Walshe, Ray (2001) The Origin of the Speeches: language evolution through collaborative reinforcement learning, submitted to the 6th European Conference on Artificial Life (ECAL-01)

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 project proposes that language evolve through reinforcement learning where agents communicate with each other and provide rewards if communication is successful. The fundamental difference between the learning mechanisms that humans use to communicate with one another and how machines learn to communicate is that the system used by humans presupposes that the adult already knows the meanings associated with the human language. Languages evolve historically to be optimal communication systems where human language learning mechanisms have evolved in order to learn these systems more efficiently. Machines in their learning of natural language, have to start at a place that humans mastered thousands of years ago. Uttering previously unheard signals and collectively establishing meaning. The question that this paper deals with is how can a communication system evolve if none of the conspirators have mastered the system previously using evolutionary computation and reinforcement learning

1 Introduction

In an artificial environment for learning there are those which suggest that genetic transmission between generations alone is capable of developing innate communication systems [1],[2],[3],[4],[5],[6],[7]. Genetic mechanisms neglect the origin of language which would be the logical starting point from an A-Life perspective. Others suggest that cultural transmission between generations alone is capable of developing and refining entirely learned communication systems [8],[9]. The border between what is signalling and what is language is often a contentious topic. Most linguists who study human language evolution are involved with the reconstruction of dead languages from structures that exist in human languages today. By analysing the current state of language it is extrapolated how previous languages may have been optimised or evolved to give us what we have today. It can never be proven if ancient languages actually had the properties being postulated as there are no records. Cangelosi [10] stated that the evolution of language implies the co-evolution of an ability to respond appropriately to signals (language understanding) and the ability to produce the appropriate signals in the appropriate circumstances (language production).Analysing how animals use signals to communicate may give us a clearer insight into how human language may have evolved but this is not the thrust of this paper. A system needs to be designed for artificial autonomous agents which has built into it the capability for a communications system to evolve. Using observed results from how the Agents behave in this model will also give insight into how other animal communications evolve. Multi-agent models of language evolution usually involve agents giving names to internal independently constructed categories [11]. Kaplan presented an approach in which the creation of categories is part of the language formation process itself. Further work by Steels and Kaplan where a lexicon emerged in a group of autonomous distributed agents situated and grounded in an open environment extends the evolutionary approach to agent communication research.. Because the agents are autonomous, grounded, and situated, the possible words and possible meanings are not fixed but continuously change as the agents autonomously evolve their communication system and adapt it to novel situations [12]. Recently there has been a lot of research into evolutionary approaches to communication systems.[13], [14].

2 The Experiment

The experiment consists of a World where Agents exist and the Agents themselves. The World has contained within it food cells, a nest and Agents. The World monitors the behaviour of the Agents and provides some rewards depending on the outcome of actions that the Agent takes. If the Agent arrives at a *Food Cell*, then the Agent will receive a reward from the World. Arriving at *the Nest* will also gain a reward from the World. The Agent receives a Zero reward (that is the equivalent of punishment) when it performs and action which does not result in finding food or returning to the nest. This environment is similar to that used by Humphrys [15],[16] with the exception that there is communication between agents. This environment can also be used here where the Agents themselves are explorers in the World and can do the following:

Table 1. Agents possible actions

Perform a random action Agent selects any action at random.
Perform a learned action Agents move based on probabilities of rewards built up from previous experience.
Instruct a random action (random signal) Agents can instruct other agents to perform an action by associating a token with it.
Instruct a learned action (learned signal) Agents can instruct other agents with pre-associated token-action pairs built up based on previous experience.
Operate in isolation(don't listen) Agents can ignore other agents.
Operate in co-operation (listen) Agents can send/receive instructions.

In order for this to be a true communications system rather than an information extraction system then the sending agent (instructor/Adult) must be obtain some reward that would not otherwise have been obtained [Burghardt 1970]. There most 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.

