Sequential Task Problem Solving using Cultural Learning in Populations of Neural Networks

Dara Curran and Colm O'Riordan

Department of Information Technology,
NUI,Galway

Abstract. Cultural learning describes the process of information transfer between individuals in a population through non–genetic means. Typically this is achieved through communication or the creation of artifacts available to all members of a population. Cultural learning has been simulated by combining genetic algorithms and neural networks using a teacher/pupil scenario where highly fit individuals are selected as teachers and instruct the next generation.
This paper explores the effect of a cultural learning approach to the development of solutions for three test–case sequential decision tasks: connect–four, tic–tac–toe and blackjack. Experiments are conducted with populations employing population learning alone and populations combining population and cultural learning.

1 Introduction

A number of learning models may be readily observed from nature and have been the focus of much study in artificial intelligence research. Population learning (i.e. learning which occurs at a population level through genetic material) is typically simulated using genetic algorithms. Life-time learning (i.e. learning which takes place during an organisms's life time through reactions with its environment) can be simulated in a variety of ways, typically employing neural networks or reinforcement learning models.

A relatively new field of study in artificial intelligence is synthetic ethology. The field is based on the premise that language and culture are too complex to be readily analysed in nature and that insight can be gained by simulating its emergence in populations of artificial organisms. While many studies have shown that lexical, syntactical and grammatical structures may spontaneously emerge from populations of artificial organisms, few discuss the impact such structures have on the relative fitness of individuals and of the entire population.

The focus of this paper is to attempt to understand the effect of cultural learning on a population of artificial organisms attempting to find solutions to three distinct sequential decision problems. The remainder of this paper is arranged as follows. Section 2 introduces background research, including descriptions of diversity measures and cultural learning techniques that have been employed for this study. Section 3 describes the experimental setup. Section 4 presents the a description of each experiment as well as their results and Section 5 provides a conclusion.

2 Background research

2.1 Cultural Learning

Culture can be succinctly described as a process of information transfer within a population that occurs without the use of genetic material. Culture can take many forms such as language, signals or artifactual materials. Such information exchange occurs during the lifetime of individuals in a population and can greatly enhance the behaviour of such species. Because these exchanges occur during an individual's lifetime, cultural learning can be considered a subset of lifetime learning.

An approach known as synthetic ethology [10, 18] argues that the study of language is too difficult to perform in real world situations and that more meaningful results could be produced by modelling organisms and their environment in an artificial manner. Artificial intelligence systems can create tightly controlled environments where the behaviour of artificial organisms can be readily observed and modified. Using genetic algorithms, the evolutionary approach inspired by Darwinian evolution, and the computing capacity of neural networks, artificial intelligence researchers have been able to achieve very interesting results.

A number of approaches were considered for the implementation of cultural learning including fixed lexicons [21, 4], indexed memory [17], cultural artifacts [8, 3] and signal–situation tables [10]. The approach chosen was the teacher/ pupil scenario [2, 7, 4] where a number of highly fit agents are selected from the population to act as teachers for the next generation of agents, labeled pupils. Pupils learn from teachers by observing the teacher's verbal output and attempting to mimic it using their own verbal apparatus. As a result of these interactions, a lexicon of symbols evolves to describe situations within the population's environment.

2.2 Sequential Decision Tasks

Sequential decision tasks are a complex class of problem that require agents to make iterative decisions at many steps throughout the task. Each decision has a direct effect on the agent's environment and in turn affect its subsequent decisions. Our selection of a number of games was driven by two main factors: games are good examples of sequential decision tasks and many artificial intelligence implementations exist for ready comparison and analysis.

The games we chose as a test-bed for cultural learning are roughly grouped in perceived order of difficulty, beginning with tic-tac-toe following with the game of blackjack and concluding with the game of connect-four.

3 Experimental Setup

The following set of experiments each employs two populations. One population is allowed to evolve through population learning (by genetic algorithm), while

the other employs both population and cultural learning. The experiments are carried out using an artificial life simulator developed by Curran and O'Riordan [6] capable of simulating population and cultural learning.

Cultural learning is implemented based on a scheme developed by Hutchins and Hazlehurst [8] and further explored by Denaro [7] where the last hidden layer (or in Denaro's case, the output layer) of a neural network functions as a verbal input/output layer. At the end of each generation, a percentage of the best individuals in the population is selected to instruct the next. Pupil networks observe teacher networks as they interact with their environment and at each stimuli, teacher networks produce an utterance through their verbal I/O layer. The pupil responds to the utterance with its own, which is then corrected by back-propagation to approximate the teacher's. After the required number of these interactions (teaching cycles) have been completed, the teachers are removed from the population and the pupils continue to interact with their environment.

