Determining major factors that influenced secondary school student final grade mark using Neural Network and Evolutionary Algorithm with student's assessment performance and supplementary questionnaires data

P. Anivan

College of Engineering and Computer Science
Australian National University
Canberra ACT 2600
AUSTRALIA

Abstract. There are many factors that can contribute toward the students grade/score for their final examination paper. I initially trained a back-propagation trained feed-forward neural network to predict the score of their final examination. However, there are a large amount of input features from the questionnaire dataset. We do not know which features are the important features that help contribute toward student performance. In this paper, we experimented with evolutionary algorithm and utilized it to find the best set of input features for the performance prediction. We successfully produced a result that is better than the pure native neural network

Keywords: neuron network · performance predication · features selection · evolutionary algorithm

1 Introduction

The goal for this experiment is to use evolutionary algorithm to improve the input features for a neural network that is uses to predict the score of the student final grade. There have been many utilizations of the genetic algorithm to help further the architecture neural network. This can be the form of the topology or hyper parameters [3]. There has been a case where evolutionary algorithm has been used for features selection in the dataset [4]. In this paper we focus on combining these two approaches and determine the set of input features that highly influence the accuracy of the neural network.

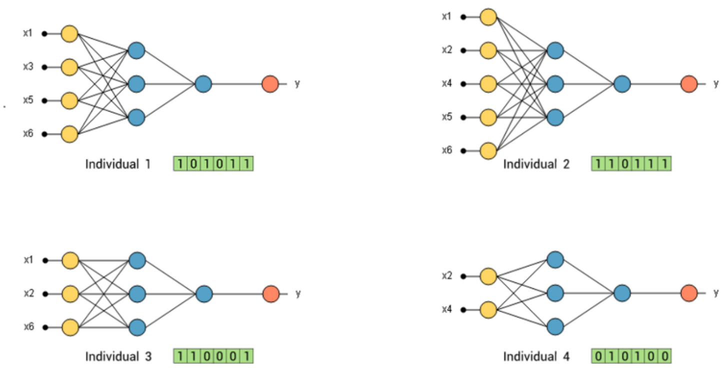


Figure 1: 4 neural networks with different input features [5]

The dataset contained 32 possible input features, we will fine-tune these inputs so we We are hoping to determine a set of input features from the dataset to that will provide the highest influence on their education performance within each student.

2 Method

Since this experiment focuses on utilizing evolutionary algorithm for features selection. Every hyper-parameters of the neural network except the input neuron were kept at constantly throughout the experiment. Throughout the experiment, the hyperparameters of the evolutionary algorithm are being tested at different value.

2.1 Dataset Features and Data Explanation

This dataset was chosen because it provides a large amount of supplementary information regarding each individual student so that we can use these external factors as features for the input neuron to help with the mark prediction.

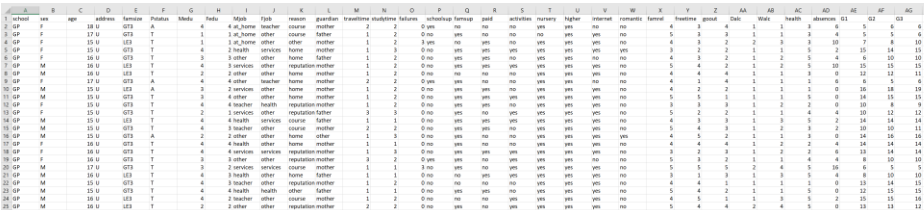


Figure 2: Example of Raw Data

The model is based of the three data mining goals [1] which was the data was previously used.

1. