

A Neuroevolutionary Approach to Draughts Playing Agents

Student Name: T. P. Nguyen

Supervisor Name: S. Dantchev

Submitted as part of the degree of BSc Computer Science to the
Board of Examiners in the Department of Computer Sciences, Durham University

Abstract —

Background

Presently, competitive Draughts AI players are currently designed to play at a fixed ability. While it has produced very competitive and intelligent players, they require some human intervention to improve their performance. This is due to their dependency on pre-defined move databases. Optimal moves are pre-calculated, and recalled when necessary. Through neuroevolution, this issue could possibly be solved by creating a player that can grow in ability over time, by learning to play itself.

Aims

The purpose of this project is to explore an neuroevolutionary approach to tackle the game of English Draughts. First, we study previous historical successes in the field, and look at the components that helped build their systems. Then, we look at how they can be adapted to suit the objective at hand. The project will establish whether neuroevolution is a viable approach to the problem.

Method

The method starts with designing a feed-forward neural network that evaluates the advantage a particular player, given a state of a checkerboard. This is then used in a algorithm that evaluates future moves to predict the best move at a given position. This, alongside a set of weights for the neural network, creates a player that can evaluate potential moves. Finally, the player is then used on an existing Draughts framework that will provide the player with the ability to play Draughts. Genetic algorithms are used to adjust the weights of the neural network, with the intention of making the system learn to evaluate checkerboards more precisely.

Results

The system has shown to perform considerably without any training, and has shown to improve over time. For different search depths, the net learning rate was positive. However, the wide scope of the adjustments afforded may have impacted the learning rate and has shown to be volatile at some situations. The use of crossovers may dramatically influence the quality of the learning.

Conclusion

Under the premise that the fitness function that is well defined, neuroevolution is a valid alternative to problems with novel domains. The solution is appropriate for machine learnt agents that where training does not require data. The robustness of the system is quite volatile, and is best paired with mutation and crossover methods that are not heavily reliant on entropy.

Keywords — Artificial Intelligence, Neural Networks, Genetic Algorithms, Draughts, Machine Learning, Monte Carlo Tree Search, Neuroevolution

I INTRODUCTION

The intention of this project is to explore the effectiveness of genetic algorithms to improve the neurons of a neural network. Neural networks can be used to evaluate the performance of two players in a zero-sum game of Draughts. We attempt to manipulate the neurons through the use of genetic algorithms to increase the accuracy of the evaluation. This would potentially allow us to create an effective Draughts playing agent that, when provided with the option to consider a set of moves, would have the ability to play and learn without human input.

A *Draughts*

English Draughts (or Checkers) is a popular 2-player board-game played on an 8x8 chess board. Players choose a colour and begin with 12 pawns each of their respective colours, and they are placed on the dark-coloured squares of the board. Beginning with the black pawns, each player takes a turn to move a pawn diagonally in one square. In the event that a pawn reaches the opposite side of the board from where the piece originated, it is promoted to a king. Kings have the ability to traverse backwards in the same diagonal motion as pawns. Assuming that there is space upon landing, A piece (pawn or king) also has the option to capture their opponents piece by moving two consecutive diagonal squares, where the opponents piece is situated immediately opposing the players piece. Pieces can be captured consecutively in a single turn if the moves afford the scenario. A player wins by capturing all of their opponents pieces. A player loses by having all of their pieces captured. A draw occurs when both players agree to draw after a three-fold repetition (where both players take three moves to simulate the same position), or a player has pieces on the board but cannot move any of them. A draw can be forced when both players fail to capture a piece after a set amount of moves.

B *Neural Networks*

Neural Networks are non-linear statistical data-modelling tools, linking inputs and outputs adaptively in a learning process similar to how the human brain operates. Networks consist of units, described as neurons, and are joined by a set of rules and weights. The units are defined with characteristics, and appear in layers. The first layer is defined as the input layer, and the last layer being the output. Layers between the two aforementioned are described as hidden layers. Data is analysed through their propagation through the layers of neurons, where each neuron influences the content of the data as they pass through. Learning takes place in the form of the manipulation of the weights connecting the units in the layers. This allows it to model complex relationships between inputs and output, and it can also find patterns in the data.

C *Genetic Algorithms*

As subset of evolutionary algorithms in general, Genetic algorithms (GAs) are a group of search techniques used to find exact or approximate solutions to optimisation and search problems. The methodology is inspired by Charles Darwin's evolutionism theory; individuals are created by the crossover of the genetic information (genome) of their parents. Genomes are metaphors of genetic information; it typically refers to a data structure, most commonly an 1D array. Genetic Algorithms generally consist of a population of agents, a tournament selection

process, algorithmic crossover mechanisms against the genomes of the agents, and the introduction of probabilistic mutation on genomes. The combination of evolutionary algorithms to train neural networks is described as a type of Neuroevolution.

D Aims

Whilst the use of evolutionary algorithms and neural networks have been explored to create draughts players, my intention is to explore the effectiveness of genetic algorithms and neural networks. The intention is to determine the possibility of developing a performing draughts playing agent by the use of Genetic Algorithms and Neural Networks.

D.1 Minimum Objective

The minimum objective is to implement a feed-forward neural network, alongside a Draughts game interface for the network to evaluate on. This will need to be paired with some decision making program that will determine the move to choose at a given state of the game. Finally, a simple genetic algorithm would also need to be implemented that interacts with the neural network.

D.2 Intermediate Objective

The intermediate objective involves the implementation of an interface to play against the end result, and to create a monte-carlo decision making algorithm, which will be used by the agents to evaluate better quality moves. Also, as the genetic algorithm includes a tournament mechanism which will play many games simultaneously, this will need to be implemented to run in parallel.

D.3 Advanced Objective

The advanced objective consists of producing a system that indicates that it is learning over time. The later sections of this paper evaluates the quality of the learning rate and other results.

To template the document, the later sections of the paper will discuss related and relevant work, the proposed solution, results and an evaluation of the project. The paper is then wrapped up with a conclusionary statement, discussing the implications of the project and potential future work.

II RELATED WORK

Historically, Draughts has been used as a testing ground for artificial intelligence and computational performance since the early introduction of computers. This is due to the relatively vast potential move space, making it difficult to brute-force, until recently. Schaeffer proposed a method to solve the game (Schaeffer & Lake 1996), and eventually followed on to prove that the game can be weakly-solved a decade afterwards. (Schaeffer et al. 2007). Weakly-solvable is defined such that 'the game-theoretic value for the initial position has been determined, and that a strategy exists for achieving the game-theoretic value from the opening position.' (Allis 1994) However, the premise of this paper is not necessarily to solve the game itself, but to rather see whether genetic algorithms are still relevant.

Research in artificial intelligence alone has been active since, targeting games with a higher move space, notably Chess and, more recently, Go. This section will summarise specific and notable techniques relevant to the discussion of the task at hand.

