*Computational Intelligence,* Volume 20, Number 4, 2004

# THE ADVANTAGES OF DESIGNING ADAPTIVE BUSINESS AGENTS USING REPUTATION MODELING COMPARED TO THE APPROACH OF RECURSIVE MODELING

THOMAS TRAN AND ROBIN COHEN
*School of Computer Science, University of Waterloo, Waterloo, ON Canada*

Adaptive business agents operate in electronic marketplaces, learning from past experiences to make effective decisions on behalf of their users. How best to design these agents is an open question. In this article, we present an approach for the design of adaptive business agents that uses a combination of reinforcement learning and reputation modeling. In particular, we take into account the fact that multiple selling agents may offer the same good with different qualities, and that selling agents may alter the quality of their goods. We also consider the possibility of dishonest agents in the marketplace. Our buying agents exploit the reputation of selling agents to avoid interaction with the disreputable ones, and therefore to reduce the risk of purchasing low value goods. We then experimentally compare the performance of our agents with those designed using a recursive modeling approach. We are able to show that agents designed according to our algorithms achieve better performance in terms of satisfaction and computational time and as such are well suited for the design of electronic marketplaces.

*Key words:* electronic commerce, adaptive business agents, trust of business agents, multi-agent systems, reinforcement learning, reputation modeling.

## 1. INTRODUCTION

A topic that is open to much debate is how best to design an electronic marketplace and how best to construct the algorithms used by buying and selling agents in that marketplace, to equip these agents with the capability of bringing satisfaction to their users. To promote the use of agent technology to do business via the Internet, it is important to develop algorithms for the behavior of these agents that engender trust in their human users.

One strategy that has been proposed as an effective method for enabling buying and selling agents to learn from past experiences is the recursive modeling approach of Vidal (1998) and Durfee (1996). In this model, there is a hierarchy of possible levels of modeling, where a 0-level agent learns only from observations about the environment and from any environmental rewards it receives, 1-level agents model others as 0-level agents, 2-level agents model others as 1-level agents, and so on. Because of the computational costs associated with maintaining deeper models, the challenge is to determine when it is best to stop maintaining these deeper models (so that the costs of the recursive modeling are balanced against the possible gains).

In this article, we present an alternative approach to the design of electronic marketplaces. In particular, we propose that selling agents be allowed to alter the quality of their goods to best meet the needs of buyers. We further propose that buying agents use a combination of reinforcement learning and reputation modeling, in order to make effective buying decisions, including the possible recognition of dishonest selling agents. Our proposed model is described in detail with specific discussion of how best to model and adjust the reputation ratings of selling agents.

We then present experimental results comparing our model with the recursive modeling strategy of Vidal and Durfee (1996). We are in fact able to demonstrate that buying and selling agents designed according to our algorithms deliver better satisfaction to their users, and in better computational time, and are thus well equipped to engender the trust of their users. Moreover, we clarify how our approach allows agents to perform successfully in open,

Address correspondence to Thomas Tran, School of Computer Science, University of Waterloo, Waterloo, Canada; e-mail: tt5tran@uwaterloo.edu

dynamic, uncertain, and untrusted electronic marketplaces. The market model and algorithms proposed in our research are therefore promoted as an effective strategy for designing adaptive business agents.

## 2. THE PROPOSED ALGORITHMS

In this section, we present our agent market model and propose the learning algorithms for buying and selling agents in electronic marketplaces, based on reputation modeling and reinforcement learning. The algorithms presented here extend an earlier version of the model, outlined in Tran and Cohen (2002). In particular, we have developed more effective strategies both for modeling reputation and for adjusting reputation ratings. These differences are reflected in the descriptions of Section 2.2 and are described in more detail in Section 2.4.

### 2.1. The Agent Market Model

We model the agent environment as an open marketplace that is populated with economically motivated agents. The nature of an open marketplace allows the economic agents, which we classify as *buyers* and *sellers*, to freely enter or leave the market. Buyers and sellers are self-interested agents whose goal is to maximize their own benefit.

Our market environment is rooted in an information delivery infrastructure such as the Internet, which provides agents with virtually direct and free access to all other agents. The process of buying and selling goods is realized via a *contract-net* like mechanism (Smith 1980; Davis and Smith 1983) which consists of three elementary phases: ($i$) when a buyer $b$ is in need of some good $g$, it will announce its request for that good to all the sellers, using multi-cast or possibly broadcast. (*ii*) After receiving the request from $b$, those sellers that have good $g$ available for sales will send a message to $b$, stating their price bids for delivering the good. (*iii*) Buyer $b$ evaluates the submitted bids and selects a suitable seller to purchase good $g$. Buyer $b$ then pays the chosen seller and receives the good from that seller. Thus, the buying and selling process can be viewed as an *auction* where sellers play the role of bidders and buyers play the role of auctioneers, and a seller is said to be *winning the auction* if it is able to sell its good to the buyer.

To make our marketplace more realistic and also more interesting, we assume that

- the quality of a good offered by different sellers may not be the same, and a seller may alter the quality (in addition to the price) of its goods (dynamic market).
- a buyer can examine the quality of the good it purchases only after it receives that good from the selected seller (uncertain market).
- it is possible that some dishonest sellers exist in the market (untrusted market).
- each buyer has some way to evaluate the good it purchases, based on the price and the quality of the good received.

### 2.2. Buying Algorithm

Consider a scenario where a buyer $b$ announces its request for some good $g$. Let $G$ be the set of goods, $P$ be the set of prices, and $S$ be the set of all sellers in the marketplace. $G$, $P$, and $S$ are finite sets.

Buyer $b$ models the reputation of all sellers in the market using function $r^b : S \mapsto (-1, 1)$, which is called the *reputation function* of $b$. Initially, buyer $b$ sets the *reputation rating* $r^b(s) = 0$ for every seller $s \in S$. After each transaction with a seller $s$, buyer $b$ will update (increase or decrease) $r^b(s)$ depending on whether or not $s$ satisfies $b$ in the transaction. A seller $s$ is considered *reputable* by buyer $b$ if $r^b(s) \geq \Theta$, where $\Theta$ is buyer $b$'s *reputation threshold* ($0 < \Theta < 1$). A seller $s$ is considered *disreputable* by buyer $b$ if $r^b(s) \leq \theta$, where $\theta$ is buyer $b$'s *disreputation threshold* ($-1 < \theta < 0$). A seller $s$ with $\theta < r^b(s) < \Theta$ is neither reputable nor disreputable to buyer $b$. In other words, $b$ does not have enough information to decide on the reputation of $s$. Let $S_r^b$ and $S_{dr}^b$ be the sets of reputable and disreputable sellers to buyer $b$, respectively, i.e.,

$$S_r^b = \{s \in S \mid r^b(s) \geq \Theta\} \subseteq S, \tag{1}$$

and

$$S_{dr}^b = \{s \in S \mid r^b(s) \leq \theta\} \subseteq S. \tag{2}$$

Buyer $b$ will focus its business on the reputable sellers and stay away from the disreputable ones.

