Reciprocating with learned models

Anish Biswas & Sandip Sen
Department of Mathematical & Computer Sciences,
The University of Tulsa

e-mail: abiswas@euler.mcs.utulsa.edu,sandip@kolkata.mcs.utulsa.edu

Abstract

In our prior work, we have demonstrated the effectiveness of a probabilistic reciprocity mechanism by which self-interested agents may learn to adopt a cooperative relationship with other similar agents. Reciprocative decisions were made based on the balance of help with another agent. In this paper, we expand that framework by making an agent explicitly model the help-giving procedure of the other agent. This learned model is then used to make decisions on requests for help received from that agent. We show that when asking for help consumes non-negligible time, the model based reciprocative agents can outperform the reciprocative agents who uses only balance of past transactions.

Introduction

If participating agents in a multiagent system can be assumed to be cooperative in nature, coordination mechanisms can be used that will realize desirable system performance. Such assumptions, however, are untenable in open systems. Agent designers have to design agents and agent environments with the understanding that participating agents will act to serve their self-interests instead of working towards group goals. We investigate the choice of interaction strategies and environmental characteristics that will make the best self-interested actions to be cooperative in nature. In our previous work (Sen 1996), we have analyzed the inadequacy of traditional deterministic reciprocity mechanisms (Axelrod 1984) to promote cooperative behavior with a fair distribution of the workload. We proposed a probabilistic reciprocity mechanism and showed that it can generate stable and cooperative behavior among a group of self-interested agents. The resultant system was found to exhibit close to optimal throughput with a fair distribution of the workload among the participating agents. This mechanism was also found to be robust in the presence of misconception about the amount of help received (Sen & Biswas 1998). Others have evaluated systems of heterogeneous agents with different, but fixed, cooperative attitudes (Cesta & Miceli 1996). We are interested in developing agents that can adapt its behavior to other agents in the group.

We had proposed that an appropriate decision mechanism based on reciprocity should have the following characteristics:

- allow agents to initiate cooperative relationships (this implies that it should be able to handle asymmetrical interactions),
- use a mechanism to compare cooperation costs,
- allow agents to be inclined to help someone with whom it has a favorable balance of help (have received more help than have offered help),
- be able to flexibly adjust inclination to cooperate based on current work-load (e.g., more inclined to cooperate when less busy, etc.).

In our previous work, we proposed that a reciprocative agent should be more inclined to help those agents from whom they have received help in the past. While seeking help, though, our reciprocative agent did not have a very good idea of who to ask for help. As such, it had to sequentially ask for help from other agents in some order until someone agreed to provide help. Since we had assumed that asking for help and getting a response takes negligible time, system performance was unaffected by such an enumerative scheme for requesting help. More realistically, we have to account for a time lag between asking for help and receiving a response. Under these situations, agents who can model the help-giving inclinations of other agents can save considerable time and effort by asking for help from those agents first who are more likely to agree to help.

In this paper, we propose an approach to developing an approximate model of the help-giving policy of other agents. Such a model will probably be a coarse approximation to the actual mechanism used by other agents, but can serve to satisfy the need for identifying agents who are likely to heed a request for help.

In the following we first visit the probabilistic reciprocity framework developed in our previous work, and then illustrate our proposed modeling scheme which is based on this framework. We conclude with experimental results highlighting when such explicit modeling of other agents' behavior is of likely benefit.

Probabilistic reciprocity

We assume a multiagent system with N agents. Each agent is assigned to carry out T tasks. The jth task assigned to the ith agent is t_{ij}, and if agent k carried out this task independently of other tasks, the cost incurred is C^k_{ij}. However, if agent k carried out this task together with its own task t_{kl}, the cost incurred for task t_{ij} is C^{kl}_{ij}. Also, the cost incurred by agent k to carry out its own task t_{kl} while carrying out task t_{ij} for agent i is C^{kij}_{kl}. In this paper, we allow an agent to carry out a task for another agent only in conjunction with another of its own tasks.

We now identify the scopes for cooperation. If an agent, k, can carry out the task of another agent, i, with a lower cost than the cost incurred by the agent who has been assigned that task ($C^i_{ij} > C^{kl}_{ij}$), the first agent can cooperate with the second agent by carrying out this task. If agent k decides to help agent i, then it incurs an extra cost of C^{kl}_{ij} but agent i saves a cost of C^i_{ij}. The obvious question is why should one agent incur any extra cost for another agent. If we consider only one such decision, cooperation makes little sense. If, however, we look at a collection of such decisions, then reciprocal cooperation makes perfect sense.