A *Arthur Samuel*

In 1959, An early design of machine learning to play Draughts was devised by Arthur Samuel (Samuel 1959). His algorithm is based on the perpetual improvement of heuristics he devised, which act as an evaluator for a given state of the checkerboard. The evaluation function is calculated using the linear combination of his heuristics, which included various parameters of relevancy of the game. This includes the number of pawns, number of pawns positioned along the central diagonal, and so on.

His evolutionary strategy consisted of agents that have different coefficients for the heuristics, and new agents are formed using a successor's coefficients with subtle mutations.

Samuel also described an early concept of Mini-Max, with Alpha Beta Pruning. The weights were then trained by the algorithm playing itself in a form of genetic programming. A later investigation by a checkers magazine evaluated Samuel's player to below Class B (Schaeffer 1997, Fogel 2000) where Class B being categorised with an ELO range of [1600-1799].

However, the work described was naturally also dependent on a set of heuristics Samuel devised. This could handicap the agent's ability to play, as they are constrained on the quality of Samuel's heuristics.

B *Blondie24*

The idea of evolving neural networks to play the game of Draughts is based on the success Chellapilla and Fogel had in evolving their own draughts playing neural networks. Their method used Samuel's system as a foundation and modified the evaluation method, using feed-forward neural networks instead of a linear combination. Their resulting player, Blondie24, was then taught to play without priori knowledge, in a similar vein, but with a higher level of competency. (Chellapilla & Fogel 1999)

Blondie24 uses an evolutionary algorithm, using fifteen agents $i = 1, \dots, 15$, each of which consisted of two properties. The first property were weights of the neural network (which are initially random), with the second property being a self-adaptive parameter vector, τ_i , which determined how random the mutations are when the agent i evolved.

In each generation, every agent plays each other in a round robin style tournament such that every agent plays an equal amount of games, and successors are chosen based on their performance from it. The best agents are chosen to mutate themselves to create new agents. Mutation is performed using the formula:

$$\tau_i(j) = \tau_i(j) \cdot e^\mu$$

Where $j = 1, \dots, N_w$ represented each weight of the neural network w , and μ representing a random decimal in the range of $[0 - 1]$.

However, crossover mechanisms were not considered (and is thus not classified as using genetic algorithms). Blondie24's neural network structure consists of a $\{32, 40, 10, 1\}$ set, such that the input layer consisted of 32 nodes, and a single output node. The output node consists of a classifier value that has a range of $[-1, 1]$, representing a particular advantage for a given side of

the board. One fallback that can be seen with this particular method is that spatial awareness is not considered as it takes an immediate input of the positions on the board. This makes it inherently more difficult for the neural network to generate heuristics based on spatial awareness. Relative distance between pieces are not encapsulated in a method that can be immediately interpreted by the neural network.

Blondie24’s decision making algorithm primarily revolved around a mini-max algorithm with a ply cap. This algorithm, with the state of the checkerboard as an input, expands a tree of potential moves. The mini-max algorithm is described using the following instructions:

1. Each node represents a state of a checkerboard. Child nodes represent a state of the checkerboard for a given move. Child nodes can only be produced from the parent node; i.e. a move from a parent node will produce a child node.
2. Child nodes are expanded until some depth is reached. From here, an evaluation function is used to calculate the value of the leaf nodes at this depth. This evaluation function produces a value that indicates a players advantage in a given state.
3. The best potential position for a given player can be deduced from the nodes at this stage, which is quantified with a value.
4. The best value propagates upwards until it reaches to a child of the root node. This represents an agent’s current state of the game.

The use of alpha-beta pruning would help to ignore evaluating unnecessary moves, reducing the number of calculations. Once the possible moves have been considered, the best child node is chosen, representing the best move at a given position.

There are inherent trade-offs to this decision algorithm, especially with requiring higher ply counts; asymptotic complexity of mini-max is exponential; it being $O(x^y)$ for x being the branching factor, and y being depth. Branching factor will consist of the moves from each agent in a given game. This exponential growth is the achilles heel to the mini-max approach.

We do not wish to repeat this work, but it is important to understand that we wish to abstract and modularise parts of the best features of this system to modify and extend them to the solution described in the paper.

C Post Blondie24

Blondie24 was influential in the field of neuroevolution during its time, leading to various studies on the theory behind its success. A notorious example is the series of publications by Al-Khateeb and Kendall, where they explore in detail the different features that comprise Blondie24. One of their studies concerning an extensive look into the Look-ahead Depth (or ply) concluded that having a higher ply-depth unsurprisingly increases the overall performance of the player, and that agents trained at a higher ply performs significantly worse when playing at a lower ply. More importantly, the results in the paper suggest that a 4ply depth produces the best value function to start with during the learning phase, and increasing the ply depth for evaluations. (Al-Khateeb & Kendall 2012)

Their study in the influence of tournaments in Blondie24 concluded that the league structure seemed more appropriate to use as opposed to the round robin tournament. (Al-Khateeb & Kendall 2009)

Within research in the game of Draughts alone, Toledo-Marin et al. performed studies on attacking strategies using the game as a reference (which however can be applied to any other zero-sum game), where they mathematically prove that for an offensive player, maximising the offensive improves their probability to win. (Toledo-Marin et al. 2016) This could be considered to guide the system to lean towards attacking moves, which could potentially accelerate the learning rate of the system.

D Return of Genetic Algorithms

Since then, there has been a non-trivial number of published papers describing the promising utility of genetic algorithms and neural networks in general. Genetic algorithms have been shown to evolve weights for a convolutional neural network in a classification task better than back-propagation, in nearly all situations, other than when the number of generations generated for neural networks are small. (Perez n.d.) Another research showed its ability to compute optimal structures for deep neural networks automatically, using an image recognition dataset as the benchmark for competitiveness against the networks. (Xie & Yuille 2017)

However, one recent paper by Clune et al. on behalf of Uber Labs returned to the use of genetic algorithms as a basis to train especially deep convolutional networks with success. (Such et al. 2017) Their neuroevolutionary system was used to play popular Atari games on the OpenAI framework, and was found to perform as well as other contemporary evolution strategies and reinforcement learning algorithms.

The success of the program was based on their system’s ability to perform safe mutations, a technique that revolves around measuring the sensitivity of the network to changes with some of the weights. (Lehman et al. 2017) This afforded the system to evolve deep neural networks (In this particular case, containing over 100 layers and over four million weights and biases,) in a manner that allowed random but safe and exploratory s.pdf. Their safe mutation exploits the knowledge of the neural network structure, by retaining sampled experiences of the network (inputs that feed into the network and its outputs). This allows the weights to be manipulated by comparing its significance with the output, by comparing it with the sampled experiences.

