

Speciation in NEAT

Timothy Nodine

5/7/2010

Turing Scholars Honors Thesis

Advisor: Risto Miikkulainen

Abstract

This paper investigates speciation in NEAT and its effect on NEAT's performance. Speciation is the set of processes by which NEAT creates, maintains and uses several disjoint groups of similar genomes for guiding reproduction. This paper also examines the issue of adoption, where mating within a species produces an individual in a different species. It then examines the relationship between NEAT's performance, the adoption rate and the number of species. The paper demonstrates that while species and adoptions are necessary, too many species and adoptions harm the algorithm's performance.

Introduction

An artificial neural network is a generalizable decision-making system. A neural network is made of nodes and weighted connections. Certain nodes are inputs activated by the environment. Node activations pass along one-way connections and cause or inhibit more activations in nodes to which the active nodes are connected. Designated output nodes have their activity interpreted as decisions.

Mathematically, a neural network is a $1 \times N$ vector representing the neurons' state, an $N \times N$ fixed matrix representing the connections between the neurons, and an activation function, usually a sigmoid, mapping real numbers to real numbers within a certain range. Some subset of the vector's values is designated as inputs, and these are set directly by the environment at each time step. The remaining part of the vector is set by multiplying the vector by the connection matrix, then applying the activation function to each neuron's value. Some subset of the nodes has its values read by the environment at each time step and interpreted as outputs.

NeuroEvolution of Augmenting Topologies (NEAT) is a genetic algorithm that produces neural networks. It was developed by UT PhD Kenneth Stanley in 2002[1]. As a genetic algorithm, it works by maintaining a large population of neural networks. It has each neural network perform a task, then evaluates its performance on the task. After each evaluation, it creates a new population by combining and mutating the most promising individuals. NEAT's most notable feature is that it varies the structure of the neural networks, producing more complex solutions over time.

NEAT needs its overall population of neural networks to be diverse enough that the mating produces a novel genome. To facilitate this, NEAT groups similar genomes into

species. Direct competition for survival happens only within a species. Species compete with each other for reproduction rights. Mating only happens within species (unless interspecies mating is explicitly allowed). Because NEAT uses species, it takes longer for the whole population to become uniform.

Related Work

This paper draws heavily on the experiments that Kenneth Stanley used when attempting to justify the various components of NEAT. In these experiments, which he called ablation experiments, he disabled one part of NEAT at a time and measured the change in performance. One of the results from these experiments, the removal of speciation from the algorithm completely, has been reproduced completely in the course of preparing this paper.[1]

Terminology

The following is a brief list of terms used in this paper:

- Organism, genome and individual - a single neural network within NEAT
- Species - one of several disjoint groups of similar genomes maintained by NEAT for population management
- Speciation – a collective term for the set of processes by which NEAT creates, maintains and uses species
- Mating – a process by which two existing genomes are combined to form a new genome. The combined genomes are called parents, and the new genome is the child
- Interspecies mating - mating where the two combined genomes are in different species
- Adoption - either an instance of mating where the child is in a different species from either of its parents or the process by which these adoptions occur

Experimental setup

This set of experiments used out-of-the-box NEAT 1.0.1, with interspecies mating disabled. Each experiment involved making minor changes to the code. These changes will be described with the corresponding experiment. With the exception of the initial adoption-tracking experiment, the changes related to each experiment were compiled out for the other experiments.