2.1 The Environment

The World is a landscape consisting of an 8x8 matrix organised as a torus where there are no edges. The World simply wraps around on itself.*(Fig. 1)*

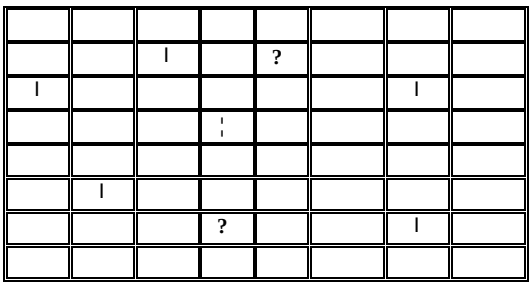


Fig.1. The World: ***Key*** **I** ***(*** *Food)* **?** ***(*** *Linguana) and* ¦ *(Nest)*
The Adult and Baby contexts have no special meaning in the diagram as both agents can be either adult or baby

This environment has contained within it randomly distributed food cells and a nest *(Figure 1).* Language Agents known as Linguanas inhabit the World. This environment rewards the Linguanas if they find food or return to the nest. Any cell in the world is a valid location for the nest, food or any of the Linguanas. When the Linguana lands on a food cell, it automatically picks up the food and gets a reward. When the Linguana is carrying food and lands on the nest cell, it automatically drops the food and gets a reward.

3 The Client Server Model

Three scenarios were investigated whereby the World (Server) would communicate with the Agent (Client). These scenarios were developed to maximise portability and re-usability of the environment and the learned brain of the agent.

3.1 Peer To Peer Socket Communication

The World reads and writes output to stdin/stdout through files (Fig.2.). The Agent reads and writes output to stdin/stdout through files. The passing of information between the World and the Agent is effected by providing a Server Process to handle the World communication and a Client Process to handle the Agent communication. This scenario allowed multiple agents to interact with the central world and establish a communication system, however the environment and the clients were restricted to the Local Area Network where the Server resided and the Client and Server Processes had to be running on different machines.

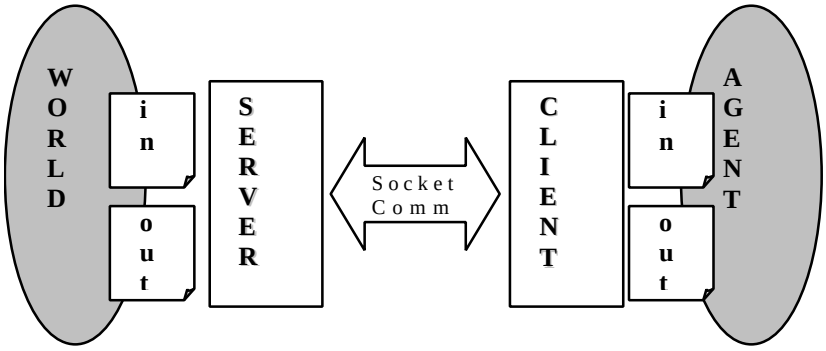


Fig.2. Communication on LAN using stdin/stdout via files. Server and Client Processes are communicating through a socket.

3.2 Peer To Peer HTTP/Socket Communication

The second scenario that was investigated to reduce the Client Process overhead was to have Web Browsers to send http instructions to the Server Process. The Server to parse the information to and from the World or Browser. This allowed access to the World by Multiple Web Clients but required a large amount of effort to write a World Specific Web Server.

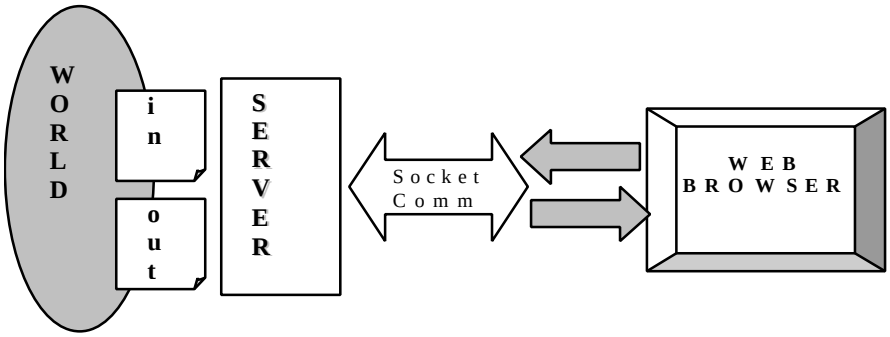
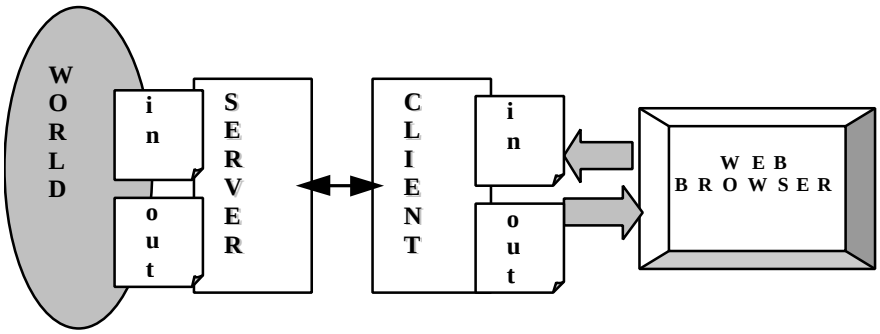