We have used a probabilistic decision mechanism that satisfies the set of criteria for choosing when to honor a request for help that we described at the end of the previous section. We define S_{ik} and W_{ik} as respectively the savings obtained from and extra cost incurred by agent i from agent k over all of their previous exchanges. Also, let $B_{ik} = S_{ik} - W_{ik}$ be the balance of these exchanges ($B_{ik} = -B_{ki}$). We now present the probability that agent k will carry out task t_{ij} for agent i while it is carrying out its task t_{kl}. This probability is calculated as:

$$Pr(i,k,j,l) = \frac{1}{1 + \exp^{\frac{C^{kl}_{ij} - \beta * C^k_{avg} - B_{ki}}{\tau}}}, \qquad (1)$$

where C^k_{avg} is the average cost of tasks performed by agent k (this can be computed on-line or preset), and β and τ are constants. This gives a sigmoidal probability distribution in which the probability of helping increases as the balance increase and is more for less costly tasks. We include the C_{avg} term because while calculating the probability of helping, relative cost should be more important than absolute cost. The constants β and τ can be used to make agents more or less inclined to cooperate. The factor β can be used to move the probability curve right (more inclined to cooperate) or left (less inclined to cooperate). At the onset of the experiments B_{ki} is 0 for all i and k. At this point there is a 0.5 probability that an agent will help another agent by incurring an extra cost of $\beta * C^k_{avg}$ (we assume that the average cost incurred is known; an approximate measure is sufficient for our calculations). The factor τ can be used to control the steepness of the curve. For a very steep curve approximating a step function, an agent will almost always accept cooperation requests with extra cost less than $\beta * C^k_{avg}$, but will rarely accept cooperation requests with an extra cost greater than that value. The level of cooperation or the inclination to help another agent dynamically changes with problem solving experience.

Learning about other agents

Based on another agent's positive and negative responses to requests for help, a model of the help-giving behavior of that agent can be approximated. If approximately accurate models of most or all other agents are available, an agent can ask help from those who are likely to help. This will save our agent considerable time, if each help request consumes a noticeable time period.

An agent A uses complementary sigmoidal curves to approximate the help-giving behavior of another agent, B, as follows:

$$aprox(x) = \frac{c_1}{I} * f(x) + \frac{c_2}{I} * (1 - f(x)),$$

where $f(x) = \frac{1}{1+\exp\frac{x-\alpha}{\tau}}$ (we have used $\alpha = 2$, $\tau = 0.75$); c_1 and c_2 are the evidences for probability distributions $f(x)$ and $1 - f(x)$ respectively, and I is the expected number of mutual interactions between these two agents.

If agent B helps agent A with y when requested for that amount of help then the coefficients c_1 and c_2 in the model of B is increased at the proportion of $f(y)$:$(1 - f(y))$, as those are the proportional chances that the help was due to the two complementary models. And if B declined help, then c_1 and c_2 are decreased in the same way. The amount of increase and decrease of c_1 and c_2 is controlled by a learning rate.

We have adopted the above-mentioned learning approach with the following assumptions:

- We have an estimate of the number of interactions between two agents. It may also be the number of expected interactions within a learning period by which an agent has to come up with a model of another agent.
- The agent being modeled is considering the helping cost to decide whether or not to help.
- There is no communication other than the answers for the help seeking questions between the agents, so there is no other way to know the parameters they are considering while deciding to help.

Experiments

We ran a set of experiments where different agents in the population were given different help-giving functions. For example, a selfish and a philanthropic agent always returns a probability of 0 and 1, respectively, for any help requested. Other agents were given sigmoidal, and exponential functions for deciding on giving help.

We used a package delivery domain where agents were delivering packets from a central depot to some location