Buyer $b$ estimates the expected value of the goods it purchases using the *expected value function* $f^b : G \times P \times S \mapsto \mathbb{R}$. Hence, the real number $f^b(g, p, s)$ represents buyer $b$'s expected value of buying good $g$ at price $p$ from seller $s$.

However multiple sellers may offer good $g$ with different qualities and a seller may alter the quality of its goods, buyer $b$ puts more trust in the sellers with good reputation. Thus, it chooses among the reputable sellers in $S_r^b$ a seller $\hat{s}$ that offers good $g$ at price $p$ with maximum expected value:

$$\hat{s} = \arg \max_{s \in S_r^b} f^b(g, p, s), \tag{3}$$

where arg is an operator such that arg $f^b(g, p, s)$ returns $s$.

If no sellers in $S_r^b$ submit bids for delivering $g$ (or if $S_r^b = \emptyset$), then buyer $b$ will have to choose a seller $\hat{s}$ from the nonreputable sellers, provided that $\hat{s}$ is not a disreputable seller:

$$\hat{s} = \arg \max_{s \in (S - (S_r^b \cup S_{dr}^b))} f^b(g, p, s). \tag{4}$$

In addition, with a small probability $\rho$, buyer $b$ chooses to explore (rather than exploit) the marketplace by randomly selecting a seller $\hat{s} \in (S - S_{dr}^b)$. This gives buyer $b$ an opportunity to discover new reputable sellers. Initially, the value of $\rho$ should be set to 1, then decreased over time to some fixed minimum value determined by $b$.

After paying seller $\hat{s}$ and receiving good $g$, buyer $b$ can examine the quality $q \in Q$ of good $g$, where $Q$ is a finite set of real values representing product qualities. It then calculates the true value of good $g$ using the *true product value function* $v^b : G \times P \times Q \mapsto \mathbb{R}$. For instance, if buyer $b$ considers the quality of good $g$ to be twice more important than its price, it may set $v^b(g, p, q) = 2q - p$.

The expected value function $f^b$ is now incrementally learned in a reinforcement learning framework:

$$\Delta = v^b(g, p, q) - f^b(g, p, \hat{s}), \tag{5}$$

$$f^b(g, p, \hat{s}) \leftarrow f^b(g, p, \hat{s}) + \alpha\Delta, \tag{6}$$

where $\alpha$ is called the *learning rate* ($0 \leq \alpha \leq 1$). Similar to $\rho$, the learning rate $\alpha$ should initially be set to a starting value of 1 and then reduced over time to a fixed minimum value chosen depending on individual buyers.

Thus, if $\Delta = v^b(g, p, q) - f^b(g, p, \hat{s}) \geq 0$ then $f^b(g, p, \hat{s})$ is updated with the same or a greater value than before. This means that seller $\hat{s}$ has a good chance to be chosen by buyer $b$ again if it continues offering good $g$ at price $p$ in the next auction. Conversely, if $\Delta < 0$

then $f^b(g, p, \hat{s})$ is updated with a smaller value than before. This implies that seller $\hat{s}$ may not be selected by buyer $b$ in the next auction if it continues selling good $g$ at price $p$.

In addition to updating the expected value function, the reputation rating $r^b(\hat{s})$ of seller $\hat{s}$ also needs to be updated. Let $\vartheta^b(g) \in \mathbb{R}$ be the product value that buyer $b$ demands for good $g$. In other words, the demanded product value $\vartheta^b(g)$ is buyer $b$'s threshold for the true product value $v^b(g, p, q)$. We use a reputation updating scheme motivated by that proposed in Yu and Singh (2000) as follows:

If $v^b(g, p, q) - \vartheta^b(g) \geq 0$, that is, if seller $\hat{s}$ offers good $g$ with value greater than or equal to the value demanded by buyer $b$, then its reputation rating $r^b(\hat{s})$ is increased by

$$r^b(\hat{s}) \leftarrow \begin{cases} r^b(\hat{s}) + \mu(1 - r^b(\hat{s})) & \text{if } r^b(\hat{s}) \geq 0, \\ r^b(\hat{s}) + \mu(1 + r^b(\hat{s})) & \text{if } r^b(\hat{s}) < 0, \end{cases} \tag{7}$$

where $\mu$ is a positive factor called the *cooperation factor*[1] ($\mu > 0$).

Otherwise, if $v^b(g, p, q) - \vartheta^b(g) < 0$, that is, if seller $\hat{s}$ sells good $g$ with value less than that demanded by buyer $b$, then its reputation rating $r^b(\hat{s})$ is decreased by

$$r^b(\hat{s}) \leftarrow \begin{cases} r^b(\hat{s}) + \nu(1 - r^b(\hat{s})) & \text{if } r^b(\hat{s}) \geq 0, \\ r^b(\hat{s}) + \nu(1 + r^b(\hat{s})) & \text{if } r^b(\hat{s}) < 0, \end{cases} \tag{8}$$

where $\nu$ is a negative factor called the *noncooperation factor*[2] ($\nu < 0$).

The set of reputable sellers to buyer $b$ now needs to be updated based on the new reputation rating $r^b(\hat{s})$, as in one of the following two cases:

– If ($\hat{s} \in S_r^b$) and ($r^b(\hat{s}) < \Theta$) then buyer $b$ no longer considers $\hat{s}$ as a reputable seller, i.e.,

$$S_r^b \leftarrow S_r^b - \{\hat{s}\}. \tag{9}$$

– If ($\hat{s} \notin S_r^b$) and ($r^b(\hat{s}) \geq \Theta$) then buyer $b$ now considers $\hat{s}$ as a reputable seller, i.e.,

$$S_r^b \leftarrow S_r^b \cup \{\hat{s}\}. \tag{10}$$

Finally, the set of disreputable sellers also needs to be updated:

– If ($\hat{s} \notin S_{dr}^b$) and ($r^b(\hat{s}) \leq \theta$) then buyer $b$ now considers $\hat{s}$ as a disreputable seller, i.e.,

$$S_{dr}^b \leftarrow S_{dr}^b \cup \{\hat{s}\}. \tag{11}$$

**Setting $\mu$ and $\nu$:** The cooperation and noncooperation factors, $\mu$ and $\nu$, are used to adjust the reputation ratings of sellers once the buyer has examined the quality of the good purchased.

To protect itself from dishonest sellers,[3] buyer $b$ may require $|\nu| > |\mu|$ to implement the traditional assumption that reputation should be difficult to build up, but easy to tear down. Moreover, buyer $b$ may vary $\mu$ and $\nu$ as increasing functions of the true product value $v^b$ to reflect the common idea that a transaction with higher value should be more appreciated than a lower one (i.e., the reputation rating of a seller that offers a higher true product value should be better increased and vice versa).