Fig.3. Communication on LAN using stdin/stdout via files. Server and Web Browser are communicating through a socket.

3.3 Peer To Peer cgi-bin Script Communication

The final method which was adopted consisted of writing a file oriented interface between client and server where cgi-bin scripts written in perl were used to transfer command to agents and return the state of the world to the user.

Fig.4. Communication unrestricted (Not LAN Based) using stdin/stdout via files. Server and Client are socket based. Web Browsers are



communicating (via cgi-bin script) through files.

This concept if distributing the Agent Minds and Agent Worlds so that they can be geographically anywhere on the planet is a new departure for AI environments and a more detailed description is given by Humphrys.[

4 Reinforcement Learning

This type of Reinforcement Learning (Q Learning) was introduced by Watkins [17] where agents exist within a world that can be modelled as Markov Decision Process (MDP). The Agent observes the state x (e X , a finite set) of the World at any given time and performs a discrete action a (e A , a finite set). The Agent is aware of the current state of the World x, takes action a and the observes the World change into state y and receives immediate reward r. This Learning Network is used to build up the brain of the Linguana based on previous experience. Transitions are probabilistic $P_{xa}(y)$ and $P_{xa}(r)$, where $P_{xa}(y)$ is the probability that taking action a in state x will lead to y and $P_{xa}(r)$ is the probability that taking action a in state x will result in reward of r . So we have

$$\mathbf{S}_y ?_{xa}(y) = 1 \text{ and } \mathbf{S}_r ?_{xa}(r) = 1$$

Linguanas have two modi operandi. They can act on their own (Cellular Mode) or they can act under instruction. (Collaborative Mode).

4.1 Cellular Mode:

The agent performs random acts (a_i e A) at first until their RL 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 E(r) for that action.

$$\mathbf{E}(r) = \mathbf{S}_r r ?_{xa}(r)$$

$$\mathbf{Q(x,a) := (1-a)\ Q(x,a) + a(r + ?max_b\ Q(y,b))}$$

4.2 Collaborative Mode:

Linguana acting under instruction, the acting Linguana (baby) carries out an action based on what the instructing Linguana (adult) has told it to do. If the baby Linguana performs the correct action (as expected by the adult Linguana) then the adult Linguana rewards the baby. If as a result of this action, food is found or the nest is located, then the Environment rewards the baby Linguana.

4.2.1 The Adult Linguana

- Performs random acts and gets a reward from the environment
- Instructs baby Linguana with “token” to perform an action and provides baby Linguana with a callback function for rewards
- Stores the “token – action” pairs
- Gets a reward if the baby Linguana gets a reward from the environment

As Agents can acts as both Adult and Baby, a type of semaphore is used for gaining control of the Master-Slave relationship i.e. Which agent is instructing and which is acting.

4.2.2 The Baby Linguana

- Performs random acts and gets reward from environment
- Stores current/previous position pairs
- Receives “token” from Adult Linguana and performs
  (a) Random Action
  (b) Learned Action based on Token
- If rewarded store “token – action” pair

4.3 The Tokens

The token sent/received by the Adult or Baby will initially be random strings. Once positive feedback has occurred the “token – action” pairs stored (*Table 1*) by both the Adult and the Baby Linguanas will be the same. If the Baby initiates the “conversation” then multiple strings (tokens) can map to the same action. Pruning will take care of the ambiguity by utilising negative feedback on “token – action” pairs.

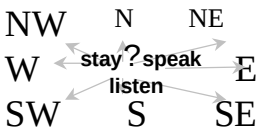
Table 2. Token-Action pair combinations.

