Heuristic Pattern Reduction Technique and Evolutionary Algorithm for Neural Networks: Application to identification of Abalone age

Jacob Wong
Research School of Computer Science,
Australian National University,
ACT 0200, AUSTRALIA

Abstract. This report uses the Heuristic Pattern Reduction technique to find that it does not improve the generalisation ability of a neural network with evolutionarily selected hyper parameters. But the statistical significance of this result is suspect as it is likely that the limited dataset is the true cause of the limited results. The neural networks also managed to perform better than a Cascor network, but the causation of that performance is again unclear. Despite the use of these techniques, the problem of identifying Abalone ages is likely to be unsolvable in its currently formulated dataset.

1 Introduction

There has been some suggestion that the reduction of the input set in a neural network can help the model better generalize after training (Gedeon and Bowden, 1993). This is inspired by a body of research which suggests that limiting the resources that a neural network has, by pruning the number of neurons or halting training with validation sets, can force the network to generalize. This report seeks to test the hypothesis that the reduction of a neural network's input set can also force the neural network to generalize.

The data used for this experiment comes from the UCI Machine Learning Repository and was donated by Sam Waugh from the University of Tasmania. The problem in this data involves determining the age of abalone. The age of abalone is usually determined by cutting through the shell, staining it and counting the number of rings through a microscope. However, this is a boring and time-consuming task (Waugh, 1995). Instead, the more easily attainable physical features of length, diameter, height etc might be able to determine the age of the abalone. So the purpose of the model is to predict the age of abalone given their measured physical characteristics.

A fully connected back-propagation artificial neural network will be used in this experiment. The rest of the report will refer to this as the 'neural net'. Given the range of ages that an abalone could be, this will be a multi-classification problem. The data that will form the input for the neural net has 8 attributes and 4177 instances. The technique that will be used to improve this network is the Heuristic Pattern Reduction Technique (HPR). HPR is a technique posed by T.D Gedeon and T.G. Bowden based the idea that a reduction of complexity in a training set can improve learning. By reducing the input set, the error surface is simplified and training duration is accelerated.

Further, this neural network's hyper parameters will be determined using an evolutionary algorithm. This will involve a guided stochastic global search to find the best hyper parameters suited for this model.

One particular reason for selecting this dataset is because past attempts at solving this problem with a neural network have resulted in low test-set performance which is an indicator of poor generalisation. By forcing the model to generalize, the test-set performance of the model might be improved. For the purposes of this experiment, generalization will be measured by the accuracy of the model on the test-set.

2 Method

2.1 Input Processing

Most of the inputs for this dataset had already been pre-processed. The length, diameter, height, whole weight, shucked weight, viscera weight and shell weight were all continuous and normalized to 0-1. The only change made to the encoding was to the sex attribute. As there is no relative relationship between the Male, Female and Infant categories, three separate inputs were used to represent this attribute. One of these inputs would be 1 depending on which category it was while the others would be 0. This prevents the neural net from learning a relationship between the inputs on that attribute where there is none.

The dataset was split into ¾ for the training set and ¼ for the testing set, which is 3161 and 1044 instances respectively. Rare examples, such as 1, 2, 25, 26, 27, and 29 where there was only 1-2 instances available, were duplicated in the training set to ensure that their features were learned. The neural network does not see any of the test set until its training is complete, this will provide a better indication on whether the neural network has become sensitive to the data and whether it is capable of generalisation.

2.2 Model Topology

Given the inputs as described above, the neural network has 10 input neurons. In this dataset, there exists labelled data for abalone aged between 1-29 years. As such, the neural network has 29 output neurons. An evolutionary algorithm was used to determine the best hyper parameters for the neural network. The evolutionary algorithm will be tasked with determining the number of hidden neurons, the learning rate, the learning rate decay multiplier and the number of epochs for training (Leung et al. 2003). For comparison, this model will be compared against a model with manually determined hyper parameters. This will have 7 hidden neurons, a learning rate of 1, a learning rate decay multiplier of 0.1 and 3000 epochs for training. These parameters were determined by trial and error guided by rules of thumb.