In previous work by Parisi *et al.*[7], it was suggested that the addition of noise to a teacher's verbal output could enhance a population's ability to retain culturally acquired information. This is parameter was implemented in the simulator and generates noise in the range [-0.5,0.5] to the teacher's output when instructing a pupil with probability n.

4 Experiments

4.1 Tic Tac Toe

Tic–tac–toe, or three in a row is a very simple two player game played on a 3x3 board. Each player is assigned either the X or O symbol and takes turns placing one symbol onto the board at a time. Each player attempts to place three of his/her pieces in a horizontal, vertical or diagonal line of three.

In order to evolve good players, it was decided that agents in the population would all compete against a perfect player rather than compete against each other. To avoid over-fitting, the perfect player's first move is randomised to provide game diversity. It was felt that populations of agents competing against each other would be likely to converge only to local maxima due to the lack of competitive pressure.

The presence of a perfect player in the population should not be construed as an example of a solution set from which the population is modeled. The perfect player is a minimax implementation and merely provides an incentive for the population to improve. It does not provide a neural network implementation of a perfect tic-tac-toe player.

Each agent's neural network structure contains 18 input nodes, 2 for each board position where 01 is X, 10 is O and 11 is an empty square. Nine output nodes corresponding to each board position are used to indicate the agent's desired move. The node with the strongest response corresponding to a valid move is taken as the agent's choice.

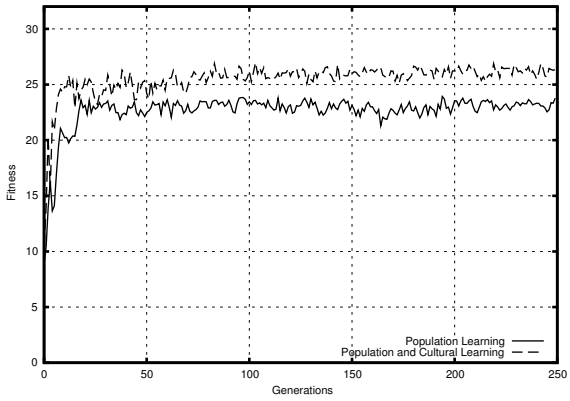

Fig. 1. Tic Tac Toe Experiment

Since the agents play against a perfect player, fitness is assigned according to how long each agent is capable of avoiding a loss situation. An agent's fitness is therefore correlated with the number of moves that each game lasts, rewarding agents capable of forcing the perfect player to as close to a draw as possible. The fitness function produces values in the range [0,32], where 32 is the maximum fitness (the situation where the agent draws all four games).

Populations of 100 agents were generated for these experiments and allowed to evolve for 250 generations. Crossover was set at 0.6 and mutation at 0.02. The cultural learning settings of teacher ratio and teaching cycles were set at 0.1 and 5 respectively. Cultural mutation was set at 0.05. These parameters were determined empirically to provide the best performance.

Experimental Results Two experiments were undertaken: one using only population learning to evolve players, and the other using population and cultural learning. Figure 1 shows the average fitness values for the two evolving populations. While both types of learning begin at similar levels of fitness, there is strong evidence (p value < 0.0001, 95% C.I.) to suggest that agents employing cultural evolution are performing better as the experiment progresses.

Population	Average Fitness	Standard Deviation
Population Learning	25.48	1.64
Cultural Learning	22.63	1.77

Discussion It is interesting to compare these results with those obtained by Angeline and Pollack [1] who used a competitive fitness function to evolve populations of neural network tic–tac–toe players. The population of evolving players was pitted against a number of 'expert' player strategies, including a perfect player. If we examine their results in terms of a draws/losses ratio, we find that their best evolved players (playing against a perfect player) obtain a ratio

of 0.2405. By contrast, the cultural learning approach presented in this paper obtains an average of 0.72 with highs of 0.94 and lows of 0.625.

4.2 The Game of BlackJack

Blackjack or twenty–one begins with the dealer dealing two cards face–up to each player and two to his/herself, with one card visible (the *up–card*) and the other face down. Cards are valued by their face value (10 for all picture cards) except for the ace which can be counted either as 11 or 1. The object of the game is to obtain a higher score (the sum of all card values) than that of the dealer's without exceeding 21. Each player can *draw* additional cards until they either *stand* or exceed 21 and go *bust*. Once all players have obtained their cards, the dealer turns over the hidden card and draws or stands as appropriate. Should the dealer's hand bust, all players win.