[1] Buyer $b$ will consider seller $\hat{s}$ as being *cooperative* if the good $\hat{s}$ sells to $b$ has value greater than or equal to that demanded by $b$.

[2] Buyer $b$ will consider seller $s$ as being *noncooperative* if the good $\hat{s}$ sells to $b$ has value less than that demanded by $b$.

[3] Dishonest sellers are those sellers who, for example, offer a good with high quality and then later offer the same good with very low quality. Some sellers may simply be unskilled in making effective decisions about setting the quality and price of their goods and as such may be more "misguided" than "dishonest."

In particular, we propose the following equations for the calculation of $\mu$ and $\nu$. If $v^b(g, p, q) - \vartheta^b(g) \geq 0$, we define the cooperation factor $\mu$ as

$$\mu = \begin{cases} \dfrac{v^b(g, p, q) - \vartheta^b(g)}{\Delta v^b} & \text{if } \dfrac{v^b(g, p, q) - \vartheta^b(g)}{\Delta v^b} > \mu_{\min}, \\ \mu_{\min} & \text{otherwise,} \end{cases} \tag{12}$$

where $\Delta v^b = v^b_{\max} - v^b_{\min}$ with $v^b_{\max}$ and $v^b_{\min}$ being the maximum and minimum value of the true product value function $v^b(g, p, q)$.[4] We prevent $\mu$ from becoming zero when $v^b(g, p, q) = \vartheta^b(g)$ by using the value $\mu_{\min}$.

However, if $v^b(g, p, q) - \vartheta^b(g) < 0$, we define the noncooperation factor $\nu$ as

$$\nu = \lambda\left(\frac{v^b(g, p, q) - \vartheta^b(g)}{\Delta v^b}\right), \tag{13}$$

where $\lambda$ is called the *penalty factor* ($\lambda > 1$) to implement the above-mentioned idea that $|\nu|$ should be greater than $|\mu|$. If applying equation (8) using $\nu$ as defined in (13) results in the updated value $r^b(\hat{s}) \leq -1$, that is, seller $\hat{s}$ is so noncooperative, then buyer $b$ will place $\hat{s}$ in the disreputable set $S^b_{dr}$ by setting $r^b(\hat{s}) = \theta$.

### 2.3. Selling Algorithm

Consider a scenario where a seller $s \in S$ has to decide on the price to sell some good $g$ to a buyer $b$. Let $B$ be the (finite) set of buyers in the marketplace, and let function $h^s : G \times P \times B \mapsto \mathbb{R}$ estimate the expected profit for seller $s$. Thus, the real number $h^s(g, p, b)$ represents the expected profit for seller $s$ if it sells good $g$ at price $p$ to buyer $b$. Let $c^s(g, b)$ be the cost of seller $s$ to produce good $g$ for buyer $b$. Note that seller $s$ may produce various versions of good $g$, which are tailored to meet the needs of different buyers. Seller $s$ will choose a price $\hat{p}$ greater than or equal to cost $c^s(g, b)$ to sell good $g$ to buyer $b$ such that its expected profit is maximized:

$$\hat{p} = \arg \max_{\substack{p \in P \\ p \geq c^s(g, b)}} h^s(g, p, b), \tag{14}$$

where in this case arg is an operator such that arg $h^s(g, p, b)$ returns $p$.

The expected profit function $h^s$ is learned incrementally using reinforcement learning:

$$h^s(g, p, b) \leftarrow h^s(g, p, b) + \alpha(\phi^s(g, p, b) - h^s(g, p, b)), \tag{15}$$

where $\phi^s(g, p, b)$ is the actual profit of seller $s$ if it sells good $g$ at price $p$ to buyer $b$, and is defined as follows:

$$\phi^s(g, p, b) = \begin{cases} p - c^s(g, b) & \text{if seller } s \text{ wins the auction,} \\ 0 & \text{otherwise.} \end{cases} \tag{16}$$

Thus, if seller $s$ does not win the auction then $(\phi^s(g, p, b) - h^s(g, p, b))$ is negative, and by (15), $h^s(g, p, b)$ is updated with a smaller value than before. This means that price $\hat{p}$ will probably not be chosen again to sell good $g$ to buyer $b$ in future auctions, but rather some lower price will. Conversely, if seller $s$ wins the auction then price $\hat{p}$ will probably be re-selected in future auctions.

[4] $v^b_{\max}$ and $v^b_{\min}$ are derived from the maximum and minimum elements of the finite sets $P$ and $Q$.

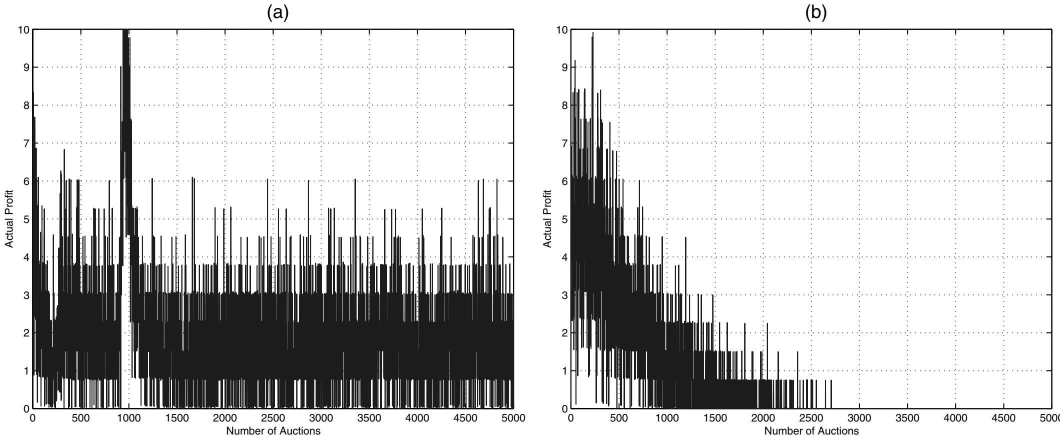


FIGURE 1. Profit values made by the dishonest sellers from a buyer using an early version of our model which did not implement the set of disreputable sellers (a), and from a buyer using the current version of the model (b).

If seller $s$ succeeded in selling good $g$ to buyer $b$ once, but subsequently fails for a number of auctions, say for $m$ consecutive auctions (where $m$ is seller $s$ specific constant), then it may not only be because $s$ has set a too high price for good $g$, but probably also because the quality of $g$ does not meet buyer $b$'s expectation. Thus, in addition to lowering the price via equation (15), seller $s$ may optionally add more value (quality) to $g$ by increasing its production cost:[5]

$$c^s(g, b) \leftarrow (1 + Inc)c^s(g, b), \tag{17}$$

where $Inc$ is seller $s$ specific constant called the *quality increasing factor*.

