An Investigation of Topological Choices in FS-NEAT and FD-NEAT on XOR-based Problems of Increased Complexity

Evgenia Papavasileiou*
(1) Vrije Universiteit Brussel (VUB), Department of Electronics and Informatics (ETRO)
Pleinlaan 2, B-1050 Brussels, Belgium
(2) imec
Kapeldreef 75, B-3001 Leuven, Belgium
epapavas@etrovub.be

Bart Jansen
(1) Vrije Universiteit Brussel (VUB), Department of Electronics and Informatics (ETRO)
Pleinlaan 2, B-1050 Brussels, Belgium
(2) imec
Kapeldreef 75, B-3001 Leuven, Belgium
bjansen@etrovub.be

ABSTRACT

Feature Selective Neuroevolution of Augmenting Topologies (FS-NEAT) and Feature De-selective Neuroevolution of Augmenting Topologies (FD-NEAT) are two well-known methods for optimizing the topology and the weights of Artificial Neural Networks (ANNs) while simultaneously performing feature selection. Literature has shown that starting the evolution with ANNs of one hidden layer can affect FD-NEAT's and FS-NEAT's performances. However, no study exists that investigates the effects of changing the networks' initial connectivity. In this paper we investigate how the choice of the number of initially connected inputs affects the performance of FD-NEAT and FS-NEAT in terms of accuracy, number of generations required for convergence, ability of performing feature selection and size of the evolved networks. For this purpose we employ artificial datasets of increasing complexity based on the exclusive-or (XOR) problem with irrelevant features. The different initial topological settings are compared using Kruskal-Wallis hypothesis tests with Bonferroni correction ($p<0.01$), while FD-NEAT and FS-NEAT are compared using Wilcoxon rank sum hypothesis tests ($p<0.01$). The results show that the initial connectivity setting does not affect the performance of FD-NEAT and FS-NEAT.

CCS CONCEPTS

•**Computing methodologies → Neural networks; Genetic algorithms; Feature selection;**

KEYWORDS

neuroevolution, topology initialization, supervised learning

ACM Reference format:
Evgenia Papavasileiou and Bart Jansen. 2017. An Investigation of Topological Choices in FS-NEAT and FD-NEAT on XOR-based Problems of Increased Complexity. In *Proceedings of GECCO '17 Companion, Berlin, Germany, July 15-19, 2017,* 4 pages.

*Research funded by a PhD grant of the Research Foundation Flanders (FWO)

978-1-4503-4939-0/17/07…$15.00
DOI: http://dx.doi.org/10.1145/3067695.3082497

DOI: http://dx.doi.org/10.1145/3067695.3082497

1 INTRODUCTION

NeuroEvolution of Augmenting Topologies (NEAT) [10] is a neuroevolutionary method that uses Genetic Algorithms (GAs) for learning the connection weights as well as the topology of Artificial Neural Networks (ANNs). Its successor, called Feature Selective NEAT (FS-NEAT) [13], performs feature selection simultaneously with the optimization of the topology and the connectivity of the underlying nodes. Feature De-selective NEAT (FD-NEAT) [12] is an alternative, promising method for classification tasks [6, 11, 12].

To our knowledge only one study exists [9] where FD-NEAT and FS-NEAT are systematically compared on the non-linear XOR problem of multiple dimensions (5, 10 and 20 inputs). In [9] the authors investigate the effect of beginning the evolution with a population of ANNs with one hidden layer. Another important topological setting that has not been studied before concerns the selection of the number of inputs that should be connected in the initial topologies. FD-NEAT starts the evolution from fully connected single layer networks (i.e. networks with all the input nodes directly connected to the outputs), whereas FS-NEAT starts with networks with only one arbitrary input node connected. As far as the feature selection is concerned, it is believed [12] that FS-NEAT outperforms FD-NEAT in problems where the majority of inputs are redundant or irrelevant, whereas FD-NEAT outperforms FS-NEAT in problems where most of the inputs are relevant. Choosing from which topology one should start, requires knowing the relevant features a priori, which is of course impossible, as it is the task in hand. In this paper we investigate whether a different connectivity setting can have an effect on FD-NEAT's and FS-NEAT's ability of performing feature selection, classifying samples as well as on their efficiency measured by the number of generations required for convergence and the size of the evolved networks.

2 METHODS

2.1 Neuroevolution

Neuroevolution (NE) is a learning method that uses GAs to optimize the parameters of ANNs. It started as a method to evolve only the connection weights of fixed topology ANNs [3, 4, 7, 8, 14] while later it advanced into a method of optimizing both the weights and the topology of the underlying nodes (Topological and Weight Evolving Artificial Neural Networks (TWEANNs)) [1, 2, 10, 15]. TWEANNs offered significant advantages, as finding the optimal topology of an ANN requires time-consuming evaluations of potential architectures.