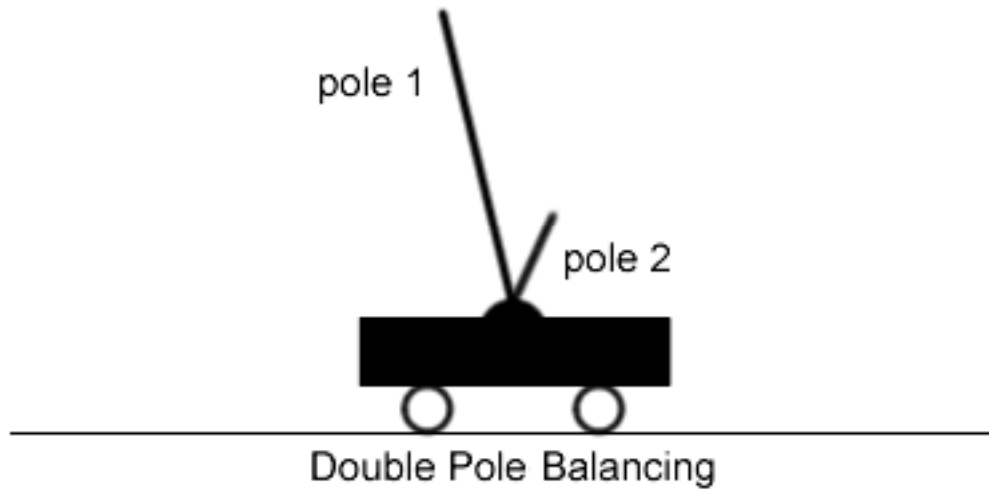


Figure 1 - Double Pole Balancing

The task used to evaluate NEAT was double pole balancing, with velocity provided. In this test, the NEAT controller operates a rail cart on a fixed length of track. The cart has two poles on it of different lengths, attached at one end by a hinge (see Figure 1). These poles are affected by gravity and quickly fall over if the controller does nothing. The controller receives information about the position of the cart and the angles and speeds of the poles. The controller's output is interpreted as a force and applied directly to the cart. The controller fails the task if either a pole falls over or the cart falls off the track. It succeeds if it manages to avoid failing within the first 100,000 input-output cycles, at which point it is assumed to be able to hold indefinitely. The individuals receive a score based on the number of input-output cycles before success or failure.

This test is a standard benchmark described in more detail in [1].

All experiments described in this paper were performed as batches of 100 independent runs of NEAT. Each run restarted from a random population and continued either until one member of the population succeeded at the task or 100 generations failed to do so. The main indicators of NEAT's performance are how many generations pass before a member succeeds at the task, and what proportion of runs produce a solution in the allotted number of generations. In these experiments, the algorithm also collected other statistics, most notably the number of species in the population after each generation and the number of adoptions during that generation.

NEAT worked across a constant-sized population of 200 genomes; 20 adoptions per generation corresponds to a 10% adoption rate. Interspecies mating is disallowed in all of these experiments except the Free Mating experiment.

Experiments

This section explains several individual experiments that built towards the eventual conclusion. The next section compares the results of those experiments

Baseline Experiment

For the first experiment, NEAT was modified to keep track of adoptions as they occurred. This modified version of NEAT provided the “Baseline” data for the various graphs and charts (Figures 2-4). Adoption is possible because NEAT assigns children to species based on raw genotypic similarity rather than parentage. It chooses a model from each existing species and compares the child's to each model. As soon as it finds a model with a similar enough topology, the child goes into that organism's species. If no model is similar enough, the child founds a new species.

The results of this experiment showed that the adoption rate was about 18%. Adoptions that created new species account for only a small fraction of these adoptions (<1%). The high adoption rate challenges the existing assumption that species remain relatively static. Ken Stanley asserts in [1, p.110] that “species do not overlap”. While true in the strictest sense, this assertion does not generalize across generations. The initial hypothesis explaining the higher than expected adoption rate was that the species' models are not kept constant between generations. In other words, the experiment showed that species are more fluid than previously realized.

Low Adoption Experiment

Just given the raw adoption rate, adoptions appear to be the main way in which species share information. Adoptions may be critically important to NEAT's overall function. For this reason the next several experiments were partial ablation experiments in which most genomes are assigned to their parent's species without even being considered for adoption. Only a small randomly-selected fraction (here, 1% or 10% of individuals) was assigned to species through the normal process. These are the 1% and 10% adoption rate experiments. The working hypothesis was that limiting the flow of information between species would cause NEAT to eventually stagnate and thus perform worse.

The results of the experiment were radically different from the hypothesis. NEAT performed much better with low adoption. The failure rate dropped significantly, to nearly nothing (~1%) with the strictest adoption policy. The actual number of adoptions in the low-adoption experiments decreased by far less than the 90 or 99 percent implied by the actual change to the algorithm. Because most adoptions were artificially prevented, species became far more diverse than the genetic comparison threshold would allow. As a result, the few individuals assigned to species the normal way were on average far more different from their parents' species' model genome and thus more likely to be adopted.