In contrast, if seller $s$ is successful in selling good $g$ to buyer $b$ for $n$ consecutive auctions, it may optionally reduce the quality of good $g$, and thus try to further increase its future profit:

$$c^s(g, b) \leftarrow (1 - Dec)c^s(g, b), \tag{18}$$

where $Dec$ is seller $s$ specific constant called the *quality decreasing factor*.

## 2.4. Design Decisions

This section details the extensions to our earlier version of the model, namely the introduction of the disreputable set and the proposed formulae for setting the cooperation factor $\mu$ and the noncooperation factor $\nu$.

An early version of our model was built with the reputation threshold $\Theta$ to form the set of reputable sellers, but without the disreputation threshold $\theta$ and its corresponding set of disreputable sellers (Tran and Cohen 2002). In an experiment later on we discovered that the proposed algorithm did not protect buyers well enough from dishonest sellers. Figure 1a shows the profit made by the dishonest sellers from a buyer following our early algorithm, obtained from an experimentation with large sized marketplaces.

[5]This supports the common assumption that it costs more to produce high-quality goods.

As we can clearly notice, the dishonest sellers were able to make profit from the buyer throughout the number of auctions tested. In other words, our early buyer could not avoid interaction with the dishonest sellers in the long run. The main reason is that although the buyer gave priority to considering reputable sellers in exploitation steps, it still chose the dishonest sellers in exploration steps, despite the fact that it had been cheated repeatedly by these sellers many times. To eliminate this undesirable situation, we introduced into our model the disreputation threshold $\theta$ to form the set of disreputable sellers, with whom the buyer would not interact even in exploration steps. Of course, $\theta$ should be set low enough so that the buyer would not mistakenly place any "innocent" seller in the disreputable set. We provide suggestions on reasonable values for $\theta$ in Section 4.5. Figure 1(b) shows the profit made by the dishonest sellers from a buyer using the current version of our model, in the same experimental settings as that of Figure 1(a), and demonstrates that the reputation mechanism provides much better protection for buyers. The dishonest sellers are no longer able to make profit from our proposed buyers in the long run.

A second extension to our earlier model is the introduction of specific formulae for setting the cooperation factor $\mu$ and the noncooperation factor $\nu$ (equations (12) and (13)). These proposed formulae implement two important assumptions: (*i*) The extent to which the reputation rating of a seller is increased or decreased should be based on the value of the transaction that it offers, and (*ii*) reputation should be difficult to build up but easy to tear down. These two assumptions are important because the first one encourages sellers to offer high value goods and also helps buyers to concentrate on the sellers that provide high-value goods, and the second one protects buyers from dishonest sellers in the market. In addition, we have theoretically proved that buyers setting $\mu$ and $\nu$ according to these formulae will not be harmed infinitely by dishonest sellers and therefore will not incur infinite loss, if they are cautious in setting their penalty factor $\lambda$.[6]

## 3. RELATED WORK

Reinforcement learning has been studied in various multi-agent problems (Littman 1994; Sen, Sekaran, and Hale 1994; Sandholm and Crites 1995; Ono and Fukumoto 1996). However, the agents and environments studied in these works are not economic agents and market environments. The reinforcement learning based strategies proposed in this article are, on the contrary, aimed at application domains where agents are economically motivated and act in open market environments.

A number of researchers have investigated the modeling of reputation. Yu and Singh (2000) develop a general model for trust, focusing on acquiring information from other agents in an agent community. Their scheme to update the trust rating of agents uses constant factors and does not take into consideration the extent to which an agent has (or has not) cooperated. In contrast, we have variable cooperative and noncooperative factors, to allow for agents who greatly disappoint to be more seriously penalized. We also specifically outline the strategies for adjusting the model of reputation within a setting of electronic marketplaces.

Perhaps, the most related to our research is Vidal and Durfee (1996) and Vidal (1998) where they develop strategies for trading agents using a recursive modeling approach. Their agents are divided into different classes depending on the agents' capabilities of modeling other agents. For instance, agents with 0-level models base their actions on the inputs and rewards they receive. Agents with 1-level models are those agents that model other agents as

[6] Details of the proof are presented in Thomas (2003), but omitted here due to the lack of space.

0-level agents. Agents with 2-level models are those that model others as 1-level agents. In theory, agents with high level models should fare better and could be recursively defined in the same manner. However, as pointed out in Vidal and Durfee (1996) and Vidal (1998), agents with deeper recursive models of others suffer from the computational costs associated with maintaining these deep models. In fact, the experimentation reported in Vidal and Durfee (1996) and Vidal (1998) is limited to only 1-level buyers and 2-level sellers. Moreover, the marketplace considered in Vidal and Durfee (1996) and Vidal (1998) does not allow for the sellers to alter the quality of their goods, nor does it address how to cope with dishonest sellers. In contrast, we model a marketplace where the quality of a good offered by different sellers may not be the same, sellers may alter the quality of their goods, and there is a possibility of having dishonest sellers in the market. To avoid heavy computational costs, we take a different approach, providing algorithms for buying agents that make use of a combination of reinforcement learning and reputation modeling techniques. Modeling sellers' reputation plays the role of a prescreening process, which partitions the set of sellers into three disjoint subsets, namely the reputable sellers, the disreputable sellers (including the dishonest sellers), and the neither reputable nor disreputable ones. Reinforcement learning is then applied to the set of reputable sellers (instead of all sellers) in exploitation steps, and to the nondisreputable sellers in exploration steps. This process helps buying agents to enhance their opportunity to purchase high-value goods from the reputable sellers, and reduce the risk of purchasing low-value goods from the disreputable sellers. In other words, reinforcement learning and reputation modeling work together as two layers of learning to improve the performance of buying agents. The algorithm we propose for selling agents enables them to learn to maximize their expected profits by not only adjusting product prices using reinforcement learning, but also by adjusting product quality to meet the buyers' specific needs. Since quality and price are the two most important factors based on which buying agents determine the value of the goods they purchase, the proposed selling algorithm obviously gives more opportunities for selling agents to make successful sales.

## 4. EXPERIMENTAL COMPARISON

We are interested in comparing the performance of our agents with those proposed in Vidal and Durfee (1996) and Vidal (1998), in terms of satisfaction and computational time.