2.2 NEAT

NEAT [10] is a TWEANN method which encodes the ANNs with two types of genomes: connection genes and node genes. The ANNs' structure is evolved over the generations by mutation and crossover operations. The evolution starts minimally with single layer networks whose structure becomes gradually more complicated over generations by mutation operators that add new connections and new nodes in the networks. The crossover between networks is facilitated by historical markings, i.e. when a new gene is added it is assigned a globally incremented innovation marker which facilitates the alignment of matching genes. Moreover, NEAT protects innovation by grouping individuals of similar topology into species so that they compete within their own niche instead of the whole population. Finally, NEAT tends to discover networks without unnecessary structure by starting the evolution with a population of minimal structures and keeping those topological innovations that are found to increase the network's performance.

2.3 FS-NEAT

FS-NEAT [13] is a NE method that extends NEAT in performing feature selection simultaneously with learning the ANNs topology and weights. It follows the three basic principles of NEAT; historical markings, speciation and starting with minimal structure. The difference lies that it starts the evolution even more minimally. The initial ANNs' topologies in NEAT consist of an input and an output layer with all the inputs directly connected to the output nodes, whereas in FS-NEAT only one random input is connected to one random output node. In the course of generations more inputs will be connected but only the connections that come from the relevant inputs tend to survive, thus performing implicit feature section.

2.4 FD-NEAT

FD-NEAT [12] is an extension to original NEAT functioning similarly to FS-NEAT in terms of performing feature selection simultaneously with topology and weight learning. The main difference between FS-NEAT and FD-NEAT lies in the way that feature selection is performed. FD-NEAT starts with the same minimal topologies as NEAT, i.e. fully connected single layer networks and drops irrelevant inputs throughout the evolution. Only the inputs that result in increasing the performance of the individual tend to survive and in this way FD-NEAT performs implicit feature selection.

3 EXPERIMENTAL SET-UP

3.1 Dataset

To be able to evaluate the effect of the investigated topological settings we use artificial benchmark datasets so that we are aware of their expected behaviour. The XOR problem is one of the first datasets a researcher would consider to verify the success of their approach [5, 10, 12]. However, XOR is a problem too easy to be learnt with no feature selection issue, so we need to construct more complex datasets. Towards this purpose, we build artificial 2 out of k datasets (referring to as $2/k$), where $k \in \{5, 10, 20\}$. The 2 inputs are assigned to the relevant features and the remaining $k-2$ inputs are assigned to irrelevant booleans. To include these irrelevant features we increase the number of samples to avoid imposing any bias, as the larger the dataset the less probable that there is an underlying correlation between the randomly generated data and the output. We use the same number of samples as in [9], equal to 700. The resulting datasets are not easy to be learnt and they constitute a challenging task for a feature selection algorithm, as they are constructed in such a way so that each of the individual attributes is equally informative for predicting the output.

3.2 Proposed Method

For each of the aforementioned 2 out of k datasets, ($k \in \{5, 10, 20\}$) we vary the number of initially connected inputs from 1 to k connected inputs in single layer ANNs. For each of the investigated connectivity settings we separate the datasets into training and test sets using 10-fold cross validation. We repeat each investigation 10 times that results in 100 repetitions for each topological choice. The parameters used for the setting of FD-NEAT and FS-NEAT are the same as in [9]. The fitness function of FD-NEAT and FS-NEAT is based on the error between the output of the ANN and the correct output of the training set and it is given by $Fitness = (N - \sum_{i=1}^{N} |O_i - T_i|)^2$ [9, 10], where N is the size of the training dataset, O_i the output of the ANN on the i^{th} pattern of the training set and T_i the real output that corresponds to the i^{th} pattern of the training set.

3.3 Performance Evaluation

The effect that the different settings of connectivity patterns have in the performance of the FD-NEAT and FS-NEAT is evaluated on different aspects. First of all, the altered algorithms are evaluated on their ability to correctly classify samples. This is measured by the accuracy on the test set which is defined by the portion of correctly classified samples and it indicates how successful the method is in finding the right relationship between input and output.

Furthermore, the algorithms are compared on their ability to perform feature selection. For this purpose we employ the measure of the average of absolute weights of the connections that initiate from each input in the best networks of the final population. It is assumed that FD-NEAT and FS-NEAT learn to assign higher weights to the relevant inputs compared to the irrelevant ones. According to this, the connections that initiate from the relevant inputs should have higher values than the connections that initiate from the irrelevant ones. If there is a statistical difference between the values among relevant and irrelevant inputs then we could argue that the algorithm has a feature selection ability.