High Threshold Experiment

A major side effect of directly preventing adoptions is to limit the rate at which new species emerge (see Figure 2). The 1% adoption experiment had fewer than a tenth the average number of species as the baseline did. Previous work indicated that long runs tended to get bogged down with many species that were too small for normal crossover reproduction to be effective. This new result led to the hypothesis that the improved performance in the low adoption experiments was the result of a smaller number of species. Thus, the next investigation looked at the possibility that the low baseline performance resulted from too many species (i.e. “over-speciation”).

The High Threshold experiment raised the species’ acceptance threshold, so that a genome could be up to twice as different from the parent species’ model and still remain in the same species. The change should directly decrease the number of species and the adoption rate, without causing any major side effects.

The predicted performance boost happened, although not quite as strongly as expected. This experiment had a one other remarkable result, that the higher threshold reduced the number of species in circulation more than the 10% adoption experiment that directly prevented most new species from forming, while reducing the adoption rate less. The remainder of the results confirmed the hypothesis that limiting the number of species improves performance. The high threshold experiment had slightly worse success rate than the 10% adoption experiment, though this number was not significant. However, it had much higher adoption rate than the 10% experiment and a much lower number of species. If the success percentage had had a larger spread, and the assumption holds that the relationship between the critical factor (speciation or adoption) and performance is monotonic over the experimental range, then this would indicate that adoption was the critical factor as opposed to number of species.

Free Mating Experiment

One other experiment performed significantly changed how mating worked. The experiment cut speciation out of the mate selection routine. The purpose of this experiment was to investigate how useful speciation is as a heuristic for choosing mates. The changes enabled the property that any parent is equally likely to mate with a member of any other species. After this change, the only effect speciation had on reproduction was giving out reproduction rights to one of the two parents, based on species fitness.

The result of the Free Mating experiment was a slight but statistically significant drop in success rate relative to the baseline experiment. This result backs up the idea that speciation is mainly helpful as a mate selection heuristic. Although not notable by itself, this experiment’s result combines with the No Speciation experiment below result to hint at a stronger result.

No Speciation Experiment

The last experiment attempted to reproduce the experiment in [1] that demonstrated that speciation was necessary. The experiment disabled speciation completely by assigning all genomes to the same species. This is equivalent to 0% adoption. The hypothesis was

that disabling speciation would cause the genetic algorithm to perform worse, but would not cause it to fail catastrophically.

The results of the No Speciation experiment backed up the hypothesis. The modified algorithm's 57% success rate was significantly lower than the baseline. Furthermore, it performed almost exactly as well as the Free Mating experiment. By comparing the two, we see that without using species as a mating heuristic, using speciation to assign reproduction rights does not have a very large benefit if any.

Summary & Discussion of Results

The table in figure 2 below summarizes the results of the experiments. It gives the average number of species per generation, average number of adoptions per generation, and the percentage of the population that succeeded at the double pole balancing task within 100 generations. The first two columns are slightly biased towards the earlier generations, as each run stops leaving collectable statistics when it succeeds.

	Average # of species	Adoptions / Generation	Success %
No Species	1	0	57
Free Mating	76.54	31.71	57
Baseline	73.13	35.9	67
High Threshold	15.18	15.84	89
10% Adoption	28.71	9.32	92
1% Adoption	7.25	1.01	99

Figure 2- Summary of Results

All pairs of experiments have a statistically significant difference in success rate except for the No Species and Free Mating pair and the High Threshold and 10% Adoption pair. The lack of an observed difference in success rate between the Free Mating and No Species experiments seems to show that speciation mainly helps because of its use as a mate selector. Its other effects do not improve performance, at least not with as many species in the population as there were in the Free Mating experiment.

Among the other four experiments, the Baseline, 1% and 10% Adoption and High Threshold experiments, a lower adoption rate always corresponded to a better performance. The difference between the High Threshold and 10% Adoption experiments indicates that the corresponding relationship between performance and speciation does not hold. These observations lead to the conclusion that adoption is indeed the critical factor affecting NEAT's performance, rather than over-speciation. If the improvement was simply a result of having fewer species, the 10% adoption experiment would have performed worse than the high-threshold experiment.