However, they also do not consider the use of crossover techniques in their paper. It is discussed that it would very well be possible to produce a safe crossover method that can exploit structures of a neural network, inspired by their research produced on safe mutations.

Unfortunately, it is unrealistic to apply the same techniques described in their paper into the proposed system. The machine used as the base (consisting of hundreds to thousands of cores) is of a magnitude greater than what is available for the task at hand, but a more simplistic and adapted model can be produced using similar ideas.

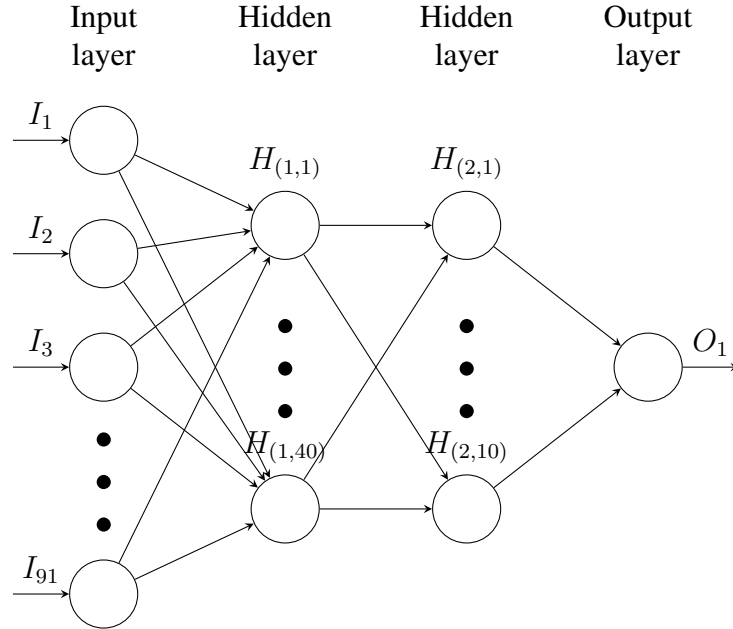
A popular trend between the aforementioned papers is that there is a trend of absence of the use of crossover algorithms. One paper describes the use of different crossover algorithms on neural networks having no clear advantage over each other, in comparison to a uniform crossover (uniform crossovers are also commonly defined as the textbook crossover) which disrupted the overall performance of the neural networks. This could be due to building blocks being disrupted as chunks of the sequential weights are swapped out of each other. (Emmanouilidis & Hunter 2000) It is important to consider this when devising a crossover method, such that the effect it has on the neural network should be subtle enough that it should pose similar results to the network prior to crossovers.

III SOLUTION

This section describes the build of the system in order of implementation. Implementing the checkerboard interface is not included in the discussion as it is not one of our objectives. However, the sections below are considered under the premise that the interface is built prior. The system is implemented using Python 3.6 and NumPy. Other machine learning libraries are not used due to the nature of the system.

A Evaluation Function

Figure 1: The chosen neural network model. The preprocessed values of the checkerboard are used as input. An output is produced after propagation that ranges from $[-1,1]$.



A feed-forward multilayer perceptron neural network was used to evaluate the board. The network contained 4 layers; the input layer consists of 91 nodes, with the output node having 1. Hidden layers had 40 and 10 nodes respectively. Figure 1 shows the network in the form of a diagram.

The inputs of the network is a preprocessed 1D array of the checkerboard, which is transformed into the set of the sums of the subsquares that comprise the positions of the pieces of the checkerboard.

The preprocessed input starts with retrieving the positions of the pieces on the checkerboard in the form of an 1D array, and calculating all possible individual subsquares of the board. The subsquares kernel size range from a 3x3 to an 8x8 size.

With the kernels, the values of the pieces in the subsquare is summed up. A new array is formed, comprising of the sums of the different kernels that comprise the board. This new array (comprised of 91 entries) represents the preprocessed input, which is fed into the neural network.

The intention is to weigh the Black pawns with a value of $v = 1$, and white pawns as $-v$. It is commonly understood that a king is worth more than a pawn piece, but it is disputed about

its precise value advantage. This could be decided by the system such that it can evolve to understand that a King is eventually worth more than the pawn but for the sake of simplicity, a king's piece value is to be weighted at $1.5v$.

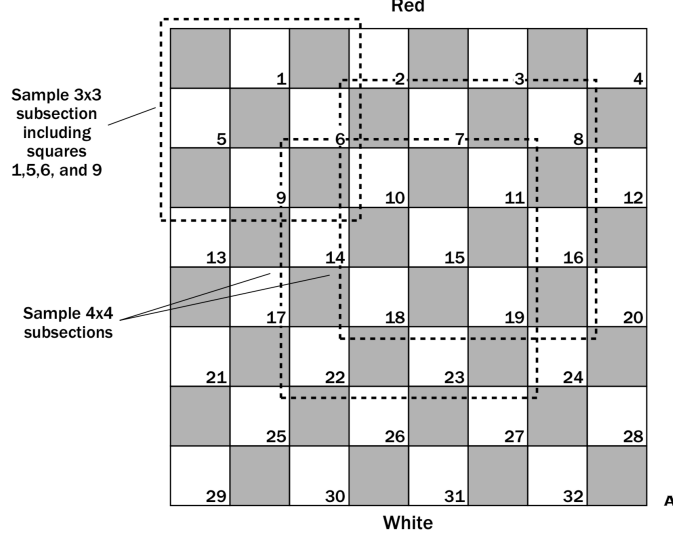


Figure 2: A diagram visualising the processing of subsquares.

Weights w of a neuron are multiplied by their input value I , summed with their bias B , and are passed through an activation function $O = f((I * w) + B)$ to become an input for another neuron. Activation functions, when visualised, usually have a sigmoid curve, but may also take the form of other shapes. Common characteristics of activation functions include values to be monotonically increasing, continuous, differentiable and bounded.

Our choice for the activation function is tanh (a type of sigmoidal function), shown in figure 3. The use of tanh facilitates the ability to simulate a zero sum formula (for a range of -1 to 1) due to its symmetric properties around the origin.

Results are cached such that inputs that have been already evaluated are saved such that they can be recalled immediately. This has shown varying levels of improvement dependent on the ply depth. As the ply increases, the size of the state space increases, and therefore the chances of recalling the same set of inputs naturally decrease. The cache is also used for our safe mutations algorithm, described later on in C.3.

