Learning Algorithms for Software Agents in Uncertain and Untrusted Market Environments

Thomas Tran and Robin Cohen
School of Computer Science, University of Waterloo
Waterloo, ON, Canada N2L 3G1
{tt5tran, rcohen}@math.uwaterloo.ca

The problem of how to develop algorithms that guide the behaviour of personal, intelligent software agents participating in electronic marketplaces is a subject of increasing interest from both the academic and industrial research communities. Since a multi-agent electronic market environment is, by its very nature, open, dynamic, uncertain, and untrusted, it is very important that participant agents are equipped with effective and feasible learning algorithms in order to achieve their goals. In this paper, we propose algorithms for buying and selling agents in electronic marketplaces, based on reputation modelling and reinforcement learning (RL).

We model the agent environment as an open marketplace which is populated with economic agents *(buyers* and *sellers),* freely entering or leaving the market. The process of buying and selling goods is realized via a 3-phase mechanism: *(i)* A buyer announces its request for a good, *(ii)* Sellers submit bids for delivering such goods. *(iii)* The buyer evaluates the submitted bids and selects a suitable seller. Thus, the buying and selling process can be viewed as an *auction* where a seller is said to be *winning the auction* if it is able to sell its good to the buyer. We assume that the quality of a good offered by different sellers may not be the same, a seller may alter the quality of its goods, and there may be dishonest sellers in the market. We also assume that a buyer can examine the quality of the good it purchases only after it receives that good from the selected seller.

In our approach, buying agents learn to avoid the risk of purchasing low quality goods and to maximize their expected value of goods by dynamically maintaining sets of reputable and disreputable sellers. Selling agents learn to maximize their expected profits by adjusting product prices and optionally altering the quality of their goods.

In our buying algorithm, a buyer *b* uses an expected value function /[6], where $f^b(g,p,s)$ represents buyer b's expected value of buying good *g* at price *p* from seller *s.* Buyer *b* maintains reputation ratings for all sellers, and chooses among its set of reputable sellers S^b a seller *s* that offers good *g* at price *p* with maximum expected value. In addition, with a small probability *p,* buyer *b* chooses to explore (rather than exploit) the marketplace to discover new reputable sellers by randomly selecting a seller 5, provided that *s* is not from the set of disreputable sellers S^b_{dr}. After paying seller *s* and receiving good *g,* buyer *b* can examine the quality *q* of *g* and calculate the true value $v^b(g,p,q)$ of *g.* The expected value function f^b is incrementally learned in an RL framework:

$$f^b(g,p,\hat{s}) \leftarrow f^b(g,p,\hat{s}) + \alpha(v^b(g,p,q) - f^b(g,p,\hat{s}))$$

where *a* is called the *learning rate* $(0 \leq \alpha \leq 1)$. The reputation rating of *s* is then updated based on whether or not the true value of good *g* is greater than or equal to the value demanded by *b.* Our reputation updating scheme implements the traditional ideas that reputation should be difficult to build up but easy to tear down, and that a transaction with higher value should be more appreciated than a lower one. The set of reputable and disreputable sellers (S^b_r and S^b_{dr}) are accordingly re-calculated based on the updated reputation rating of *s.*

In our selling algorithm, a seller *s* makes use of an expected profit function h^s, where $h^s(g,p,b)$ represents seller *s* expected profit if it sells good *g* at price *p* to buyer *b.* Seller *s* chooses a price $\hat{p}$ to sell good *g* to buyer *b* such that its expected profit is maximized. After the transaction, the expected profit function h^s is learned incrementally using RL:

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

where $\phi^s(g,\hat{p},b)$ is the actual profit of seller *s* when it sells good *g* at price *p* to buyer 6, and is defined as follows:

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

where $c^s(g,b)$ is the cost of seller s to produce good *g* for buyer *b.* Our selling algorithm also allows sellers to alter the quality of their goods in order to meet the buyers' needs and to further increase their future profit, depending on the success of their previous sales with the buyers.

We believe that our approach should lead to improved satisfaction for buyers and sellers, since buyers should be less at risk of purchasing low quality goods when maintaining sets of reputable and disreputable sellers, and since sellers are allowed to adjust both price and quality to meet the buyers' demands.

We have performed experimentation to measure the value of our model on both microscopic and macroscopic levels. On the micro level, we were interested in examining the individual benefit of agents, particularly their level of satisfaction. Our experimental results show that in both modest and large sized marketplaces, buyers following the proposed buying algorithm (i.e., using the combination of RL and reputation modelling) will achieve better satisfaction than buyers using

RL alone, and that sellers following the proposed selling algorithm (i.e., using RL and adjusting product quality) will make better profit than sellers using only RL. On the macro level, we studied how a market populated with our buyers and sellers would behave as a whole. Our results demonstrate that such a market can reach an equilibrium state where the agent population remains stable (as some sellers who repeatedly fail to sell their goods will decide to leave the market), and this equilibrium is optimal for the participant agents.