### 4.1. Selecting Agents for Comparison

We would like to experimentally compare our buyers and sellers with 1-level buyers and 0-level sellers proposed in Vidal and Durfee (1996) and Vidal (1998), respectively. We choose these specific agents for comparison because of the following reasons:

– As explained in Vidal and Durfee (1996) and Vidal (1998), as a buyer receives bids from the sellers, there is no need for the buyer to try to out-guess or predict what the sellers will bid. The buyer is not concerned with what other buyers are doing either because it is assumed that there will be enough supply in the market. Thus, buyers do not need to keep models of others deeper than level 1. In other words, 1-level buyers are the buyers with deepest models of others. We were therefore interested in challenging our buyers with 1-level buyers.
– We would like to compare our sellers with 0-level sellers because both our sellers and 0-level sellers learn from the observations they make about the environment and from any environmental rewards they receive. In addition, it is not relevant to compare our

sellers with sellers of deeper levels (i.e., 1 or 2-level sellers), because these levels of sellers make use of two assumptions which, we think, are unrealistic and therefore do not implement in our market mechanism. These two assumptions are:

($i$) *The bid submitted by a seller to a buyer is known by other sellers in the market.* This assumption is unrealistic because the bid submitted by a seller to a buyer should be treated as private information between that seller and buyer; and therefore, should not be made known to other sellers in the marketplace. Moreover, a seller would not have any incentive or interest to broadcast the bid it is submitting to a buyer to all other sellers in the market.

($ii$) *The price accepted by a buyer at each auction is known by all sellers in the market.* This assumption is also unrealistic because the buyer would not want everybody know the price at which it purchases the good from a particular seller, and neither would the seller involved in the transaction; otherwise, its behaviors would be modeled and exploited by other sellers in the market.

### 4.2. How 1-Level Buyers and 0-Level Sellers Work

Let us have a brief look at how 1-level buyers and 0-level sellers work, as described in Vidal and Durfee (1996) and Vidal (1998).

- 1-level buyers model sellers in the marketplace by keeping a history of the qualities of the goods they purchased from each seller. In particular, a 1-level buyer $b$ remembers the last $N$ qualities offered by a seller $s$ for the good $g$ that it purchases from $s$. It then defines a probability density function $q_s^g(x)$ over the quality $x$ offered by seller $s$ for good $g$. Function $q_s^g(x)$ returns the probability that seller $s$ will offer an instance of good $g$ that has quality $x$. Buyer $b$ then uses the expected value of this probability density function to calculate which seller will offer good $g$ with highest expected product value:

$$s^* = \arg\max_{s\in S} E\left(V_b^g\left(p_s^g, q_s^g(x)\right)\right) \tag{19}$$

$$= \arg\max_{s\in S} \frac{1}{|Q|}\sum_{x\in Q} q_s^g(x) V_b^g\left(p_s^g, x\right), \tag{20}$$

  where $Q$ is a finite set of values representing product qualities.
- A 0-level seller $s$, when requested by some buyer $b$ for the price of some good $g$, will choose a price $p_s^*$ greater than or equal to its cost $c_s^g$ to produce $g$ such that its expected profit is maximized:

$$p_s^* = \arg\max_{p\in P} h_s^g(p), \tag{21}$$

  where $h_s^g(p)$ returns the profit seller $s$ expects to get if it offers good $g$ at price $p$. Depending on the success of the transaction, $h_s^g(p)$ is learned as follows:

$$h_s^g(p) \leftarrow (1-\alpha)h_s^g(p) + \alpha Profit_s^g(p), \tag{22}$$

  where $\alpha$ is the learning rate ($0 \le \alpha \le 1$) and $Profit_s^g(p)$ is the actual profit:

$$Profit_s^g(p) = \begin{cases} p - c_s^g & \text{if } s \text{ is able to sell } g, \\ 0 & \text{otherwise.} \end{cases} \tag{23}$$

### 4.3. Buyers' Comparison

We experimentally compare the performance of our proposed buyers with 1-level buyers, in terms of satisfaction and computational time. Toward this goal, we simulate a marketplace populated with 32 sellers and 40 buyers, using Java 2. The seller population is equally divided into four groups (each having eight sellers): Group A offers goods with quality chosen randomly from interval [32, 42]. Group B consists of dishonest sellers who attract buyers with high-quality goods ($q = 45$) and then cheat them with really low-quality ones ($q = 1$). Sellers in group C offer goods with fixed quality $q = 40$. These sellers do not consider adjusting the quality of their goods. Sellers in group D offer goods with relatively lower starting quality $q = 38$, compared to sellers in group C. However, these sellers will consider improving product quality up to value 45 to meet the buyers' needs, according to our proposed selling algorithm.

The buyer population is equally divided into two groups: Group I consists of the 1-level buyers. Group II consists of our proposed buyers. Other parameters are set as follows:

- The number of qualities $N$ offered by a seller $s$ for some good $g$ that a 1-level buyer remembers is 50.
- The quality $q$ of a good is chosen to be equal to the cost for producing that good. This supports the common assumption that it costs more to produce high-quality goods.
- The true product value function $v^b(g, p, q) = 3q - p$, where $p$ and $q$ represent the price and quality of the good $g$ purchased, respectively.
- The reputation threshold $\Theta = 0.5$ and the disreputation threshold $\theta = -0.9$.
- The demanded product value $\vartheta^b(g) = 80$. Thus, even when a seller sells at cost, it must offer goods with quality of at least 40 to meet the buyers' requirement.[7]
- The cooperation factor $\mu$ is defined as in equation (12), where $\mu_{\min} = 0.005$, $v^b_{\max} = 3q_{\max} - p_{\min}$, $v^b_{\min} = 3q_{\min} - p_{\max}$, $q_{\max} = p_{\max} = 49.0$, and $q_{\min} = p_{\min} = 1.0$. We prevent $\mu$ from becoming zero when $v^b = \vartheta^b$ by using value $\mu_{\min}$.
- The noncooperation $\nu$ is defined as in equation (13), where we choose $\lambda = 3$. The use of factor $\lambda > 1$ indicates that a buyer will penalize a noncooperative seller $\lambda$ times greater than it will award a cooperative seller. This implements the traditional assumption that reputation should be difficult to build up, but easy to tear down.
- The exploration probability $\rho$ and the learning rate $\alpha$ are both set to 1 initially, and decreased over time (by factor 0.998) down to $\rho_{\min} = 0.1$ and $\alpha_{\min} = 0.1$.
- The number of consecutive unsuccessful auctions (after which a seller following our proposed algorithm may consider improving the quality of its goods) $m = 10$, and the number of consecutive successful auctions (after which a seller following our proposed algorithm may consider reducing the quality of its goods) $n = 10$.
- The quality increasing factor $Inc = 0.05$, and the quality decreasing factor $Dec = 0.05$.

The results we report here are based on the average of 100 runs each of which has 5000 auctions.

Because the higher product value a buyer receives, the better satisfied it is, we record and present in Figure 2 the true product values obtained by a 1-level buyer (graph (i)) and by a buyer following our proposed algorithm (graph (ii)).

As shown in the figure, the buyer following the proposed algorithm receives goods with higher true product values and is therefore more greatly satisfied. In fact, the product values this buyer obtains are reaching the highest possible value (90) that could be offered in the

[7] Because $3(40) - 40 = 80$.

