The Origin of the Speeches: Language Evolution through Collaborative Reinforcement Learning

Ray Walshe

Artificial Intelligence Research Group
Dept. of Computer Applications, Dublin City University, Ireland
Ray.Walshe@CompApp.DCU.ie http://www.compapp.dcu.ie/~ray

Abstract. This research focuses on the artificial creation of the conditions that were necessary or facilitated language in its evolution. I propose a model for a system that has the capability to allow a complex communication protocol to evolve. In human language learning, the adult humans have already mastered the language and use their knowledge to teach infant humans. 50.000 B.C. (or whenever) when there were no adult masters of language, what conditions were necessary for the language-less homo sapiens to start developing the first language? Can machines evolve a language in a similar manner across generations? This research deals with facilitating the genesis of a communication system using evolutionary computation and reinforcement learning when initially none of the conspirators have mastered the system.

1 The Model for Language Evolution

Task Orientated Networked environment for the Acquisition of Language (TONAL) consists of a World where Agents exist and the Agents themselves. The World(s) and Agent(s) themselves are servers that can exist geographically anywhere on the Internet where there is WWW connectivity. Agents can travel to the World across the Internet while their minds remain where they are hosted [1]. The World has contained within it food cells, a nest and Agents. The World monitors the behavior of the Agents and provides some rewards depending on the outcome of actions that the Agent takes. The World server responds to some action by an Agent by updating the state of the World and sending the new state back to the Agent. The Agent(s) respond to the current state of the World by selecting an action to be taken and sending this new action to the World. If the Agent specifies an action that results in the Agent arriving at a Food Cell, then the Agent will receive a reward from the World and information pertaining to the new state of the World. If the Agent specifies an action in the World that results in the Agent arriving at the Nest, this will also result in the World rewarding the Agent. The World will respond with a Zero reward to the Agent, (that is the equivalent of punishment), when the Agent has specified an action to perform in the World which when enacted by the World does not result in (a) finding food or (b) returning to the nest. This environment is similar to that used by Humphrys [2][3] with the exception that there is also communication between agents. In order for this to be a true communications system rather than an information extraction system then the sending agent (Instructor/Speaker) must obtain some

A. de Antonio, R. Aylett, and D. Ballin (Eds.): IVA 2001, LNAI 2190, pp. 239-241, 2001.

reward that would not otherwise have been obtained [4]. There must be some sense of fairness built into this system where cheating is not rewarded (but is allowed) otherwise the population could never learn. Conflicts between rewards that come from the World and rewards that come from obeying the other Agents must be resolved to allow the system to evolve. This can be achieved by summing coincidental rewards from both World and Speaker or also by weighting the rewards accordingly.

The agent performs random acts until their Reinforcement Learning Network brain has developed so that learned actions can be performed. After they have visited a cell more than once they can "decide" whether an action should be taken or not depending on the Agents estimated reward for that action. Q Learning deals with delayed reinforcement and is described in [3]. In summary if a transition leads to a reward, then the Q values of intermediate steps that are taken which lead to this state are also increased but not by the same amount. This diminished reward decreases the further the intermediate steps are away from the rewarding state. Linguana acting under instruction, the acting Linguana (Hearer) carries out an action based on what the instructing Linguana (Speaker) has told it to do. If the Hearer Linguana performs the correct action (as instructed by the Speaker Linguana) then the Speaker Linguana rewards the Hearer. If, as result of this action, food is found or the nest is located, then the Environment rewards the Hearer Linguana. The Speaker constructs a "token-action" pair to map the instruction to an action and stores this mapping in a table. The Hearer receives the token and checks to see if it already has an action (meaning) linked to that token in its own table. If a mapping exists on the Hearers end then it performs the associated action, otherwise it selects an action which will be linked to the token if it elicits a reward from the Environment and the Speaker.

This symbiotic relationship between the Speaker and Hearer reinforces truthfulness as the reward functions for instructor actions only provide positive feedback when the outcome of the action benefits the actor. This means that only the instructions from the Speaker, which result in a reward for both Hearer and Speaker persist and malicious or random instructions are not rewarded. It is not in the Speakers interest to instruct the Hearer Linguana on a "fools errand" as the Speaker receives a punishment (zero reward) in this case. (Although initial instructions by the Speaker are random and can be termed malicious, they may result in random environmental rewards for the Hearer which in turn results in Lingua rewards for the Speaker). Malicious instructions are however permitted as the model better reflects real world communication.

Humphrys architecture of the WWM separates the Agent from the Environment and allows multiple agents to have their own view of the World [5]. It also facilitates inter-agent communication through actions in the world. Although the TONAL implementation of the WWM only requires a subset of its capabilities, the architecture model lends itself to many types of distributed intelligence and collaborative agent projects. The capacity to embed "foreign" language learning agents in the Environment can be used to test for noise, dialects and robustness of system where multiple tokens map to the same action.

References

[1] Humphrys, Mark (to appear), Distributing a Mind on the Internet: The World-Wide-Mind, to appear in *ECAL-01*, Prague, Czech Republic, September 10-14, 2001

[2] Humphrys, Mark (1996). Action Selection Methods using Reinforcement Learning. *SAB '96*. http://www.compapp.dcu.ie/~humphrys/

[3] Humphrys, Mark (1997) Action Selection Methods using Reinforcement Learning. PhD thesis, Cambridge University. http://www.compapp.dcu.ie/~humphrys/PhD

[4] Burghardt, G. M. (1970). Defining 'communication.'. *In Johnston, J., Moulton, D., and Turk, A., editors, Communication by Chemical Signals, pages 5--18*, New York. NY: Appleton-Century-Crofts.

[5] Walshe, Ray and Humphrys, Mark (to appear), First Implementation of the World-Wide-Mind, poster to appear in *ECAL-01*, Prague, Czech Republic, 2001.