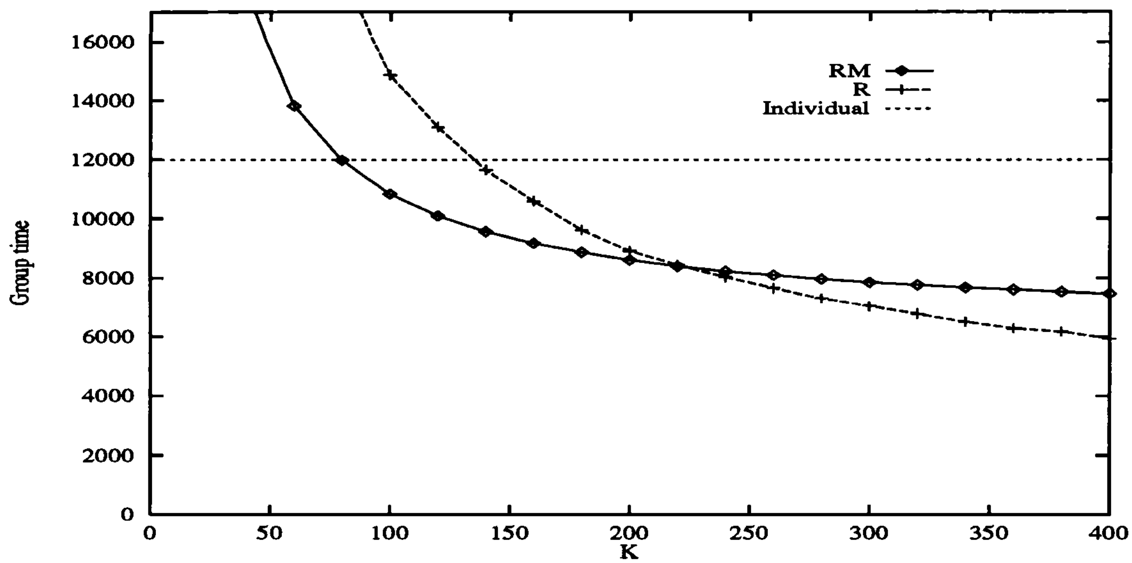

Figure 1: Total time taken by individual (I), reciprocative (R), reciprocative agents using models of others (RM) to deliver their packet as interaction time is varied (as $\frac{1}{K}$ is the unit of time, i.e. higher k implies lower interaction time).

along one of several radial roads. The goal of the agents was to complete the deliveries in minimal time. Agents who deliver their own packets without considering others in the world (individual, or I, agents) were used provide baseline performance. Agents developed in our previous work are called reciprocative, or R, agents. Agents proposed in this work are reciprocative agents who build explicit models of other agents' help-giving behavior and are called RM agents.

In our experiments, the RM agents use the first few interactions to learn the model of other agents, and thereafter use such models to select the order in which they will ask other agents for help. Also, it asks for help from only those agent who it thinks are highly likely to help (probability of help greater than 0.9). Thus, at times the RM agents will fail to secure help where an R agent would have secured it. This is because the R agents ask for help from all agents who can provide help irrespective of whether they are likely to help or not.

In a preliminary set of experiments, we observed that the RM strategy was able to approximate the help-giving function of other agents within a few interactions. Next, we ran experiments in the package delivery problem where each agent was assigned to deliver 100 packages. The number of agents, N, was varied from 1000 to 4000.

Our expectations from the experiments were as follows: if the time spent in requesting for help from another agent and getting a response (we will call this *interaction time*) was negligible, R agents would do the best; if interaction time is comparable to average time taken to deliver packets, the I agents would be most effective; in between, the RM agents would outperform others. We varied the interaction time by varying k, where $\frac{1}{k}$ is the interaction between two agents. Results from this set of experiments is presented in Figure 1. The performance of I, R, and RM agents confirm our expectations about their relative advantages as interaction time is varied. I, RM, and R agents perform the best in the ranges $k < 80$, $80 \leq k \leq 230$, and $k > 230$ respectively. This empirical observation also allows agents to adaptively choose the appropriate strategy to adopt as they find out about interaction time.

Future Work

In this work the modeling agents use separate learning phase and a phase where it uses the learned model. We plan to extend the learning mechanism to be a continuous, life-long scheme. This will be particularly useful when the strategy being modeled is not a fixed one, but is changing over time. To speed up the learning process we would like agents to use observation of interaction history between other agents. Agents can also share learned models of others. We are also interested in experimenting with domains where all the factors used in decision making by an agent are not accessible to the modeling agent. Resultant models developed, therefore, would necessarily be rough approximations. We plan to study the degradation in accuracy of such models as the influence of non-observable parameters is varied.

References

Axelrod, R. 1984. *The Evolution of Cooperation.* Basic Books.

Cesta, A., and Miceli, M. 1996. Help under risky conditions: Robustness of the social attitude and system performance. In *Proceedings of the Second International Conference on Multiagent Systems*, 18–25. Menlo Park, CA: AAAI Press.

Sen, S., and Biswas, A. 1998. Effects of misconception on reciprocative agents. to appear in the Proceedings of the Second International Conference on Autonomous Agents: AGENTS98.

Sen, S. 1996. Reciprocity: a foundational principle for promoting cooperative behavior among self-interested agents. In *Proceedings of the Second International Conference on Multiagent Systems*, 315–321. Menlo Park, CA: AAAI Press.