Token	Action
Oomph	Go_north()
Blah	Go_south_west()
Aargh	Stay_still()

4.4 The Actions

Table 3. The possible actions are:

Action N (north)	Linguana moves north , Position co-ordinate Y increases.
Action S (south)	Linguana moves south , Position co-ordinate Y decreases
Action E (east)	Linguana moves east , Position co-ordinate X increases
Action W(west)	Linguana moves west , Position co-ordinate X decreases
Action NE(north-east)	Linguana moves north-east , X increases, Y increases
Action SE (south-east)	Linguana moves south-east , X decreases, Y decreases
Action NW (north-west)	Linguana moves north-west , X decreases, co-ordinate Y increases
Action SW (south-west)	Linguana moves south-west , X decreases, Y decreases
Action STAY (stay)	Linguana stays put , Position co-ordinates X and Y no-change.
Action SPEAK(send token)	Linguana stays put , X/Y no-change, Token-Action pair saved. Action reward function provided.
Action LISTEN(get token)	Action is based on token , X/Y may change. Token-Action pair saved. Reward function executed



4.5 Current State

The state of the World at any given time as perceived by the agents is *x*. The Agent needs to be able to sense the nest, food, other agents and if it is currently carrying food. It needs to inform other agents when speaking. Adopting a notation from Humphrys [15] the following is assigned.

- ***i*** is whether Agent is carrying, not carrying food (1 = C ,0 = NC)
- ***n*** (0-8) is the direction of the nest. (Known at any time)
- ***f*** (0-9) is the direction of the nearest visible food. (Vision is limited, 9 = not visible)
- ***l*** (0-9) is the direction of the nearest visible Linguana. (Vision is limited, 9 = not visible)

Thus state **x = (*i, n, f, l*)**

9(Not Visible)

0	**1**	**2**
7	**8 Here**	**3**
6	**5**	**4**

Fig.5. The Agent senses direction of things in a small radius.

Actions that the creature can take *a* take values 0-7 (move in that direction), 8 (stay), 9 (Speak) and 10 (Listen)

4.6 Cheating and malicious instructions.

This symbiotic relationship between the Adult and Baby reinforces truthfulness as the reward functions for instructor actions only provide positive when the outcome of the action benefits the actor. This means that only the instructions from the Adult which result in a reward for both Baby and Adult persist and malicious or random instructions are pruned via negative feedback.

5 Feudal Q Learning

Watkins' Feudal Q-learning [Watkins, 1989] shows another way of having agents use other agents - by sending explicit orders. In *Feudal Q-learning*, a master agent sends a command to a slave agent, telling it to take the creature to some state (which the master agent may not know how to reach on its own). The slave agent receives rewards for succeeding in carrying out these orders. To formalise, the master has a current command *c* in operation. This actually forms part of the `state' of the slave. Using the notation

$$(x,c),a \rightarrow (y,c)$$

the slave will receive rewards for transitions of the form

$$(*,c),* \rightarrow (c,c)$$

Note that immediately the command *c* changes, we jump to an entirely new state for the slave and it may start doing something entirely different. Given a slave that has learnt this, the master will learn that it can reach a state by issuing the relevant command. Using the notation

$$x,a \rightarrow y$$

it will note that the following will happen:

$$(*),(c) \rightarrow (c)$$

6 Action Rewards

6.1 Environmental Rewards

When the Linguana performs an action of relocation and calls the environment reward callback function, the World will reward the Linguana if:

- The result of the action brings the Linguana to a Food Cell.
- The Result of the action brings the Linguana to The Nest.

6.2 Lingua Rewards

When the baby Linguana responds to an instruction it has two sources of reward

- From the Environment (see above)
- From the Adult Linguana

The Adult Linguana will reward the baby Linguana via its reward callback function if

- The action performed was the one instructed by the Adult Linguana

Conclusions

Although the system is stable the communication system that evolves is restricted. The limits on the system are primarily due to the Agents inadequacy as the Agent has limited mobility and dexterity. With more actions available to the agent and a greater number of combinations and permutations of actions the greater the complexity of the evolved communications system. Collaboration is successful as agents achieve results (rewards) faster. The same results would have taken longer in Cellular Mode.

Further Work

A larger system needs to be coded which allows interaction between multiple Agents in the World. This is already underway and will be the subject of future publications.