B Decision Making Algorithm

The decision making algorithm of choice is a modified Monte-Carlo Tree Search (MCTS). A basic MCTS differs from mini-max where future moves are randomly played to the end of the game. It acts as a sampling of possible outcomes, and does not depend on an evaluation function at all. The ran-

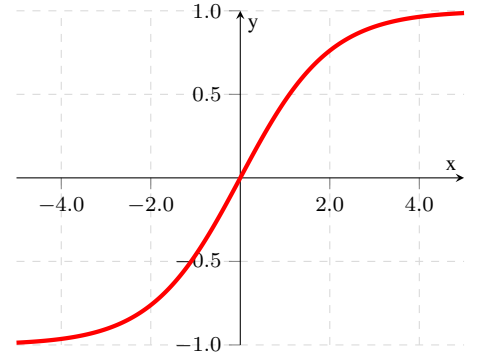


Figure 3: Graph of tanh function $f(x) = \frac{2}{1+e^{-x}} - 1$ (An example of an activation function with a sigmoid curve).

dom simulation of games are skewed such that more reasonable moves are chosen. A survey by Browne et al. found that MCTS can dramatically outperform mini-max based search engines in games where evaluation functions are otherwise difficult to obtain (Browne et al. 2012). MCTS are played in rounds that consist of four operations:

- Selection: A selection strategy is played recursively until some position is reached.
- Play-Out: A simulated game is played. Again, a basic form of MCTS would play until an end-game is reached.
- Expansion: A node is added to the tree.
- Back-propagation: The result of the game is back-propagated in the tree of moves.

These rounds are repeated until a certain period of time is met or a certain number of rounds are computed. Once finished, a node is chosen based on its total probability of reaching an winning outcome.

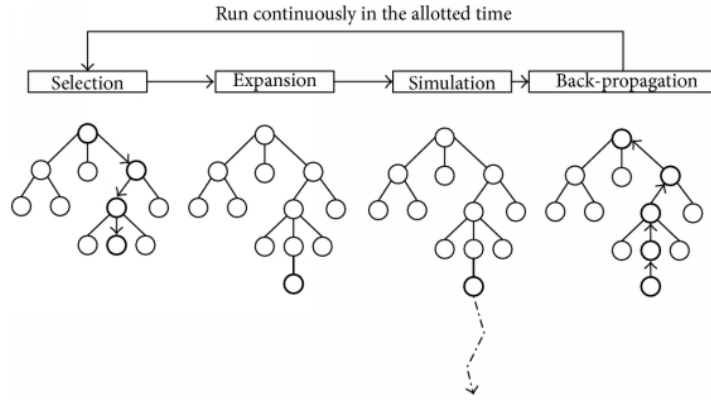


Figure 4: A diagram describing the MCTS process.

MCTS-EPT (or MCTS with early playout termination) is introduced by Lorentz (Lorentz 2016). MCTS-EPT modifies the MCTS random play-out approach; Instead of allowing the random moves play to the end of the game, the number of moves traversed are capped and an evaluation function (which in our case is the neural network) is applied at that future state instead. The termination ply would be hard capped at 6ply, but the system has been adapted such that if there are a limited number of moves the ply depth is extended by 1ply. This reduces the need to depend on random playouts of moves to the end, and ensures that the quality of the moves depends on the evaluation function. For our results, we use the ply depths 1,3 and 6 to determine the system is consistent. As opposed to using a time threshold, the number of MCTS rounds r is based on the ply depth p , such that $r = 300p$. This allows the system to process moves in a manner that the performance of the system is hardware-independent, as faster machines could process more rounds in a fixed set of time.

C Genetic Algorithms

The genetic algorithm (GA) is the premise of the system's learning strategy. GA's are used to train the neural network. We discuss the various algorithms that form the collection of GA strategies below. For the system, a population size of 15 will be chosen.

C.1 Population Generation

In genetic algorithms, the population serves as a base that allows agents in the pool to play each other. Every generation has its own population, with all having the same number of agents. The initial population will consist of randomly generated weights and biases of the neural network, with values from $[-1,1]$.

For a population size of 15, the next generation is created using the best five agents from the current generation. This is discussed in *C.2 Tournament Selection*. They will continue to play in the next generation; this strategy is described as elitism. These agents are also used as a base to create 10 new agents from.

The next eight players are generated through the use of crossover strategies, explained in *C.4 Crossover Strategy*. The weights of the 1st and 2nd place agents are used as input to the crossover strategy and will generate 4 offsprings. Two are reciprocal crossover representations from each other, and the other two being directly mutated from the parents themselves. Another four children will be created using the same strategy, with the 2nd and 3rd agent's weights. The remaining two will be direct mutations of the 4th and 5th place agents.

C.2 Tournament Selection

To sort the quality of the players in a population, a tournament selection process is deduced. This allows us to choose the best players who will continue to play in the next generation.

Currently, each agent in the population would play 5 games as Black, against randomly selected opponents. Each game lasts a maximum of 100 moves from both players. If a winner is not deduced at this stage, a draw is called. Draws are also called when there is a three-fold move repetition from both players. A win is worth 2 points, a draw being none and a loss being -1 points. Both the agent and its opponent receives a score from the game. Scores are tallied up at the end of the tournament. Players are sorted by the number of points they scored. The best players would have the highest number of points.

Talk about how this is running in parallel thanks to python's multiprocessing library'

C.3 Coefficient Mutation

To create variation between agents in preparation for the next generation, we create statistical anomalies through the use of mutations. This is used as one of the learning mechanisms that help change the decision factors of the neural network.

Weight and biases of an agent's neural network would increment by a random value that is created using the following formula, where w_p is the current weight, K represents the number of weights and biases in the neural network, and m representing a random floating point in the range of $[-1,1]$. Equation 1 describes the mutation formula.

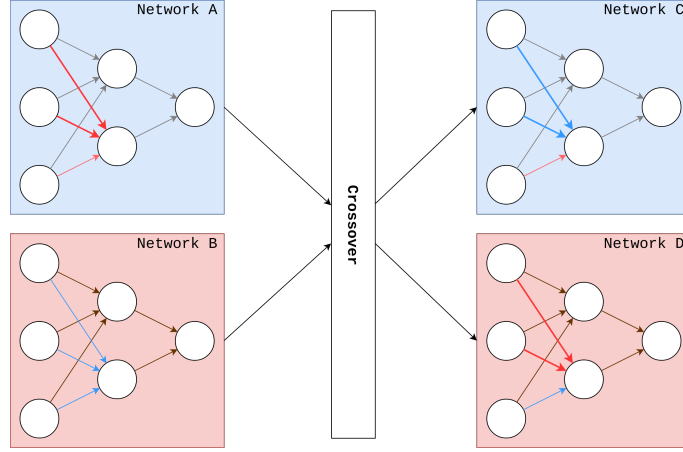


Figure 5: A diagram visualising the crossover process.

$$w_n = w_p + \frac{m}{\sqrt{2 * \sqrt{K}}} \quad (1)$$

Like the activation function, the weights will have a soft maximum of $[-1, 1]$. This would consequently mean that the mutation is not controlled, and dependent on the number of weights in the system; The more weights in the network implies a less significant mutation.