The dealer is at considerable advantage because he/she only enters the game once all players have fully completed their play. Thus, it is probable that some players will have bust even before the dealer reveals the hidden card. In addition, the fact that only one of the dealer's cards is visible means that players must make judgements based on incomplete information. As a rule, the dealer follows a fixed strategy, typically standing on a score of 17 or more and drawing otherwise.

All aspects related to betting such as doubling down and splitting have been removed from this implementation and only one deck is used in each game. This is in order to facilitate comparison with previous work which employs a similar approach.

Several attempts have been made to develop high performing blackjack strategies with populations of neural networks using reinforcement learning techniques[12, 13]. The nature of the game means that there is no perfect set of neural network outputs from which to perform back–propagation. It is for this reason that we wish to show that the introduction of cultural learning can generate superior strategies than reinforcement learning methods and provides the learning framework required without knowledge of the perfect strategy.

In this implementation, each agent's neural network is given information about the card value currently held, as well as a flag indicating the presence of an ace. In addition, each neural network is given the value of the dealer's upcard. Each experiment allows 100 agents to evolve over 500 generations. At each generation, agents play 100 games against a dealer strategy and an agent's fitness is determined by the percentage of wins obtained scaled to [0.0,1.0]. Crossover was set at 0.6 and mutation at 0.02. The cultural learning settings of teacher ratio and teaching cycles were set at 0.1 and 5 respectively. Cultural mutation was also added with probability 0.05.

Experimental Results The graph in figure 2 shows that the addition of cultural learning, allows the population to perform substantially better than population learning alone, achieving highs of over 0.45 (45% wins) versus 0.44 for population learning. The evolved strategy outlined below was extracted from

the population by examining the neural network response to all possible card values.

Population	Percentage Wins	Standard Deviation
Cultural Learning	43.0149	1.97
Population Learning	42.6455	2.24

```
if (an Ace is held)
{
    if (dealer has a 6 or higher)
        stand on 16
    else
        stand on 17
}
else
{
    if (dealer has a 7 or higher)
        stand on 17
    else
        stand on 13
}
```

This time there is strong evidence ($p < 0.05$, 95% C.I.) to suggest that cultural learning agents are indeed statistically different than population learning agents. It is clear from the strategy that the evolved agents are employing the new dealer and ace information to the full extent and have identified a threshold value for the dealer up–card. The strategy is tested in the next section to ascertain its performance with respect to the bench–marked strategies.

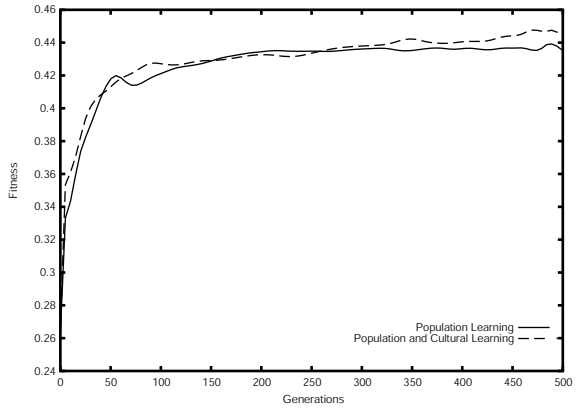

Fig. 2. *Blackjack Experiment*

It is interesting that cultural learning only appears to provide improvement in this last set of experiments. On examination of evolved strategies for both

populations in the previous two experiments, it is clear that both populations have reached the optimum possible given the amount of information provided. However, the addition of dealer information creates greater complexity which population learning alone is incapable of tracking.

Strategy Testing In order to assess the performance of the evolved strategy, a set of bench–marks must be obtained for comparison purposes. This was achieved using a blackjack simulator consisting of a dealer, who employs the traditional dealer strategy of standing on 17 or greater, and a single player whose strategy can be set at the beginning of the simulation. The evolved strategy was compared to a number of strategies, including an evolved strategy developed by Uribe and Sanchez[13] and 1000 runs of 1000 games were performed for each strategy to produce statistically significant results.