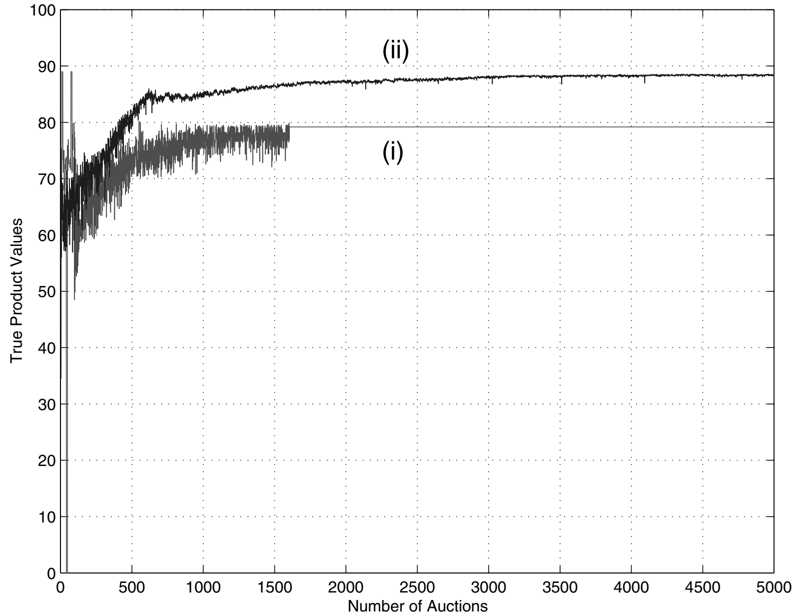


FIGURE 2. Comparison of true product values obtained by a 1-level buyer (graph (i)) and by a buyer following our proposed algorithm (graph (ii)).

marketplace.[8] The highest product value obtained by the 1-level buyer is about 80 only, indicating that it selects sellers in group C as its favorite sellers.[9] Clearly, the 1-level buyer is not able to discover that sellers in group D are actually the best sellers to purchase from. The reason is that although the 1-level buyer may try these sellers with their improved quality products, the history of low initial quality products offered by these sellers earlier keeps the buyer from selecting them as sellers with maximum expected value, according to the buyer's probability density function model shown in equations (19) and (20).

We are also interested in investigating the performance of the two buyers in terms of computational time. The run time needed for a buyer to complete an auction is composed of communication and computational time. The communication time accounts for the time needed for communication between the buyer and the sellers (e.g., the buyer broadcasting its request to sellers, the sellers responding with their bids, etc.). The computational time accounts for the time needed by the buyer to compute the seller that it will purchase the good from, according to its buying algorithm. Clearly, the communication time depends on the specific network underlying the marketplace and is therefore not relevant for comparison. The computational time, however, depends on the complexity of the buying algorithm and can be compared between agents using different algorithms. Obviously, the shorter the computational time the better the algorithm. This is especially important in application domains where the buyer is required to calculate a suitable seller within a constrained time frame.

[8] Since the highest quality offered (by sellers in group D) in our marketplace is 45 and since we assume cost equals quality, the highest possible product value offered in our market (by sellers in group D if they sell at cost) would be $3(45) - 45 = 90$.

[9] These sellers offer goods with fixed quality 40. So, the highest product value they could offer if they sell at cost is $3(40) - 40 = 80$.

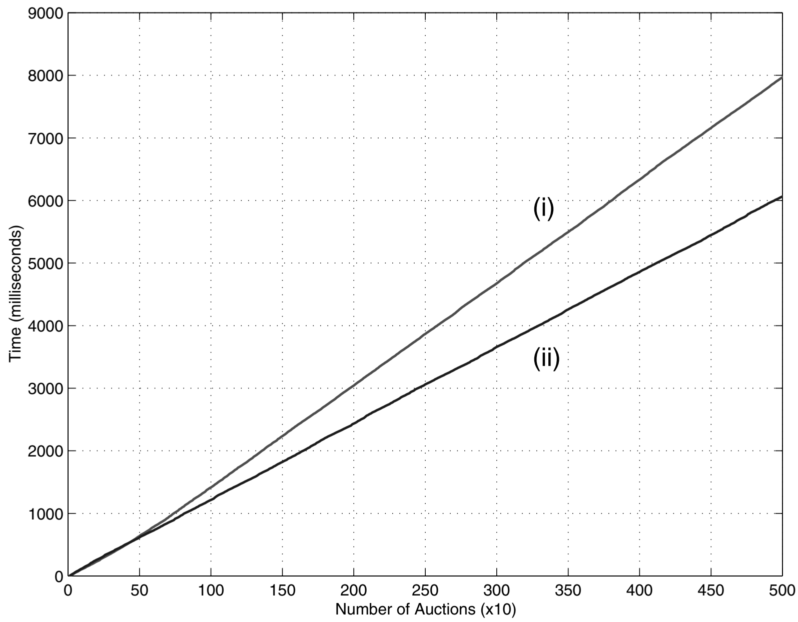


FIGURE 3. Comparison of computational time over the number of auctions taken by a 1-level buyer (graph (i)), and by a buyer using our proposed algorithm (graph (ii)).

Figure 3 shows the computational time over the number of auctions taken by a 1-level buyer (graph (i)) and by a buyer using our proposed algorithm (graph (ii)). The figure indicates that the buyer following our proposed algorithm outperforms the 1-level buyer. This is argued even more convincingly by looking at the respective algorithms governing the behaviors of these two buyers. To calculate the seller with highest expected value, the 1-level buyer has to examine every seller in the market (equation (19)). Moreover, for each seller $s$, the 1-level buyer also needs to calculate the product of the expected value and the probability of having that value at each quality $q$ (equation (20)). Thus, the order of growth of the algorithm underlying the 1-level buyer is $O(|S||Q|)$, where $|S|$ and $|Q|$ are the cardinalities (sizes) of the set of sellers and the set of quality values, respectively. The order of growth of our proposed algorithm is $O(|S|)$, because a buyer $b$ only needs to examine the set of sellers to compute a suitable seller. In the long run when every seller may be placed into either the set of reputable sellers or the set of disreputable sellers, this order will be reduced to $O(|S_r^b|)$, where $S_r^b$ is the set of reputable sellers to buyer $b$ and $|S_r^b|$ should be a lot smaller than $|S|$. Obviously, $O(|S_r^b|)$ beats $O(|S||Q|)$, especially when $|S|$ and $|Q|$ are sufficiently large.