We report some experimental results confirming the satisfaction of buyers following the proposed algorithm in a large sized marketplace. The simulated market is populated with 160 sellers and 120 buyers where each buyer participates in 5000 auctions. The seller population is equally divided into four groups: Group A offers goods with quality chosen randomly from the quality interval [32, 42]. Group B consists of dishonest sellers, who try to attract buyers with high quality goods (q = 45) and then cheat them with really low quality ones (q = 1). Sellers in group C use RL but do not consider adjusting product quality; they offer goods with fixed quality 39. Sellers in group D follow the proposed selling algorithm and offer goods with initial quality 39. The buyer population is divided equally into two groups: Group I uses *RL* alone and group II follows the proposed buying algorithm. We set quality equal to cost to support the common idea that it costs more to produce high quality goods. Also, we set the true product value $v^b = 3.bq$ - p, the demanded product value tf^6 = 100, the learning rate a = 1 and exploration rate p = 1 initially with both decreased over time by factor 0.9997 down to a_{min}— p_{min} = 0.1. The results reported here are based on the average taken over the respective populations of the two groups of buyers.

Figure 1(a) and (b) present the histograms of true product values obtained by a buyer using RL alone, and by a buyer following the proposed buying algorithm, respectively. Clearly, the buyer following the proposed algorithm receives more goods with higher value (v^b = 110) and fewer goods with lower values (v^b = 65... 105), and is therefore better satisfied. In particular, the buyer following the proposed algorithm makes about 2350 more purchases with high product value of 110 (or 15.67 times greater) than those purchases made by the buyer using RL alone.

We are also interested in seeing how much better the buyer following the proposed algorithm is able to avoid interaction with the dishonest sellers. Figure 2(a) and (b) show the profits made by the dishonest sellers from the buyer using RL alone, and from the buyer following the proposed algorithm, respectively. We notice that graph (a) is higher than graph (b), indicating that the dishonest sellers are able to make more profit from those buyers that only use RL but do not model sellers' reputation. Moreover, the profit in graph (b) is reduced to zero after about 2700 auctions, implying that in the long run the dishonest sellers are not able to make any profit from the buyer following the proposed algorithm, because they are rated as disreputable sellers and therefore no longer chosen by the buyer.

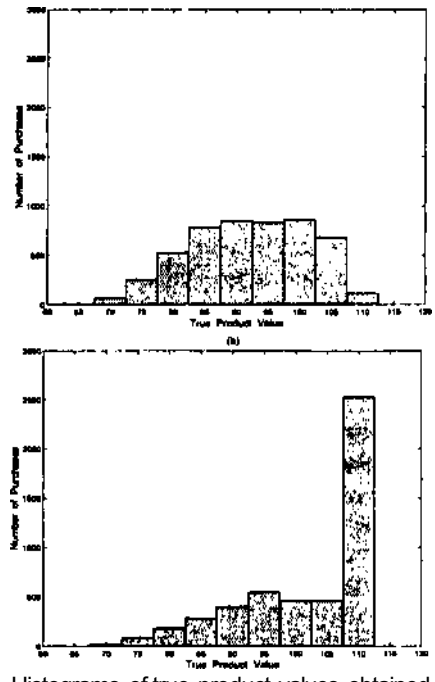

Figure 1: Histograms of true product values obtained by a buyer using RL alone (a), and a buyer using the proposed algorithm (b).

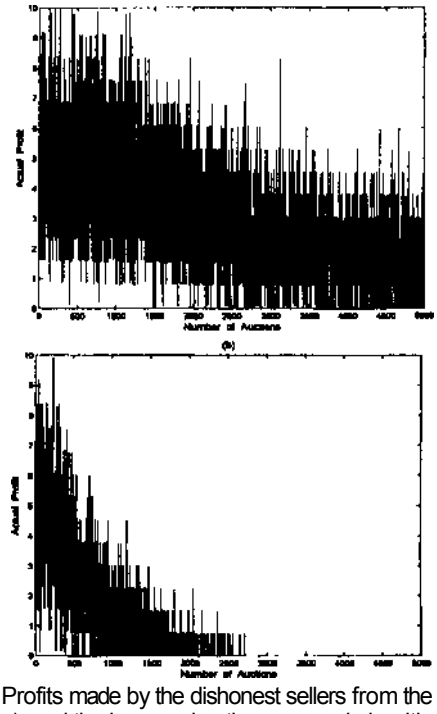

Figure 2: Profits made by the dishonest sellers from the buyer using RL alone (a), and the buyer using the proposed algorithm (b).

In general, our work demonstrates that reputation modelling can be used in combination with reinforcement learning to design intelligent learning agents that participate in electronic marketplaces. This research also aims to provide a principled framework for building effective economic agents and desirable electronic market environments.