Strategy	**Percentage Wins**	**Standard Deviation**
Hoyle	43.70	1.587
Evolved Strategy	**43.67**	**1.582**
Dealer	41.52	1.576
Sanchez et al	38.43	1.505
Always Stand	38.00	1.531
Random	30.67	1.511

The results of the simulation show that the evolved strategy does not quite reach the level of Hoyle's strategy but is very close. On examination of the standard deviations, it is clear that the top two strategies are very similar, suggesting that the population has evolved an optimum strategy given the information available. It is likely that in order to out–perform Hoyle's strategy it is necessary to keep track of cards that have been played during a game, something which would only become truly useful if the number of players was increased.

Discussion The results presented show that cultural learning provides a modest improvement on population learning, provided that sufficient environmental information is present. It is clear that the addition of dealer information to the population significantly improves the performance of both population learning and cultural learning.

While these improvements are small, it is worth remembering that the game of blackjack is inherently very noisy and odds are very much stacked in favour of the dealer. Consequently, any statistically significant improvement such as we have shown, represents an achievement on the part of the evolutionary process. Through the bench-marking process we have shown that the evolved strategy is equivalent to the best human strategy which does not incorporate card-counting.

4.3 Connect Four

The game of connect–four is a two–player game played on a vertical board of 7x6 positions into which pieces are slotted in one of seven available slots. Each player

is given a number of coloured pieces (one colour per player) and must attempt to create horizontal, vertical or diagonal piece–lines of length four. Players place one piece per turn into one of the seven slots. The piece then falls onto a free position in the chosen column, creating piles, or towers, of pieces. If a column is full, the player must select an available slot.

Some research has been undertaken in the evolution of connect–four players employing a library of existing games to train the neural networks by back–propagation [15] as well as reinforcement learning methods [16].

While the game appears simple, a certain amount of tactical knowledge is required to play proficiently. The most obvious approach is to scan the board for existing lines of three and either finish them to create four–in–a–line, or if the line is the opponent's, block it. However, as is the case in many games, the best approaches focus on forcing the opponent to contribute to the player's victory, requiring more complex strategies.

In order for a population of neural networks to play games of connect–four, a method must be developed to encode both the board's current position and decode the network's output into a valid move. Following a number of empirical trials examining a number of techniques, the best approach, dubbed Multiple Board Selection, was chosen for this set of experiments.

Multiple board selection, presents a neural network with all possible board positions resulting from each of the moves available. At each move, the neural network's single output node records the network's estimation of the board position's worth. The move producing the best board position (according to the neural network output value) is taken to be the agent's preferred choice and is chosen as the agent's next move.

Experimental Results A population of 20 agents were allowed to evolve for 100 generations. At each generation, agents play in a tournament against all other players. In addition, each agent plays a minimax player with three levels of difficulty. In total, each agent plays 22 games of connect–four in its lifetime. Agents are assigned fitness according to each game's result: 3 points for a both a win and a draw and 0 points for a loss. This gives a fitness range of [0,66]. Teachers play a full tournament while pupils observe, and at each move, the teacher corrects the pupil's perception through back–propagation.

Crossover was set at 0.6 and mutation at 0.02. The cultural learning settings of teacher ratio and teaching cycles were set at 0.1 and 5 respectively. Cultural mutation was also added with probability 0.05. The results in Fig. 3 show that the addition of cultural learning provides the best performance and that the fitness levels show an upward trend at the end of the experiment, suggesting that the population is capable of further improvement.

Discussion There is strong evidence (p value < 0.05, 95% C.I.) to suggest that the increase in performance brought by the addition of cultural learning to the population is statistically significant . It is therefore possible to conclude from

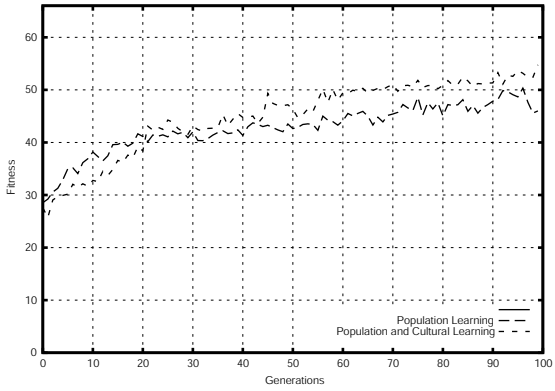

Fig. 3. *Connect Four Experiment*

this set of results that cultural learning improves the performance of agents in the connect–four environment.