We also choose to experiment with soft mutations, inspired by Lehman et al. (Lehman et al. 2017) We use a subset of precalculated neural network calculations for inputs and outputs and use this as a premise to guide our mutation. The method is as follows:

1. With the current set of weights w , choose a subset of precalculated input and output tuples $\phi_w = (I, \lambda_{I,w})$ where I represents the input values of the network, and $\lambda_{I,w}$ as the output.
2. Generate a perbutation of the weights calculated using Equation 1. Call this y . Use y as a basis of a neural network.
3. Using the inputs I in ϕ_w , calculate a new set of input and output tuples $\phi_y = (I, \lambda_{I,y})$.
4. Calculate the divergence $\delta = \lambda_{I,y} / \lambda_{I,w}$.
5. Calculate the quality of y based on the number of inequalities $\delta \geq 1$.
6. Repeat s.pdf 2-5 some n times, keeping note of the the best y .

The intuition behind this is that if ϕ_w is a subset of moves that led to a winning game, then the manipulation of the weights is left open as long as it can replicate the performance of winning games. One of the concerns about this approach, also mentioned during the similar works, is the efficiency of the approach.

C.4 Crossover Strategy

Another learning mechanism provided from the use of genetic algorithms is the use of crossovers. This combines the traits that build two parent agents to create children. In our scenario we use the weights and biases of the parent's neural networks.

Two offsprings would be created from a pair of parents, with each offspring being the reciprocal crossover of each other. A random node from the dominant parent's hidden nodes are chosen. A dominant parent is the agent who gained more points in the tournament round than the other (who would be the submissive parent). If a dominant parent cannot be deduced it would be randomly chosen.

Once a random node is chosen, The weights and biases that feed into the chosen node are swapped with the submissive parents values at the same positions. As values in the weights can range from -1 to 1, dominant input weights are values that are greater than 0.75 or less than 0.75. Figure 5 shows this in more detail. In the diagram, Network A and B are parents, creating reciprocal offspring networks C and D. A is the dominant network, and A's node with the red inputs is chosen. The weights and biases of the red inputs are swapped with the values at the same position in network B. Network C is formed with Network B's dominant weights, with the rest being composed from A. Network D is the reciprocal, having Network A's dominant weights and the rest from B.

This crossover process would create a subtle modification to a neural network, by swapping a small subset of weights and biases at a time. Having a more dramatic crossover could potentially change the structure of the network flow, reducing its effectiveness. A visualisation of the crossover process can be found in Figure 5.

D Testing

At the end of a given generation, we measure growth of performance by using the generation's champion. Presently we will use the mean of means approach. When a new champion is generated, it is played against the previous 5 champions from earlier generations. 6 games are played for each previous champion, with 3 being as Black, and 3 being White. A mean score is calculated from those 6 games. The overall performance of the current champion is the mean of the 5 sets of games. A positive improvement is when the mean of means are greater than 0.

Point Score for the champion games are measured by 1,0,-1 where a Win counts as 1 point and -1 for a loss. The weights are scaled differently to the regular tournament in order to accurately portray the difference between previous champions.

Evaluation Method

The end player is used to create comparison games against its ancestors. This player is to be first tested against an agent who is choosing random moves, an agent who is using pure monte-carlo tree searches, and a subset of older agents prior to its current state. These statistics will be used as evidence in order to determine whether the agent is learning over time, is used as the premise to determine whether neuroevolution can be considered as a technique for non-specific domain problems.

E Setup and Issues

Initial runs will operate on a 1-ply load to determine the stability of the system on a linux (Ubuntu) machine containing a 4-core Intel i5 6600u processor with 12GB's of memory. Development and debugging will also occur on this machine.

Once testing has proven to be stable, the system would run on Durham's MIRA distributed system (Debian) with a 6-ply move depth. This machine utilises 4 Intel Xeon E7-8860 processors (Each CPU contains 16 physical cores running in 2.2Ghz, with hyperthreading). This comes to a total of 64 Cores, and 128 threads. MIRA also comes with 256GB of DDR3 memory.

To keep simulations running on MIRA, MOSH is used to maintain a consistent connection to MIRA. The end champion is then transferred to the initial machine in order to be played against by human input. Statistical evaluations and calculations are also calculated on MIRA to reduce computational time.

- finding a computer that would run the simulations - appropriately saving files - testing neural network was difficult - testing move decisions were difficult - optimising the learning rate

Verifying that the moves are properly chosen was difficult and slow as games need to be played out first, with the quality assurance handled afterwards. The system was implemented first with the use of minimax search, but this cannot be used for the final measurements as the scalability of the decision making algorithm scales exponentially, as described prior.

Interestingly, the machine used for training did not perform necessarily as efficiently. Although the system is built such that the games in the tournaments are "embarrassingly parallel", the machine does not scale accordingly. Most of the computation is spent on calculating floating point matrix operations (Which represents the neural network evaluations.) Later research has shown that the use of hyperthreading does not necessarily show to scale the performance, even though more threads exist. (Leng et al. 2002) This is due to the shared floating point arithmetic registers in the CPU, where two threads utilise one physical cores property. The use of GPU acceleration is not considered due to the lack of access.

IV RESULTS

For the experiment, we create three different neuroevolutionary agents with separate ply-depth: 1,3 and 6. Ply Depths 1 and 3 are run for 100 generations, and 6 is run for 200 generations to measure the overall potential of the system. For sections containing three charts, each chart represents the results of the agents playing at 1-ply, 3-ply, and 6-ply, respectively, in that order. Each simulation utilises a population of 15 agents.

A Learning Rate

The learning rate is derived by comparing a generation's champions with its predecessor champions from the previous five generations. For each older champion, six games are played, with the outcome of the current champion's performance scored and tallied up. An average of the six outcomes are calculated, representing the learning rate. We do not take the learning rate of the first 2 champions as this would not produce a sufficient comparison. This measurement is performed for every generation until the end of training. It is also the case that this does not influence the actual learning process of the system, but is used as one of the indicators of whether the system is learning in ability over time.