The graphs in Figures 3 and 4 plot the number of species and adoptions per generation, respectively, against the amount of time in generations since the start of the run. They show the average of all runs from each separate experiment. Neither graph takes into account runs on generations after they succeeded. As a result, the 1% Adoption line is unreliable after about generation 50 because the high success rate decreased the sample size to an unreliable level. The High Threshold and 10% Adoption line also suffer from the same effect at the far-right end of the graph. Also, the “No Species” line is not shown because by definition it always has one species and no adoptions.

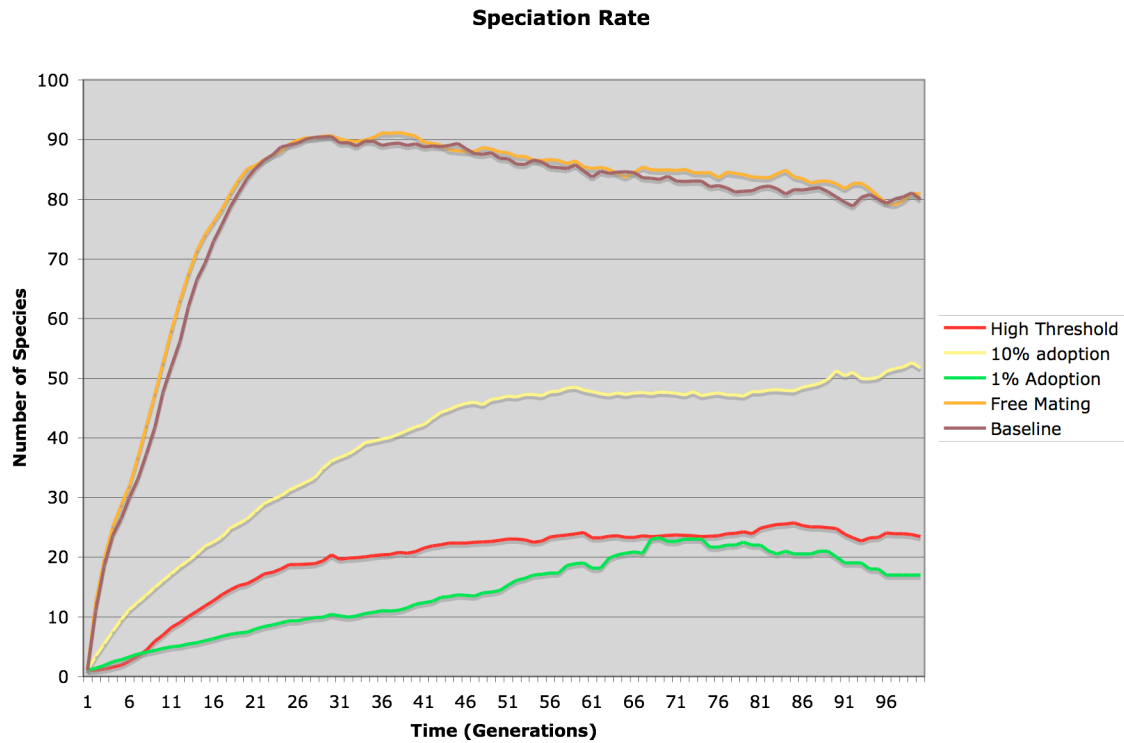


Figure 3 - Number of species changing over time

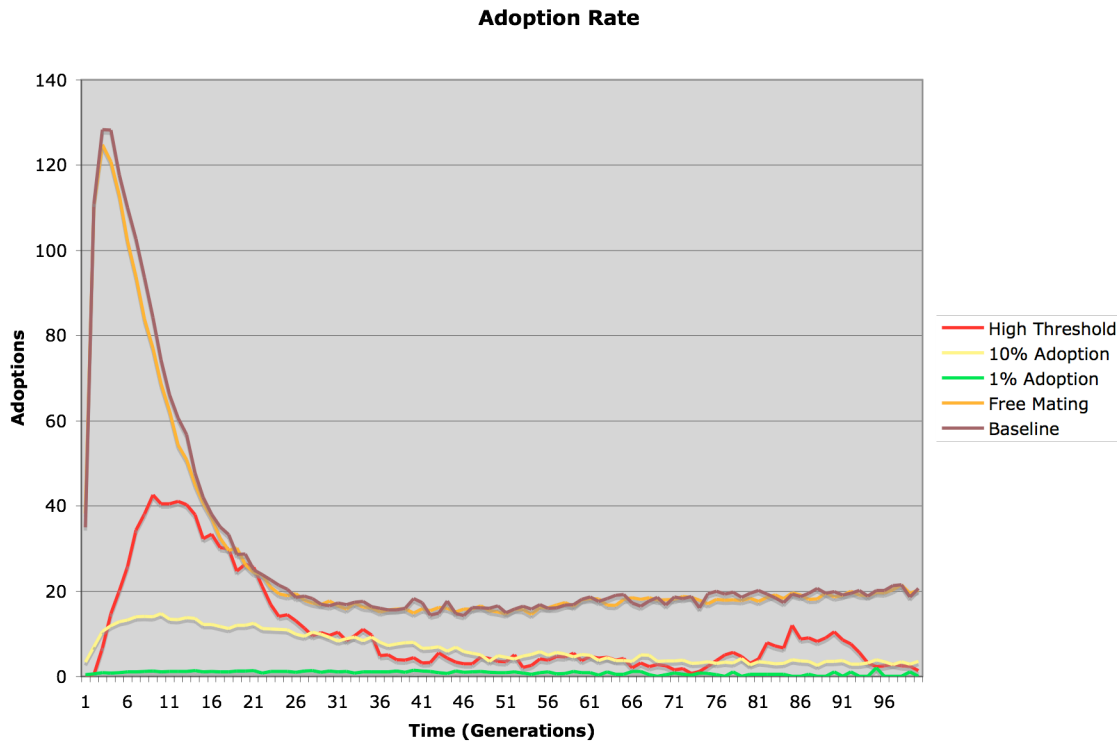


Figure 4 - Number of adoptions changing over time

The adoption rate approximates the derivative of the Speciation Rate early on, suggesting that most of the adoptions early on are a result of new species being populated by adoption. Later on, when the number of species is relatively constant, adoptions still occur despite the presence of few new species. The late anomaly in the High Threshold experiment might be the result a major event in a single run, such as two large species becoming overlapped so that one of them absorbed the other over a period of several generations.

Conclusions and Future Work

The experimental results in this paper demonstrate that species are an effective heuristic for mate selection in NEAT. A version of NEAT that does not use species for mate selection (Free Mating) performed worse than unmodified NEAT, and did not appear to perform any better than NEAT without species.