References

1. A.Cangelosi and D. Parisi. The emergence of language in an evolving population of neural networks. In Proceedings of the 18th Conference of the Cognitive Science Society, San Diego, 1996
2. De Bourcier, P. and M Wheeler (1997). The truth is out there : The evolution of reliability in aggressive communications systems. In P.Husbands and I. Harvey (Eds); Fourth European Conference on Artificial Life; pp. 444 453. Cambridge; MA: MIT Press.
3. Di Paulo, E.(1997). An investigation into the evolution of communication. Adaptive Behaviour 6, 285 324.
4. Werner, G. and P. Todd (1997). Too manty love songs: Sexual selection and the evolution of communication. In P.Husbands and I. Harvey (Eds); Fourth European Conference on Artificial Life; pp. 434 443. Cambridge; MA: MIT Press.
5. Noble, J. (1998). Evolved signals: Expensive hype vs. conspiratorial whispers. In C. Adami, R. Belew, H. Kitano, and C. Taylor (Eds), Artificial Life 6: Proceedings of the Sixth International Conference on Artificial Life. Cambridge, MA: MIT Press.

6. Oliphant, M. The dilemma of Saussurean communication. Biosystems, 37 (1-2):31-38.
7. Bullock, S. (1997). An exploration of signalling behaviour by both analytic and simulation means for both discrete and continuous models. In Husbands, P., & Harvey, I. (Eds.), 14 Proceedings of the Fourth European Conference on Artificial Life (ECAL'97), pp. 454--463. MIT Press / Bradford Books, Cambridge, MA.
8. Hutchins, E. and B. Hazelhurst (1995). How to invent a lexicon: The development of shared symbols in interaction. In N. Gilbert and R. Conte (Eds), Artificial Societies: the computer simulation of social life. London: UCL Press.
9. Steels, L. and P. Vogt (1997) Grounding adaptive language game sin robotic agents. In P.Husbands and I. Harvey (Eds); Fourth European Conference on Artificial Life; pp. 474 482. Cambridge; MA: MIT Press.
10. Cangelosi, A. & Parisi, D. (1996). The emergence of a "language" in an evolving population of neural networks. 18th Conference of the Cognitive Science Society, San Diego.
11. F. Kaplan. A new approach to class formation in multi-agent simulations of language evolution. In Proceedings of the Third International Conference on Multi Agent Systems ICMAS'98. IEEE Computer Society Press, 1998.
12. L. Steels and F. Kaplan. Situated grounded word semantics. In T. Dean, editor, Proceedings of the Sixteenth International Joint Conference on Articial Intelligence IJCAI'99, pages 862-867, San Francisco, CA., 1999. Morgan Kaufmann Publishers.
13. Vogt, P. Grounding Language About Actions: Mobile Robots Playing Follow Me Games (2000)
14. De Jong, E. D. (1998). Coordination developed by learning from evaluations. AI-MEMO 98-13, Vrije Universiteit Brussel, AI Lab.
15. Humphrys, Mark (1996). Action Selection Methods using Reinforcement Learning. In P. Maes et. al. (Ed s), From Animals to Animats: Proceedings of the Fourth International Conference on Simulation of Adaptive Behaviour: MIT Press/Bradford Books.
16. Humphrys, Mark (1997) Action Selection Methods using Reinforcement Learning. PhD thesis, Cambridge University. http://www.compapp.dcu.ie/~humphrys/
17. Watkins, C. J.C.H. (1989). Learning from delayed rewards, PhD thesis, University of Cambridge, Psychology department.
18. Humphrys, Mark (2001), Distributing a Mind on the Internet: The World-Wide-Mind, submitted to the *6th European Conference on Artificial Life (ECAL-01)*

The Origin of the Speeches: language evolution through collaborative reinforcement learning.

Table of Contents

Abstract....3
1 Introduction....3
2 The Experiment....3
2.1 The Environment....3
3 The Client Server Model....4
3.1 Peer To Peer Socket Communication....4
3.2 Peer To Peer HTTP/Socket Communication....4
3.3 Peer To Peer cgi-bin Script Communication....4
4 Reinforcement Learning....5
4.1 Cellular Mode:....5
4.2 Collaborative Mode:....5
4.2.1 The Adult Linguana....6
4.2.2 The Baby Linguana....6
4.3 The Tokens....6
4.4 The Actions....6
4.5 Current State....7
4.6 Cheating and malicious instructions.....7
5 Feudal Q Learning....7
6 Action Rewards....8
6.1 Environmental Rewards....8
6.2 Lingua Rewards....8
Conclusions....8
Further Work....8
References....8