Finally, the algorithms are compared on their efficiency. This is measured by the number of generations that are required in order for the algorithm to converge to the solution. Smaller number of generations indicate faster algorithms. Finally, we compare the structure of the final evolved neural networks in terms of the number of connections and nodes that are evolved.

3.4 Analysis of the Results

The first goal of this study is to investigate the influence that a different connectivity setting has in the different measures described

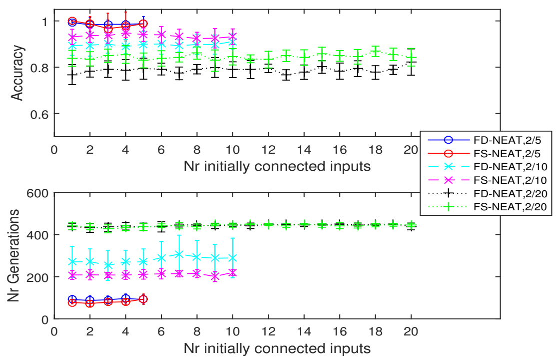


Figure 1: Average accuracy on the test set and average number of generations of FD-NEAT and FS-NEAT for different numbers of initially connected inputs on the 2/k XOR problems.

in the previous section. Towards this goal we perform multiple comparisons with hypothesis tests (Kruskal-Wallis test with Bonferonni correction, $p<0.01$) to investigate whether a connectivity setting results in networks whose performance is significantly different than the performance of another connectivity setting. In order to examine the feature selection ability of the altered FD-NEAT and FS-NEAT for each of the different connectivity settings we perform hypothesis tests (Wilcoxon rank sum test, $p<0.01$) on the values between the relevant and the irrelevant inputs of the average of absolute weights. Finally, we compare the performance between FD-NEAT and FS-NEAT by performing hypothesis tests (Wilcoxon rank sum test, $p<0.01$) between all the performance measures of the two algorithms.

4 RESULTS

In Figure 1 (top) we present the results of accuracy on the test set for the 2/5, 2/10 and 2/20 XOR problems as a function of the different number of connected inputs in the initial topologies. It is observed that the ability of the algorithms to solve the problem decreases as the complexity of the problem increases, e.g. the accuracies in the 2/20 problem are lower than the accuracies in the 2/10 which are lower than the ones of the 2/5. For all the three problems there is no statistical difference among the different connectivity patterns (Kruskal-Wallis, $p<0.01$). Overall, FS-NEAT seems to perform better than FD-NEAT. FS-NEAT's accuracy is statistically different than FD-NEAT's at the 2/20 inputs problem (Wilcoxon, $p<0.01$) for all the chosen connectivity patterns except for the ones of 5, 9 and 20 connected inputs, from which the last one is the default initial connectivity topology for FD-NEAT.

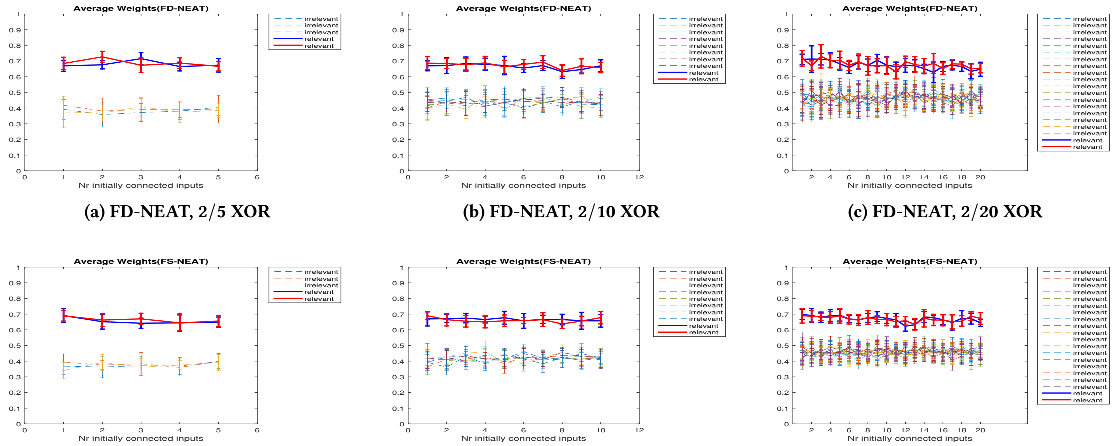
Figure 1 (bottom) shows the number of generations required by FD-NEAT and FS-NEAT to converge, or in case of no convergence the maximum number of generations they were allowed to evolve. From the graphs it is clear that the harder the problem the more generations are required for the convergence. Also, the maximum limit of 450 generations was not enough for the convergence of the 2/20 XOR problem. The comparisons among the different connectivity settings reveal that there is no significant difference (Kruskal-Wallis, $p<0.01$) among them. Comparing FD-NEAT and FS-NEAT, it seems that FD-NEAT in general requires more generations but a significant difference (Wilcoxon, $p<0.01$) exists only at the 2/10 XOR for setting the initial connectivity to 2, 4, 7-9 connected inputs.