## 4.4. Sellers' Comparison

We also experimentally compare the performance of our proposed sellers with 0-level sellers, in terms of satisfaction level and computational costs. Toward this objective, we simulate a marketplace populated with 20 sellers and 40 buyers. We let half of the sellers be the 0-level sellers, who offer goods with fixed quality of 40. The other half are our proposed sellers, who provide goods with lower initial quality of 38 but consider adjusting product quality to meet the buyers' needs, according to the proposed algorithm. All buyers follow a simplified learning version of our proposed buying algorithm, that is, they only

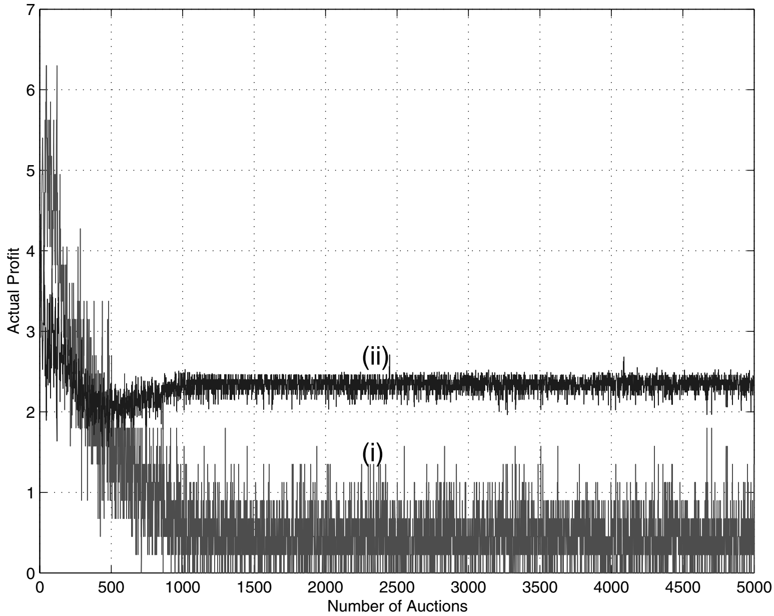


FIGURE 4. Actual profits made from a buyer by the group of 0-level sellers (graph (i)), and by the group of sellers following our proposed algorithm (graph (ii)).

use reinforcement learning and do not model sellers' reputation.[10] Other parameters such as $v^b(g, p, q)$, $\rho, \alpha, m, n, Inc$, and $Dec$ are chosen as in the previous experiment. The following reported results are based on the average taken over the buyer population.

However, the higher profit a seller makes the more greatly satisfied it is, we show in Figure 4 the profits made over the number of auctions from a buyer by the 0-level sellers (graph (i)), and by the sellers following our proposed algorithm (graph (ii)). We notice from the figure that, at the beginning the 0-level sellers are often chosen by the buyer because they offer goods with higher quality. However, as sellers of the other group improve the quality of their goods, the 0-level sellers lose more and more sales in the long run. This is indicated by a sharp decline in the profit graph, reaching the mean of approximately 0.5 after about 1000 auctions. In contrast, as the sellers following our proposed algorithm improve their product quality, they are selected more and more often by the buyer, resulting in their improved profit. In fact, they outperform the 0-level sellers after 1000 auctions with their profit reaching the mean of about 2.25, which is 4.5 times greater than that of the 0-level sellers.

Although the sellers following our proposed algorithm achieve better satisfaction than the 0-level sellers, they do not incur more computational time. This is because both algorithms underlying these two seller types have the same order of growth $O(|P|)$ where $|P|$ is the cardinality of the set of prices, as the sellers of both types search this set for the price that maximizes their expected profits. Indeed, Figure 5 shows that the difference in computational time spent by the two types of sellers is negligible.

[10] This is because it would be even more advantageous for our proposed sellers if buyers were to also model reputation of sellers.

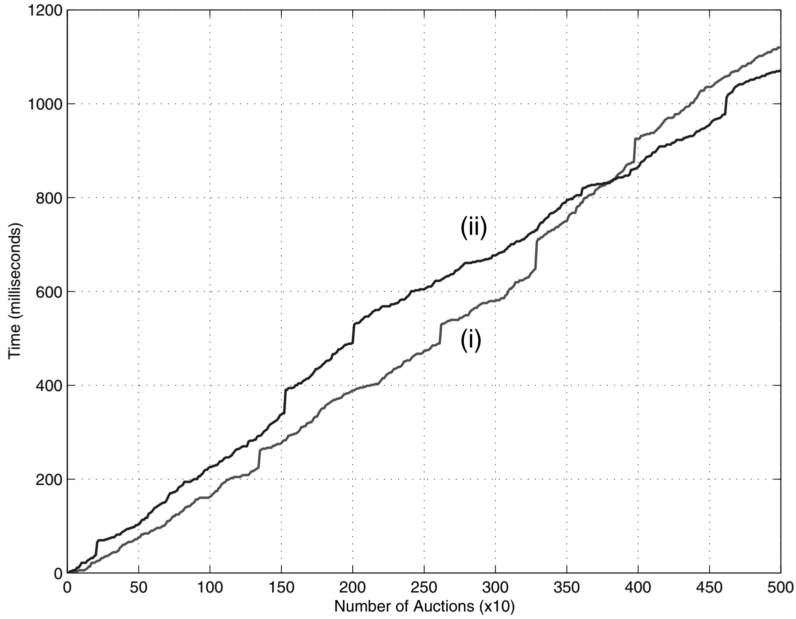


FIGURE 5. Computational times spent by a 0-level seller (graph (i)) and by a seller following the proposed algorithm (graph (ii)).

## 4.5. Discussion of Parameters

This section provides some justification for our choices of such parameters as the reputation threshold $\Theta$, the disreputation threshold $\theta$, and the true product value function $v^b$.[11]

The reputation threshold $\Theta$ ($0 < \Theta < 1$) is a buyer $b$'s specific constant, which buyer $b$ uses to determine whether it should consider a seller $s$ as a reputable seller. Consequently, the stricter (or more conservative) $b$ is, the higher value it would choose for $\Theta$. In addition, the more untrustful the market environment is, the higher the value $b$ should set $\Theta$ to. As the range of $\Theta$ is (0, 1), a buyer $b$ of medium strictness acting in a market of medium trust probably chooses $\Theta$ to be 0.50. This explains why we used this value for $\Theta$ in our experiments.

The disreputation threshold $\theta$ ($-1 < \theta < 0$) is also buyer $b$'s specific constant. Buyer $b$ uses this constant to decide whether a seller $s$ should be rated as a disreputable seller. Obviously, if $b$ chooses $\theta$ to be too low, dishonest sellers will not be placed in the disreputable set as they should be, resulting in $b$'s frequently purchasing unsatisfactory value goods. In contrast, if buyer $b$ sets $\theta$ to be too high, more sellers will be placed in the set of disreputable sellers, with the extreme case where all sellers in the market are rated as disreputable sellers. Moreover, due to the fact that $b$ will not re-select disreputable sellers to do business with according to the proposed algorithm, $\theta$ should be set low enough in order for $b$ to avoid situations where it may carelessly place a seller $s$ in the disreputable set without having enough evidence of $s$'s being noncooperative. This also gives those sellers, who are willing to improve their products, opportunities to make good offers to $b$. Considering these reasons, we suggest that $\theta$ should take values in the range $[-0.9, -0.7]$. In fact, in our experiments

[11] A more complete discussion of parameters can be found in Thomas (2003).

we set $\theta = -0.9$ to make sure that a seller is placed in the disreputable set only when it is a really noncooperative or dishonest seller and therefore deserves that treatment.