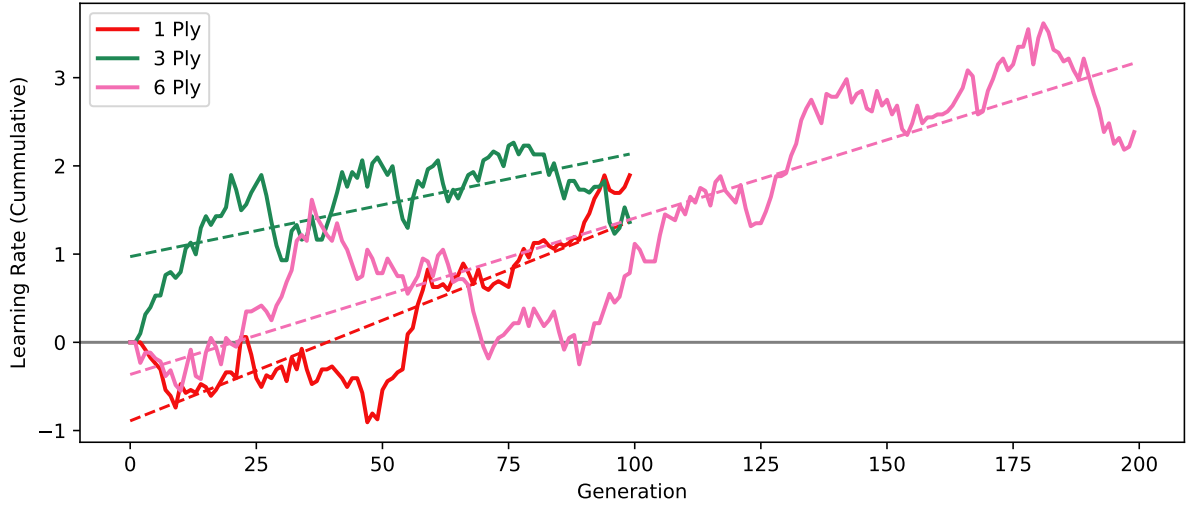


Figure 6: A chart showing the cummulative learning rate across the generations for the different ply depths.

For all ply depths, we can see a cummulative growth such that returns are in the net positive as more generations are added. However, even with safe mutations, we can also see that the learning is very volatile. As the champion is chosen from a tournament, it may be related to the quality of the following generations having a lower ability to evaluate the board than it's parent generation. This could potentially be prevented by reusing the previous champion and restarting a particular generation if a loss can be seen. This could however tunnel the system towards a local maximum in terms of learning. Further measurements are taken to understand the reasoning of the potential troughs in learning.

B Performance

To measure the system performance at the end of training, the champion of the final generation is compared against a subset of their predecessors. For each precessor, the trained system plays 128 games (half as black and the other for white), and the number of wins, losses and draws are counted to produce the following statistics. For debugging purposes, all agents are also played against a random program to determine whether the system is evaluating moves properly. Fortunately, all agents successfully won every game.

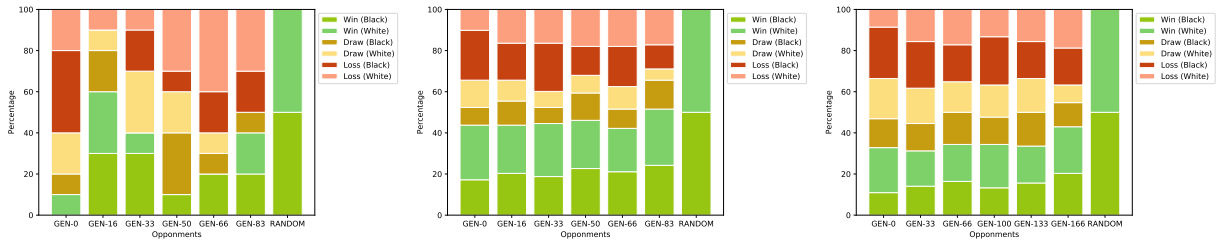


Figure 7: Charts showing the performance of the trained systems against their ancestral counterparts.

For the ply depth of 1, we can see an immediate net trend, with the end system beating earlier generations except for the first generation. This may make sense, as heuristic retention would be lost as the search depth is so low. This does however suggest that the system is worse off than playing with an initialised system (without training). For the higher ply depths (3,6) however, the agents manage to consistently maintain a higher chance of winning/drawing against their earlier counterparts. It is difficult to see a trend line. For the 6-ply, it is evident that there is an increase of draws caused against the earlier agents, which could be seen as an advantage.

C Champion Distribution

The following charts shows the following distribution charts the champions of the generations. The intention behind this measurement is to determine whether champions are inherently chosen based on the influence of mutations, crossovers or simply elitism. Measurements are taken such that the champion of each generation is looked at, and their genomic properties are collated to show a distribution of the chances of a particular agent being the champion. Note that while every offspring is mutated, some of the offsprings are created using genomic crossovers, described in section C.4.

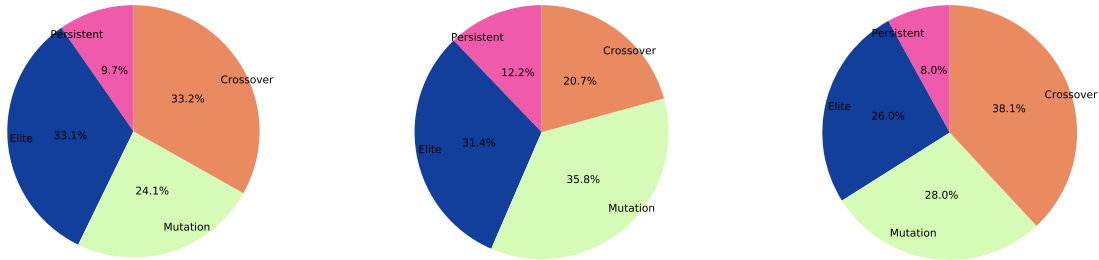


Figure 8: Distributions of the different genomic identities across the tournament champions.

The influence of offsprings become more apparent for larger plys, with mutation and crossover appearing more frequently, coming at 66.1% of the distribution. Interestingly, for ply depths 1 and 6, crossovers tend to be the most frequent occurrence when it comes to producing champions at every generation. This suggests that crossover methods tend to produce better agents than their counterparts.

D Influence of Inheritance Methods

Building on from the chart provided by Figure 8, we look into the scores retrieved by the different genomic identities when they are compared against their earlier counterparts. This is to determine whether the use of genetic algorithms can overall improve the quality of neural networks. To measure this, we correlate the generation's champion with their score (the same measurement used to evaluate the learning rate in Figure 6) produced when measured against their predecessors. The results are shown in Figure 9.

Interestingly, for the 3-ply, the crossover method is most likely to produce a growing system, producing the highest mean. However, across all of the ply depth agents, we can see that the use of mutation produces the best possible agents. For the higher 6-ply depth we can see that it did not significantly influence the learning rate of the system.

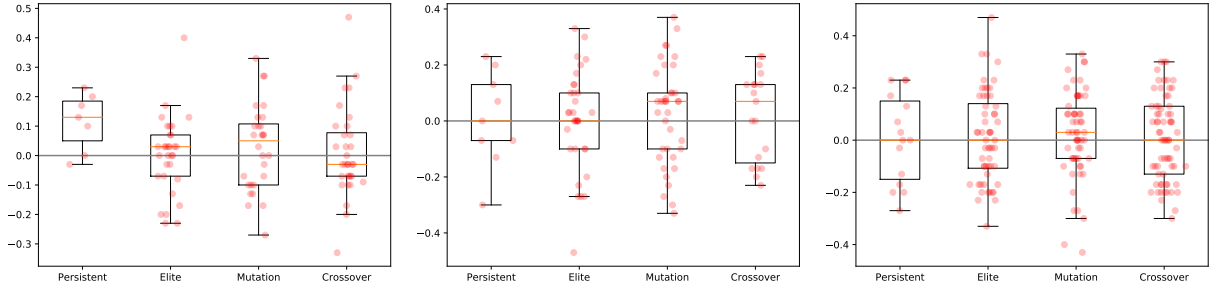


Figure 9: Box-Plots of the different genomic identities and their distribution of scores. The yellow line in the box represents the mean score.

E Other Observations

E.1 Number of Moves

The number of moves is measured to verify that the system is producing realistic and quality moves over the training phase. The following charts show the distribution of moves for the games played over the generations, alongside the mean. The results are shown in Figure 10:

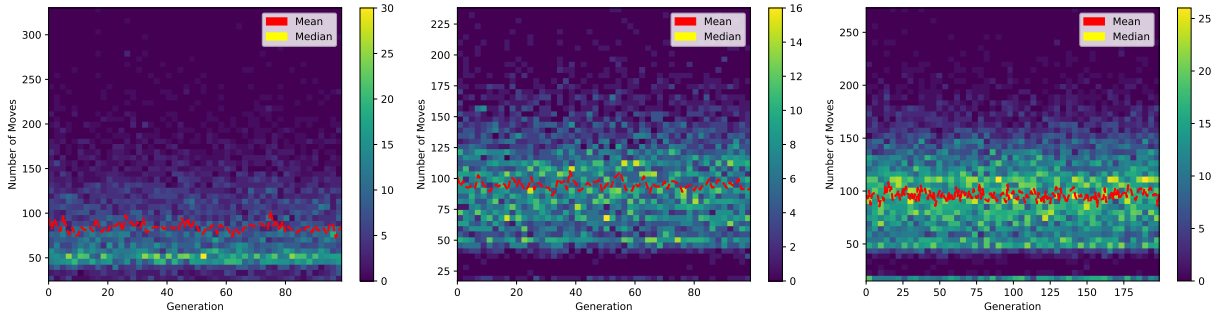


Figure 10: 2D Histograms representing the number of moves played in the games played in the generations.

For the 1-ply simulations, the average mean of moves is lower, but is counteracted by the much larger range of games that have a large number of moves. Understandably, the lower quartile for the 1-ply is lower than the others. This may be due to the level of precision, where it does not think further ahead enough to make better moves. This explanation applies inversely for the larger ply-depths, notably the 6-ply chart, where the level of precision is higher, leading to a tighter distribution of games having similar numbers of moves played.

As the system is shown to play fairly consistently across the generations, this verifies that the system is choosing consistent decisions across the board. This confirms that the decision making algorithm is implemented precisely, reducing the number of unknowns for evaluation.

E.2 CPU Times

The system is shown to grow in running time linearly, as this is the cause of growing the search limit for MCTS in a linear fashion according to the ply-depth.

Table 1: Mean and Net running times of system training for the various depths.

	1-Ply	3-Ply	6-Ply
Mean	0:04:24	0:17:32	0:38:45
Net	7:21:25	1 day, 5:14:41	5 days, 9:12:25

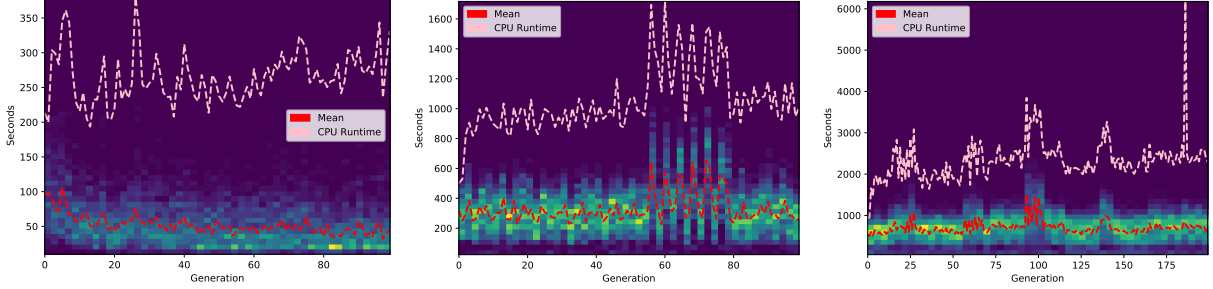


Figure 11: 2D Histograms of the average times spent on the games during the simulations.

The results suggest that the CPU time is very unstable, and is one of the downfalls of using a time-shared machine. The performance increase does not show to correlate with the learning rate.

V EVALUATION

We now reflect back to the original research question: "is it possible to create a performing draughts playing agent by the use of Genetic Algorithms and Neural Networks?". We answer this through the examinations of the strengths and limitations of the system.

A Strengths

The statistics indicate that the system is indeed learning over time. This suggests that neuroevolution is a viable method for programs with vast search spaces. While this is where both genetic algorithm can shine. Although this is the case for Deep learning and evolutionary computation in general, the benefit of using genetic algorithms is that the fitness function can be any function, whereas for a neural network, its function needs to be translatable to some form of a differential sum of partial derivatives.

The system itself relies on a relatively simple learning heuristic that is open to domain specific adaptation. Because the agents are essentially playing itself to learn over time, no training data is necessary for the program to evolve itself.

embarrassingly parallel tournaments make it possible to reduce the overall system running time on the system.

The use of crossovers, which differentiates genetic algorithms and its other evolutionary algorithms has shown to improve the quality of the players, especially the case in the 3-ply.

B Limitations

The overall training rate is slow. Naturally slower learning rate than typical ventures as there is an immediate comparison to make against, and requires many more samples (often exponentially more) to get a weight of equivalent quality. The quality of the neuroevolutionary algorithm boils down to the quality of the fitness function, which in our case is the tournament method. Round-Robin is not necessarily the most efficient (from a number of games played perspective, and not necessarily parallelism) and does not eliminate cases where there is an ambiguous decision between the champions (which could be due to several players having an equal number of points). It also computes the most number of games. - creating a high quality fitness function is difficult and its domain specific, or rather, there is no universal fitness function for all problems. This poses as a benefit and a weakness for the neuroevolutionary approach. - the fitness function defines the size of the search space and how easily a globally optimal solution can be derived. - using a best of champions approach is quite difficult! - reality rarely fits into good fitness functions The crossover mechanism is a hit and miss, and is shown to impede the learning rate for some of the longer ply-depths. This may be due to its disruptive nature, as heuristics are destroyed. As feared by findings by (Emmanouilidis & Hunter 2000), we can see the potential fears of the negative influence of the textbook crossover mechanism, where it has shown in the results to discourage learning.