In Figure 2 we present the graphs of the average absolute weights of the connections that initiate from the input nodes. By performing Wilcoxon hypothesis tests ($p<0.01$) on the difference between the relevant and the irrelevant inputs we observe that a significant difference between them always exists independently of the choice of the initial connectivity setting. This means that both FD-NEAT and FS-NEAT are suitable algorithms for performing feature selection. Next, we examine whether a connectivity pattern results in a difference of relevant-irrelevant inputs which is significantly different than the one of another connectivity setting. At the 2/5 and 2/10 XOR no significant difference exists among the different connectivity settings (Kruskal-Wallis, $p<0.01$). At the 2/20 XOR, the connectivity setting of 1 initially connected input is statistically different (Kruskal-Wallis, $p<0.01$) than the settings of 11 and 19 connected inputs in FD-NEAT and 13 connected inputs in FS-NEAT. Comparing the performances between FD-NEAT and FS-NEAT, no significant differences were found (Wilcoxon, $p<0.01$).

Finally, in Figure 3 we investigate the size of the evolved networks in terms of number of nodes and connections. At the 2/5 and 2/10 XOR problems there is no significant difference (Kruskal-Wallis, $p<0.01$) among the different connectivity settings. At the 2/20 XOR problem, there is a significant difference for FD-NEAT between the setting of 20 inputs and the one of 1 and 3 inputs. In comparison to FS-NEAT, FD-NEAT evolves networks with statistically (Wilcoxon, $p<0.01$) more nodes than FS-NEAT for connectivity setting of 1 and 2 initially connected nodes at the 2/10 XOR problem. On the other hand, FS-NEAT evolves networks with statistically (Wilcoxon, $p<0.01$) more nodes than FD-NEAT's for connectivity setting of 1 and 11 initially connected inputs at the 2/20 XOR problem. At the same problem, FS-NEAT evolves networks with significantly more connections (Wilcoxon, $p<0.01$) than FD-NEAT for most of the connectivity settings.

5 CONCLUSION

Our analysis shows that different numbers of inputs connected in the initial topology does not affect the performance of FS-NEAT and FD-NEAT. This means that no initial connectivity pattern is proven better than another. So, even if FS-NEAT starts the evolution having irrelevant inputs connected it will be still able to reach the same accuracy in the same amount of generations. Also, it is observed that both FD-NEAT and FS-NEAT have a feature selection ability which is independent of the number of inputs that are initially connected. In addition, both algorithms learn to assign weights of similar magnitude to the relevant and irrelevant inputs independent of the number of inputs initially connected. This means that even when FD-NEAT starts with one connected input or FS-NEAT starts from fully connected networks they both learn to assign smaller weights to the irrelevant inputs and higher weights to the relevant ones. Furthermore, by comparing the feature selection abilities of FD-NEAT and FS-NEAT we observe that FS-NEAT's is significant better than FD-NEAT's at the cases when FD-NEAT does not start from fully connected networks. We can therefore conclude that



(a) FD-NEAT, 2/5 XOR (b) FD-NEAT, 2/10 XOR (c) FD-NEAT, 2/20 XOR

(d) FS-NEAT, 2/5 XOR (e) FS-NEAT, 2/10 XOR (f) FS-NEAT, 2/20 XOR

Figure 2: Average absolute weights of input connections of FD-NEAT and FS-NEAT for different numbers of initially connected inputs on the 2/k XOR problems.

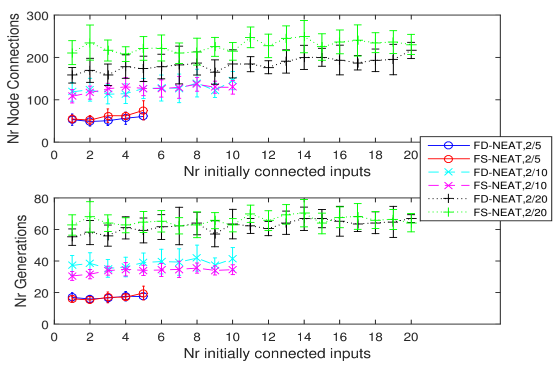


Figure 3: Average number of connections and average number of nodes in the champion network of the final population in FD-NEAT and FS-NEAT for different numbers of initially connected inputs on the 2/k XOR problems.