Each buyer $b$ has its own way to evaluate the good it purchases using the true product value function $v^b$. Basically, $v^b$ is a function of the price $p$ that buyer $b$ pays for the good, and also of the quality $q$ that $b$ examines the good after receiving it from the seller. Buyer $b$ formulates $v^b$ based on its idea of the relative importance of these two factors. For example, if $b$ considers quality to be more important than price, it may set $v^b = kq - p$ with $k > 1$. Since $p$ and $q$ are elements in the finite sets of prices and quality values, respectively, there exist the maximum and minimum values ($v^b_{\max}$ and $v^b_{\min}$) of the true product value function $v^b$. If we continue with the above-mentioned example then $v^b_{\max} = kq_{\max} - p_{\min}$ and $v^b_{\min} = kq_{\min} - p_{\max}$. The existence of $v^b_{\max}$ and $v^b_{\min}$ justifies their use in equations (12) and (13).

## 5. DISCUSSION AND CONCLUSION

To promote the use of adaptive business agents for electronic commerce, it is important to design a market model and processing algorithms that serve to engender trust in their human users.

In our model, sellers learn to maximize their expected profits by using reinforcement learning to adjust product prices and also by altering product quality to provide more customized value for their goods.

Buying agents learn to maximize their expected value of goods by using reinforcement learning to make effective future purchases. In particular, our buying agents keep track of which seller has offered the good it has purchased, to appropriately learn the best seller from which to purchase that good, in the future.

A critical component of the buying agents' algorithm is the modeling of the reputation of the selling agents. Buyers learn to maximize their expected value of goods by selecting appropriate sellers from among the reputable ones. Moreover, the use of a disreputation threshold and its corresponding disreputable set provides a mechanism for buying agents to recognize and eventually ignore any dishonest sellers in the marketplace.

The algorithms proposed here allow adaptive business agents to operate in marketplaces with all of the following characteristics: open (new agents are allowed to enter the market), dynamic (selling agents may alter product quality and buying agents may alter their needs), uncertain (buying agents evaluate the quality of goods only after purchase), and untrusted (dishonest agents may exist).

In this article, we have demonstrated how to effectively model reputation as part of the design of adaptive business agents on the Internet. In addition, we have demonstrated the value of our proposed market model and the buying and selling algorithms for its agents, in comparison with a competing popular design: that of recursive modeling. We have discussed both the importance of user satisfaction and the computational time, in our approach. We are therefore providing an important new approach for the design of adaptive business agents that engender trust in their users.

Two possible avenues for future research are worth noting. First, it is possible to explore an additional version of the proposed selling algorithm in which a seller divides buyers into groups that use similar true product value functions and keeps track of groups of buyers' behaviors, instead of individual buyers' behaviors. The main advantage of this approach is that it allows a seller to significantly reduce the number of customized versions of a good to be maintained. However, it also presents several issues to be addressed: ($i$) How to measure the similarity between buyers' true product value functions. ($ii$) How to update the models of groups of buyers appropriately over time. (*iii*) How to detect if any buyer has changed the way it evaluates the goods it purchases and therefore should be removed from its current group and placed in another group, resulting in both groups to be updated. Obviously, a formal analysis is needed to justify the approach, considering its advantages and the complexity in addressing the above issues.

Second, in our proposed buying algorithm, buyers select sellers based on their own experience without communicating with other buyers. The advantage of this approach is that buyers can act independently without being affected by communication delays, the failure of some key buyers, and the reliability of the information exchanged. An alternative approach is to consider the case where buyers in the market form neighborhoods such that within a neighborhood they exchange knowledge about sellers. The buyers then use their own knowledge combined with the exchanged knowledge to make purchase decisions. Several issues need to be addressed to realize this approach: (*i*) How neighborhoods should be formed. (*ii*) What knowledge to exchange and how to make use of it. (*iii*) How often a buyer should communicate and how much a communicating buyer gains (compared to the case where the buyer does not communicate and considering the cost of communication).

## REFERENCES


DAVIS, R., and R. G. SMITH. 1983. Negotiation as a metaphor for distributed problem solving. Artificial Intelligence, **20**(1):63–109.

LITTMAN, M. L. 1994. Markov games as framework for multi-agent reinforcement learning. *In* Proceedings of the Eleventh International Conference on Machine Learning, pp. 157–163.

ONO, N., and K. FUKUMOTO. 1996. Multi-agent reinforcement learning: A modular approach. *In* Proceedings of the Second International Conference on Multi-Agent Systems, pp. 252–258.

SABATER, J., and C. SIERRA. 2001. REGRET: A reputation model for gregarious societies. *In* Papers from the Fifth International Conference on Autonomous Agents Workshop on Deception, Fraud and Trust in Agent Societies, pp. 61–69.

SANDHOLM, T. W., and R. H. CRITES. 1995. Multi-agent reinforcement in the iterated Prisoner's dilemma. Biosystems, **37:**147–166.

SEN, S., M. SEKARAN, and J. HALE. 1994. Learning to coordinate without Sharing Information. *In* Proceedings of the Twelfth National Conference on Artificial Intelligence, pp. 426–431.

SMITH, R. G. 1980. The contract net protocol: High level communication and control in a distributed problem solver. IEEE Transactions on Computers, **C-29**(12):1104–1113.

TRAN, T. 2003. Reputation-Oriented Reinforcement Learning Strategies for Economically-Motivated Agents in Electronic Market Environments. Ph.D. Thesis, School of Computer Science, University of Waterloo.

TRAN, T. T., and R. COHEN. 2002. A reputation-oriented reinforcement learning strategy for agents in electronic marketplaces. Computational Intelligence (CI) Journal Special Issue on Agent Technology for Electronic Commerce, **18**(4):550–565.

VIDAL, J. M. 1998. Computational Agents That Learn About Agents: Algorithms for Their Design and A Predictive Theory of Their Behavior. Ph.D. Thesis, Department of Computer Science & Engineering, University of Michigan.

VIDAL, J. M., and E. H. DURFEE. 1996. The impact of nested agent models in an information economy. *In* Proceedings of the Second International Conference on Multi-Agent Systems, pp. 377–384.

YU, B., and M. P. SINGH. 2000. A social mechanism of reputation management in electronic communities. *In* Cooperative Information Agents IV; Lecture Notes in Artificial Intelligence, Vol. 1860. *Edited by* M. Klusch and L. Kerschberg. Springer-Verlag, Berlin, pp. 154–165.