The evolutionary algorithm will have a population size of 50 and will be run for 100 generations. The replacement strategy used here is the replace worst strategy, where the offspring replaces the worst individuals of the current population. Here, the worst half of the population will be replaced with new offspring. But there is a 5% chance that an individual that is supposed to be removed will be kept, this is to avoid a reduction of diversity by removing the worst offspring which might all have similar features. The offspring are also given a mutation chance of 15% initially. This mutation chance reduces with each passing generation proportionately to the number of generations past, resulting in 0% mutation by the final generation. This mutation rate introduces diversity and allows the algorithm to perform a guided global search. As the algorithm approaches an acceptable solution, the mutation rate is reduced to ensure that convergence occurs and the solution can be more finely optimised.

Table 1. Evolutionary Algorithm Progress

Generation	Mutation chance (%)	Highest Fitness (%)	Average Fitness (%)
1	15	29.30	27.20
25	11.25	29.91	29.46
50	7.5	30.03	29.47
75	3.75	29.91	29.53
100	0	29.84	29.52

Although the algorithm was run for 100 generations, the improvement to fitness in the population quickly plateaued as seen in the average fitness of the population. After 25 generations, most individuals were very similar and there seemed to be no more noticeable improvements to fitness.

Upon the algorithm's completion, the best solution found was:

Table 2. Final hyper parameter values for neural network

Hyper parameter	**Value**
Hidden Neurons	13
Learning Rate	1
Learning Rate Decay Multiplier	0.5
Epochs	3000

For both models, there is only one hidden layer and leaky ReLU has been used as the activation functions. This is to mitigate the phenomena of dead neurons that are possible when using ReLU by giving them a chance to recover. The loss function uses Cross Entropy Loss which penalizes answers that are confidently wrong more, optimizing the neural network to guess more conservatively. This is helpful particularly since a close approximation is still useful in the context of the problem.

2.3 Heuristic Pattern Reduction and Evaluation Technique

Similar to T.D. Gedeon and T.G. Bowden's Heuristic Pattern Reduction technique, the input set is reduced by a fifth, a fourth, a third and a half. The inputs that were removed were chosen randomly. Each of these configurations were then run 15 times with different initial weights.

For each run's results, the loss of training set, accuracy on training set and accuracy on test set was recorded. This evaluation would provide an indication of the neural network's learning and whether any over-fitting had occurred. The results would then be aggregated to the best test set accuracy for each of different input sizes.

3. Results and Discussion

Table 3. Aggregation of average accuracy from 15 attempts on test set using manually determined hyper parameters

Input Size	Loss on training set	Accuracy on training set (%)	Accuracy on test set (%)
100%	1.9207	28.84	27.13
80%	1.9028	28.38	27.97
75%	1.9312	28.30	27.40
66%	1.9119	29.94	26.46
50%	1.9167	29.42	27.02

Table 4. Aggregation of average accuracy from 15 attempts on test set using evolutionary algorithm determined hyper parameters

Input Size	Loss on training set	Accuracy on training set (%)	Accuracy on test set (%)
100%	1.8948	29.38	27.50
80%	1.8936	28.86	27.62
75%	1.8975	28.64	27.58
66%	1.8915	30.45	26.46
50%	1.8701	30.15	27.33

As seen in the tables above, there is very limited improvement in the neural network's accuracy in the test set as the input set decreases. Nor is there a significant difference between the test set results produced by both models.

In T.D. Gedeon and T.G. Bowden's case, their best result was found in removing half of the input set (Gedeon and Bowden, 1993). Here, the 'best' result would be the 80% input size on the manually determined model. Even then, this improvement is likely to be statistically insignificant. The reduction of the input size does seem to improve the accuracy on the training set, but is due to the reduction of the problem complexity and is actually causing over fitting in both cases.

Between the two models, the model with evolutionary determined hyper parameters has a lower overall loss and greater training accuracy. But this did not translate to a higher accuracy on the test set which again indicates that over fitting had occurred.

