A Reputation-Oriented Reinforcement Learning Approach for Agents in Electronic Marketplaces

Thomas Tran
Computer Science Department, University of Waterloo
Waterloo, ON, Canada N2L 3G1
tt5tran@math.uwaterloo.ca

The problem of how to design personal, intelligent agents for e-commerce applications is a subject of increasing interest from both the academic and industrial research communities. In our research, we consider the agent environment as an open marketplace which is populated with economic agents (buyers and sellers), freely entering or leaving the market. The problem we are addressing is how best to model the electronic marketplace, and what kinds of learning strategies should be provided, in order to improve the performance of buyers and sellers in electronic exchanges.

Our strategy is to introduce a reputation-oriented reinforcement learning algorithm for buyers and sellers. We take into account the fact that multiple sellers may offer the same good with different qualities. In our approach, buyers learn to maximize their expected value of goods and to avoid the risk of purchasing low quality goods by dynamically maintaining sets of reputable sellers. Sellers learn to maximize their expected profits by adjusting product prices and by optionally altering the quality of their goods.

In our buying algorithm, a buyer b uses an expected value function f^b, 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 sellers, and chooses among its set of reputable sellers S_r^b a seller $\hat{s}$ that offers good g at price p with maximum expected value. After paying seller $\hat{s}$ and receiving good g, buyer b can examine the quality q of g. It then calculates the true value $v^b(p, q)$ of good g. The expected value function f^b is incrementally learned in a reinforcement learning framework:

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

where α is called the *learning rate* ($0 \leq \alpha \leq 1$). The reputation rating of $\hat{s}$ is then updated based on whether or not the true value of good g is greater than or equal to the desired value. The set of reputable sellers S_r^b is also re-calculated based on the updated reputation rating of $\hat{s}$.

In our selling algorithm, seller s tries to sell good g to buyer b to maximize its expected profit h^s, where $h^s(g, p, b)$ represents the expected profit for seller s if it sells good g at price p to buyer b. The expected profit function h^s is learned incrementally using reinforcement learning:

$$h^s(g, p, b) \leftarrow h^s(g, p, b) + \alpha(Profit^s(g, p, b) - h^s(g, p, b))$$

where $Profit^s(g, p, b)$ is the actual profit of seller s when it sells good g at price p to buyer b. $Profit^s(g, p, b)$ is defined as follows:

$$Profit^s(g, p, b) = \begin{cases} p - c^s(g, b) & \text{if } s \text{ is able to sell } g \text{ to } b, \\ 0 & \text{otherwise.} \end{cases}$$

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, depending on the success of previous sales with buyers.

The work of (Vidal & Durfee 1996) on modeling buying and selling agents in an information economy motivates our work. Instead of focusing on having agents maintain recursive models of other agents, we believe that reputation is an important factor for buyers to exploit, and that it is important to allow for sellers to alter the quality of their goods to satisfy buyers' needs.

We feel that our approach should lead to improved satisfaction for buyers and sellers, since buyers should be less at risk of receiving low quality goods when maintaining sets of reputable sellers, and sellers are allowed to adjust both price and quality to meet buyers' needs. In addition, it should lead to improved performance for buyers (in terms of computational cost), since buyers are focusing on the subset of reputable sellers.

For future work, we plan to conduct some experimentation to measure the value of our model. Our plan is to compare the proposed algorithm with a simplified version where buyers do not use a reputation mechanism and sellers do not consider altering the quality of their products.

Other extensions of the model that we are considering exploring are: *(i)* sellers not tracking individual buyers' behaviour; *(ii)* sellers dividing buyers into groups and tracking groups of buyers' behaviour; and *(iii)* allowing buyers to receive advice from other buyers in their neighbourhoods.

Our work aims to demonstrate that reputation mechanisms can be used in combination with reinforcement learning to design intelligent learning agents that participate in market environments. We also hope to provide some general guidelines for AI-systems designers in building effective economic agents.

References

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