5 Conclusion

This paper presents a set of test–cases which highlight the potential of cultural learning in sequential decision problems. The results indicate that cultural learning provides improved performance over population learning in each test–case. We have shown that unlike traditional life–time learning techniques of neural network optimisation such as gradient descent, cultural learning does not require explicit solution information.

Cultural learning gives populations the opportunity to sample acquired information within the population itself. This allows weaker members of the population to gain access to environmental information which would otherwise be impossible to attain without incurring possible fitness losses. In addition, experiments such as these provide a possible explanation of complex behaviour in nature: since no perfect solution is possible for a given environmental situation, organisms are not capable of receiving direct error feedback in the manner of synthesised life–time learning simulations. Instead, they must either rely on purely genetic information, or develop a mechanism for imparting useful knowledge to the next generation.

Future work will concentrate on more complex sequential decision problems and examine the effect of dynamic environments on populations employing cultural learning.

References

1. P. J. Angeline and J.B. Pollack. Competitive environments evolve better solutions for complex tasks. In S. Forrest, editor, *Proceedings of the Fifth International Con-*

ference on Genetic Algorithms, pages 264–270, San Francisco, CA, 1993. Morgan Kaufmann.

2. A. Billard and G. Hayes. Learning to communicate through imitation in autonomous robots. In *7th International Conference on Artificial Neural Networks*, pages 763–738, 1997.
3. A. Cangelosi. Evolution of communication using combination of grounded symbols in populations of neural networks. In *Proceedings of IJCNN99 International Joint Conference on Neural Networks (vol. 6)*, pages 4365–4368, Washington, DC, 1999. IEEE Press.
4. A. Cangelosi and D. Parisi. The emergence of a language in an evolving population of neural networks. *Technical Report NSAL–96004, National Research Council, Rome*, 1996.
5. N. Chomsky. On the nature of language. In *Origins and evolution of language and speech*, pages 46–57. Annals of the New York Academy of Science, New York. Vol 280, 1976.
6. D. Curran and C. O'Riordan. On the design of an artificial life simulator. In V. Palade, R. J. Howlett, and L. C. Jain, editors, *Proceedings of the Seventh International Conference on Knowledge-Based Intelligent Information & Engineering Systems (KES 2003)*, University of Oxford, United Kingdom, 2003.
7. D. Denaro and D. Parisi. Cultural evolution in a population of neural networks. In *M.Marinaro and R.Tagliaferri (eds), Neural Nets Wirn-96.New York: Springer*, pages 100–111, 1996.
8. E. Hutchins and B. Hazlehurst. Learning in the cultural process. In *Artificial Life II, ed. C. Langton et al.* MIT Press, 1991.
9. E. Hutchins and B. Hazlehurst. How to invent a lexicon: The development of shared symbols in interaction. In N. Gilbert and R. Conte, editors, *Artificial Societies: The Computer Simulation of Social Life*, pages 157–189. UCL Press: London, 1995.
10. B. MacLennan and G. Burghardt. Synthetic ethology and the evolution of cooperative communication. In *Adaptive Behavior 2(2)*, pages 161–188, 1993.
11. A. H. Morehead and G. M. Smith. *Hoyle's Rules of Games*. Plume, 1963.
12. D. K. Olson. *Learning to Play Games from Experience: An Application of Artificial Neural Networks and Temporal Difference Learning*. Pacific Lutheran University, 1993.
13. Andrs Prez-Uribe and Eduardo Sanchez. Blackjack as a test bed for learning strategies in neural networks. In *International Joint Conference on Neural Networks, IJCNN'98*, pages 2022–2027, 1998.
14. H. Maisel R. R. Baldwin, W. E. Cantey and J. P. McDermott. The optimum strategy in blackjack. In *Journal of the American Statistical Association*, 1956.
15. M. O. Schneider and L. Garcia Rosa J. Neural connect four, a connectionist approach to the game, 2002.
16. P. Sommerlund. Artificial neural networks applied to strategic games, 1996.
17. L. Spector and S. Luke. Culture enhances the evolvability of cognition. In *Cognitive Science (CogSci) 1996 Conference Proceedings*, 1996.
18. L. Steels. The synthetic modeling of language origins. In *Evolution of Communication*, pages 1–34, 1997.
19. E. O. Thorp. *Beat the Dealer*. Random House, 1966.
20. E. O. Thorp. *The Mathematics of Gambling*. Lyle Stuart, 1984.
21. H. Yanco and L. Stein. An adaptive communication protocol for cooperating mobile robots, 1993.