3.1 Comparison with Cascor

Unfortunately, there was no other previous research on this dataset that involved an evolutionary algorithm. However, Cascor remains a comparable network as both Cascor and the evolutionary algorithm used here determines the number of hidden neurons to be used.

Cascor is a technique whereby the hidden neurons are generated and added during training. The goal of which is to systematically determine the number of hidden neurons that the model should have. In Sam Waugh's paper, this dataset was used train a Cascor network and the result was 27.66% accuracy on training and 24.90% accuracy on the test set (Waugh, 1995).

Both neural networks here produced results on the test set that were generally better than the Cascor network. However, it is difficult to say that this is because the HPR technique and evolutionary algorithm forced the models to generalize better. There are a few caveats to this comparison.

The first is that the neural network was given 3000 epochs to learn from the training set data whereas the Cascor network was restricted to 100 epochs. This likely had a significant effect on the learning of the Cascor network. While pre-processing the input, the rare examples were duplicated to give the neural network a better chance of learning those cases. There is no indication of this in Waugh's paper, which is likely another factor involved in the result. In light of these caveats, it is probably unlikely that the neural network performed better because of the HPR technique or the evolutionary algorithm.

3.2 Limitations with the evolutionary algorithm

The result of the evolutionary algorithm in this experiment suggested that 13 neurons and 3000 epochs were the best fit. However, there are a few limitations in the way that the evolutionary algorithm was run which likely negatively influenced the results.

The fitness function of the algorithm was determined by training a network with an individual's parameters and comparing by fitness. However, this failed to take into account the possibility of over fitting. Both 13 neurons and 3000 epochs were the highest options available and they would naturally have the highest training accuracy. But this high training accuracy, as shown, does not necessarily result in high testing accuracy or good generalisation ability.

3.3 Limitations with dataset

One big limitation, as stated by Waugh, is the problem of overlapping classes in the input data (Waugh, 1995). This hinders learning as very similar input might have two different labelled answers, preventing proper learning. Further, it is also likely that more features are required for this problem to be solved. Features such as the region captured may provide further generalisation information like whether the abalone grew in an area exposed to cold water (Waugh, 1995). Until additional information is provided, it is unlikely this problem can be solved.

4. Conclusion and Future Work

In this experiment, a neural network was used to classify the age of abalone using their physical characteristics. The neural network's hyper parameters were determined using an evolutionary algorithm to provide the most acceptable settings. The input set was then reduced using the HPR technique to determine if the technique could improve the model's generalisation ability.

From the results, it is unclear that the reduction of the input size or use of evolutionary algorithm served to improve the neural networks' generalisation ability. Both models over fitted which compounded the effect of the limited data set to produce very limited results. In comparison to a Cascor network, the neural networks performed generally better, but this again unlikely to have been caused by the HPR technique or evolutionary algorithm.

In future, the Heuristic Pattern Reduction technique and evolutionary algorithm could be further tested on a wider range of datasets in order to see if these techniques can improve generalization generally. The Heuristic Pattern Reduction technique and evolutionary algorithm could also be applied to models with different topologies like a Cascor network or a convoluted network. This would determine whether the generalization improvement is amplified or mitigated by the effects of those topologies.

4. References

Gedeon, T.D. and Bowden, T.G. (1993) "Heuristic Pattern Reduction II," Proceedings ICCS, Invited Paper, Beijing, vol. 3: 43-45.

Sam Waugh (1995) "Extending and benchmarking Cascade-Correlation", PhD thesis, Computer Science Department, University of Tasmania.

Sam Waugh, (1995). Abalone Data Set, [https://archive.ics.uci.edu/ml/datasets/Abalone], University of Tasmania.

Leung, F. H., Lam, H. K., Ling, S. H., & Tam, P. K. (2003). Tuning of the structure and parameters of a neural network using an improved genetic algorithm. Neural Networks, IEEE Transactions on, 14(1), 79-88.