Another conclusion from these experiments is that adoption is significant and worth paying attention to. Proper use of adoption controls had a large positive effect on the algorithm's performance, as shown in the 1% Adoption experiment. An overabundance of species also harms performance as shown in the High Threshold experiment, though this effect may be linked to the adoption effect.

There are several courses that future work might take to expand on these results. The simplest would be to take a more thorough sample of adoption rate/threshold pairs. More

data points are necessary to find the ideal adoption rate to maximize NEAT's performance.

Another line of future work is to check the same experiments on a more complex task, such as one with several co-dependant outputs. It may be the case that a task so complicated that the no-speciation version cannot solve it at all forces adoption to take a more active, helpful role. This would demonstrate whether the results from this one domain generalize well.

Finally, the plot of adoption rate by generation shows that early adoptions are used to populate new species while later adoptions are exchanges between existing species. It may be worthwhile to experiment with restricting adoptions only at specific times in order to better understand what role adoption plays.

Bibliography

- [1] Kenneth O. Stanley. "Efficient Evolution of Neural Networks Through Complexification" Ph.D. Thesis; Department of Computer Sciences, The University of Texas at Austin. Technical Report~AI-TR-04-314, August 2004
- [2] Kenneth O. Stanley and Risto Miikkulainen. "Evolving Neural Networks Through Augmenting Topologies" *Evolutionary Computation* 10(2):99-127, 2002.