the exact value iteration method by choosing a small set of representative belief points and calculating the value of these belief points and their derivative. However, the assumption that the observation and transition functions are known is not a realistic one, since this would imply, among other things, that we knew which advisors were most knowledgeable in advance.

To determine the value of possible policies without knowledge of the dynamics of the environment (in this case the observation and transition functions) we can use some of the tools provided by reinforcement learning. One such tool is that of gradient ascent. Essentially gradient ascent begins by constructing a class of parameterized policies and finding the parameters θ that optimize $\eta(\theta)$ the expected discounted reward when following a given policy. The optimization is achieved by finding the gradient $\nabla\eta(\theta)$ with respect to the parameters θ and taking a step in the uphill direction by adding $\gamma \cdot \nabla\eta(\theta)$ to θ , where γ is some step size. This gradient can be found exactly for POMDPs with small state spaces, however the state space of the Advisor-POMDP can grow quite large. This could be addressed by using the GPOMDP algorithm developed by Baxter and Bartlett [3] which provides a way to approximate the gradient and calculate local optima in the using conjugate-gradient procedures.

One of the issues with the gradient ascent approach is that it requires significant exploration of the environment as for each possible policy many samples must be taken. In the context of our social reputation model each sample would translate into the purchase and we would like to limit the purchases required while finding the best policy for our buying agent. One approach to limiting the samples of the environment uses what Peshkin and Shelton call a proxy environment [4]. This proxy environment is positioned between an agent and the real environment and uses likelihood ratio estimation to reuse data gathered from one policy to estimate the results of following another. The result is that existing reinforcement learning algorithms such as the one proposed by Baxter and Bartlett [3] can be plugged into the proxy environment to reduce the number samples needed. Peshkin and Shelton present a simple method to balance exploration versus exploitation in their proxy environment but mention that more sophisticated methods such as maintaining a distribution over policies could be used.

5 Conclusion

This work examines the problem of reasoning under the uncertainty present in social reputation system for electronic markets with buying and selling agents. A brief survey of other reputation models was presented and degree to which 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 a purchase. This framework captures both the *stochastic* and *epistemic* uncertainty that is inherent in the problem posed.

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 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 to be undertaken and the policies generated compared to simpler heuristic approaches.