C Approach

The modularity of the system implementation was pivotal to debugging the system. The program itself is quite sizable, and it helps that individual tests can be conducted in most cases. It also helps that the order of the implementation helps.

Measuring the learning rate is quite difficult and arbitrary, and may best be compared to against systems that are classified to play at a particular level.

As most of the experiments are constrained by computational power and time, it may be best to re-evaluate this concept until either computational resources are more available, such as the use of GPU acceleration. The system should also be left to run for longer generations as presently the results do not show a notably strong correlation. The lack of computational resources also influenced the number of experimental testings that can be measured, such as the population size, the number of generations and the learning rate.

Another consideration for the project is to experiment with different crossover algorithms that utilise the temporal difference between inputs and outputs. There is an incredibly vast scope of future work that would extend the project. One interesting avenue is to consider using a safer method to produce mutation, exercising other contemporary methods in machine learning. It may also be the case that more time could be spent in examining references outside of computer science. In this particular example, the algorithms used are inspired by various biological and neuroscientific understandings. More research into this field may assist in finding more creative methods of solving some of the issues concerning genetic crossover mechanisms and mutation rates.

VI CONCLUSIONS

In terms of fulfilling its objectives, the project was an overall success. A system was implemented using evolutionary methods and it successfully plays checkers in a manner in which it can learn from over time. To the extent of its growth remains suspect however. Efforts to measure the learning rate is one of the non-trivial challenges that makes neuroevolution a particular field of interest, as it is inherent that it can tackle problems in a somewhat novel manner.

The main findings of this project is as follows:

1. It is evidently possible to create a checkers playing agent that learns to play itself, with little human intervention.
2. neuroevolution is relatively inefficient compared to their gradient based counterparts. This however foreseen due to the heavy dependency of entropy disguised as learning.
3. The fitness function can be anything, but it is important to derive a high quality fitness function.
4. The quality of crossovers can be a deciding factor in how the agents learn to play over time.

The major implication is that it is a possible contender for tasks that require unsupervised learning, and that the fitness function cannot be trivially decomposed into set of derivatives that allow regular derivative based learning to shine.

Provided a high quality fitness function can be designed, the applications of neuroevolution expand beyond its applications in zero-sum games. Neuroevolution in general is useful based on its ability to solve problems without the use of training data, and generating evolving content over time. When applied in a successful manner, its applications can go towards the understanding of biological neural networks and the evolution of intelligence itself.

References

- Al-Khateeb, B. & Kendall, G. (2009), 'Introducing a round robin tournament into blondie24', *CIG2009 - 2009 IEEE Symposium on Computational Intelligence and Games* **44**(0), 112–116.
- Al-Khateeb, B. & Kendall, G. (2012), 'Effect of look-ahead depth in evolutionary checkers', *Journal of Computer Science and Technology* **27**(5), 996–1006.
- Allis, V. (1994), Searching for solutions in games and artificial intelligence, PhD thesis, University of Limburg, S.I. OCLC: 905509528.
- Browne, C. B., Powley, E., Whitehouse, D., Lucas, S. M., Cowling, P. I., Rohlfshagen, P., Tavener, S., Perez, D., Samothrakis, S. & Colton, S. (2012), 'A Survey of Monte Carlo Tree Search Methods', *IEEE Transactions on Computational Intelligence and AI in Games* **4**(1), 1–43.
- Chellapilla, K. & Fogel, D. (1999), 'Evolving Neural Networks to Play Checkers without Expert Knowledge', *IEEE Transactions on Neural Networks* **10**(6), 1382–1391.
URL: <http://ieeexplore.ieee.org/abstract/document/809083>

- Emmanouilidis, C. & Hunter, A. (2000), A Comparison of Crossover Operators in Neural Network Feature Selection with Multiobjective Evolutionary Algorithms, in 'GECCO-2000 Workshop on Evolutionary Computation in the Development of Artificial Neural Networks, Las Vegas'.
URL: <http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.152.7867>
- Fogel, D. (2000), 'Evolving a Checkers Player Without Relying on Human Experience', *Intelligence* **11**(2), 20–27.
URL: <http://doi.acm.org/10.1145/337897.337996>
- Lehman, J., Chen, J., Clune, J. & Stanley, K. (2017), 'Safe Mutations for Deep and Recurrent Neural Networks through Output Gradients', *arXiv:1712.06563 [cs]* . arXiv: 1712.06563.
URL: <http://arxiv.org/abs/1712.06563>
- Leng, T., Ali, R., Hsieh, J., Mashayekhi, V. & Rooholamini, R. (2002), 'An empirical study of hyper-threading in high performance computing clusters', *Linux HPC Revolution* **45**.
- Lorentz, R. (2016), 'Using evaluation functions in Monte-Carlo Tree Search', *Theoretical Computer Science* **644**, 106–113.
URL: <http://linkinghub.elsevier.com/retrieve/pii/S0304397516302717>
- Perez, S. (n.d.), 'Apply genetic algorithm to the learning phase of a neural network'.
- Samuel, A. (1959), 'Some studies in machine learning using the game of checkers', *IBM Journal of Research and Development* **3**, 210–229.
URL: <https://ieeexplore.ieee.org/document/5392560/>
- Schaeffer, J. (1997), *One Jump Ahead*, Springer-Verlag.
- Schaeffer, J., Burch, N., Bjornsson, Y., Kishimoto, A., Muller, M., Lake, R., Lu, P. & Sutphen, S. (2007), 'Checkers Is Solved', *Science* **317**(5844), 1518–1522.
URL: <http://science.sciencemag.org.ezphost.dur.ac.uk/content/317/5844/1518>
- Schaeffer, J. & Lake, R. (1996), 'Solving the Game of Checkers', *Games of No Chance* **29**, 119–133.
URL: <http://library.msri.org/books/Book29/files/schaeffer.pdf>
- Such, F. P., Madhavan, V., Conti, E., Lehman, J., Stanley, K. & Clune, J. (2017), 'Deep Neuroevolution: Genetic Algorithms Are a Competitive Alternative for Training Deep Neural Networks for Reinforcement Learning', *arXiv:1712.06567 [cs]* . arXiv: 1712.06567.
URL: <http://arxiv.org/abs/1712.06567>
- Toledo-Marin, J. Q., Diaz-Mendez, R. & Mussot, M. d. C. (2016), 'Is a good offensive always the best defense?', *arXiv:1608.07223 [cs]* . arXiv: 1608.07223.
URL: <http://arxiv.org/abs/1608.07223>
- Xie, L. & Yuille, A. (2017), 'Genetic CNN', *arXiv:1703.01513 [cs]* . arXiv: 1703.01513.
URL: <http://arxiv.org/abs/1703.01513>