ADVICE-EXCHANGE IN HETEROGENEOUS GROUPS OF LEARNING AGENTS

Luis Nunes
ISCTE/FEUP/LIACC-NIAD&R
ISCTE, Av. Forças Armadas,
1649-026 Lisbon, Portugal
Luis.Nunes@iscte.pt

Eugénio Oliveira
FEUP/LIACC-NIAD&R
FEUP, Av. Dr. Roberto Frias,
4200-465, Porto, Portugal
eco@fe.up.pt

ABSTRACT
This research aims at studying the effects of exchanging information during the learning process in Multiagent Systems. The concept of advice-exchange, introduced in [3], consists in requesting extra feedback, in the form of episodic advice, from other agents that are solving similar problems. This work is concerned with the exchange of information in heterogeneous groups of learning-agents that either share the same environment or are solving problems with similar structure. Concepts, such as self-confidence, trust and advisor preference, were introduced in the experiments that led to the results discussed in this paper.

Categories and Subject Descriptors
I.2.6 [Learning]: Connectionism and neural nets, Parameter Learning; I.2.8 [Problem solving, Control Methods and Search]: Dynamic Programming; I.2.11 [Distributed Artificial Intelligence]: Intelligent agents, Multiagent systems

General Terms
Algorithms, Performance, Experimentation

Keywords
Machine Learning, Multiagent Systems, Advice-Exchange

1. INTRODUCTION
The question that drives this research is: “(How) can heterogeneous learning-agents cooperate during the learning process?” The approach taken in this work aims at improving the individual and global learning performance by providing learning-agents with more feedback concerning their current hypothesis than the environment can give them. The groups of agents referred in this paper are heterogeneous in the sense that they may use different learning techniques. The main difficulty in having heterogeneous groups is that internal parameters cannot be exchanged, because agents are not aware of which learning algorithm their peers are using. This fact restricts communication capabilities to knowledge that has meaning outside the agents' internal context, like states of environment, actions, or values of the quality function. The exchange of other types of data, such as “trust agent X”, was also found useful and it requires that all agents share a (small) common vocabulary. We will assume that the agents are working in domains in which the problem has the same structure and is presented in the same way to the agent, i.e. the meaning and order of the variables that compose both state and actions, are known to be the same for all environments used in an experiment. The key questions of this research are: how to select, incorporate, and benefit-from all the available information without re-processing all the data gathered by each of an agent’s peers.

2. ADVICE-EXCHANGE
The first version of this algorithm was introduced in [3] where it was evaluated in solving a simplified traffic-control problem. Descriptions of the most recent work on this subject with a detailed explanation of several variants of the concept as well as experimental results and conclusions can also be found in [2, 4].

When an agent senses the state of the environment and is prompted for an action it decides either to use its own knowledge to select the best action or to request advice to a peer concerning that state. The decision on whether or not to get advice is based on a comparison of an agent’s performance with that of its peers and in the degree of certainty an agent has concerning the next action to take. The self-confidence parameter is used in the decision of whether or not to get advice. The agent will compare its own performance statistics, weighted by the self-confidence parameter, with those of other peers. This parameter will make the agent overrate or underrate its performance, depending on its value. The self-confidence of an agent is increased at the end of each epoch if the agent’s performance is close enough to the best agents’ performance and decreased otherwise.

The trust parameters are used in the selection of the most suitable advisor for a given agent. The trust parameter measures “how good was the advice from a given agent in the past”. The trust that an advisee agent has on a given agent
(agentX) will be increased if the performance of the advisee increased during an epoch were it was advised by agentX. Trust in a certain advisor can also be influenced by a partner. This enables a group of agents to learn joint strategies used by other teams by influencing each of their partners to seek advice with a different member of a successful team.

When an agent decides to request advice it selects the peer with the highest performance-trust product and sends its current state to the selected agent, giving preference to the last advisor when this agent’s performance is similar to the best. When an advice request is received an agent will make a context switch, replacing its current parameters with those that were used in the epoch where it was most successful, and evaluate the state presented by the advisee according to those parameters. The process of incorporating the received knowledge in an agent’s current hypothesis depends on the algorithm each agent is using. In this work there are two types of agents, EA-agents and QL-agents, the first uses Evolutionary Algorithms (EAs) [1] to adjust the weights of an Artificial Neural Network (ANN), the second uses one-level Q-Learning (QL) [6]. When the advisee is an EA-agent it will backpropagate the advice as desired response using standard online Backpropagation [5] on the weights of the ANN. When the advisee is a QL-agent it will give a bonus to the advised action and decrease the expected quality of all other actions starting at that state. A similar technique, labelled Biasing-Binary was reported in [7].

The experiments this work refers to were conducted in the Pursuit Problem (a.k.a. Predator-Prey Problem). A detailed explanation of the variant used can be found in [2, 4]. The problem consists in two predator-agents learning to catch a prey in a square arena. In each experiment there were several scenarios, each with several arenas. Each arena contains two predators and one prey. The scenarios used in these experiments are the following:

1. Individual: Four arenas. In each arena all predators use the same learning algorithm and they do not exchange any information with their peers. Two arenas have EA-agents, the other two have QL-agents.
2. Social Heterogeneous: Equal to previous, except that agents are able to request advice to any of its seven peers in the same or other arenas.
3. Social Heuristic: Similar to the previous scenario but with an extra arena where two Heuristic agents are performing the same task and may also be chosen as advisors. Heuristic agents are preprogrammed to perform a fixed optimal individualistic policy.
4. Social Homogeneous: Similar to the Social Heterogeneous scenario, except that all agents use the same learning algorithm.

For each of the above scenarios 11 trials were made (x4, or x8, agents of each type) with different random seeds. Each trial ran for 9000 epochs and each epoch has 150 turns. For each trial there is a corresponding test which runs for 1000 epochs without learning or exchange of advice.

### 3. CONCLUSIONS AND FUTURE WORK

Table 1 shows, in the first column, the average performance, in all tests, for each algorithm-scenario pair. The second column of the same table shows the standard deviation of the results. We can see that in the scenarios where advice-exchange was used (Social scenarios) the learning agents achieve better performances, than the same algorithms in Individual scenarios. This is only a small example of the results achieved by this technique, further results can be found in [2, 4]. These results refer to a 20x20 toroidal arena with a reward function that is inversely proportional to the distance between predator and prey.

The results obtained so far in this and other experiments show that advice-exchange can be beneficial for the performance of learning agents even when the exchanged information is very simple.

The process of human learning has several stages of growing autonomy. Current work is focusing on the introduction of these stages of evolution in the process of advice-exchange to enable an agent to shift the way in which advice is being used. A development of the way trust is used, associating it with agent-situation pairs, is also under study. This may allow the advisee to differentiate who is the expert on the particular situation it is facing and improve the effectiveness of advice-exchange.

### 4. REFERENCES