FD-NEAT should start the evolution with its default fully connected setting, while FS-NEAT seems not to be affected by the number of inputs originally connected. Finally comparing the default connectivity settings of FD-NEAT and FS-NEAT, even though FS-NEAT seems to perform better, a statistical difference does not exist.

At this point, we should take into account that these conclusions are limited to datasets with similar format as the 2/k XOR, i.e. with fewer relevant inputs and more irrelevant ones. Therefore, the behaviour of FS-NEAT and FD-NEAT in datasets with different ratio between relevant and irrelevant inputs still needs to be investigated. For this reason, we are going to extend the set of experiments on problems of increased complexity containing more features to approach more realistic problems by also employing benchmark datasets, such as the spiral plots with irrelevant features.

REFERENCES

[1] Peter J Angeline, Gregory M Saunders, and Jordan B Pollack. 1994. An evolutionary algorithm that constructs recurrent neural networks. *IEEE Transactions on Neural Networks* 5, 1 (1994), 54–65.

[2] David T Cli, Inman Harvey, and Philip Husbands. 1992. Incremental evolution of neural network architectures for adaptive behaviour. In *Proceedings of the European Symposium on Artificial Neural Networks (ESANN'93)*. 39–44.

[3] Faustino Gomez and Risto Miikkulainen. 1997. Incremental evolution of complex general behavior. *Adaptive Behavior* 5, 3-4 (1997), 317–342.

[4] Faustino Gomez and Risto Miikkulainen. 1998. 2-d pole balancing with recurrent evolutionary networks. In *Proceedings of the 8th International Conference on Artificial Neural Networks*. Springer, 425–430.

[5] Derek James and Philip Tucker. 2004. A comparative analysis of simplification and complexification in the evolution of neural network topologies. In *Proceedings of Genetic and Evolutionary Computation Conference*.

[6] B. Jansen, M. Tan, I. Bautmans, B. Van Keymolen, T. Mets, and R. Deklerck. 2011. Accelerometer based gait analysis. In *Proceedings of the International Conference on Bio-inspired Systems and Signal Processing*. 138–143.

[7] David J Montana and Lawrence Davis. 1989. Training Feedforward Neural Networks Using Genetic Algorithms.. In *IJCAI*, Vol. 89. 762–767.

[8] David E Moriarty and Risto Mikkulainen. 1996. Efficient reinforcement learning through symbiotic evolution. *Machine learning* 22, 1-3 (1996), 11–32.

[9] Evgenia Papavasileiou and Bart Jansen. 2016. A comparison between FS-NEAT and FD-NEAT and an investigation of different initial topologies for a classification task with irrelevant features. In *2016 IEEE Symposium Series on Computational Intelligence (SSCI)*. 1–8. DOI: http://dx.doi.org/10.1109/SSCI.2016.7850127

[10] Kenneth O Stanley and Risto Miikkulainen. 2002. Evolving neural networks through augmenting topologies. *Evolutionary Computation* 10, 2 (2002), 99–127.

[11] Maxine Tan, Rudi Deklerck, Bart Jansen, and Jan Cornelis. 2012. Analysis of a feature-deselective neuroevolution classifier FD-NEAT in a computer-aided lung nodule detection system for CT images. In *Proceedings of the 14th Annual Conference Companion on Genetic and Evolutionary Computation*. ACM, 539–546.

[12] Maxine Tan, Michael Hartley, Michel Bister, and Rudi Deklerck. 2009. Automated feature selection in neuroevolution. *Evolutionary Intelligence* 1, 4 (2009), 271–292.

[13] Shimon Whiteson, Peter Stone, Kenneth O Stanley, Risto Miikkulainen, and Nate Kohl. 2005. Automatic feature selection in neuroevolution. In *Proceedings of the 7th Annual Conference on Genetic and Evolutionary Computation*. ACM, 1225–1232.

[14] Darrell Whitley, Timothy Starkweather, and Christopher Bogart. 1990. Genetic algorithms and neural networks: Optimizing connections and connectivity. *Parallel computing* 14, 3 (1990), 347–361.

[15] Byoung-Tak Zhang and Heinz Muhlenbein. 1993. Evolving optimal neural networks using genetic algorithms with Occam's razor. *Complex systems* 7, 3 (1993), 199–220.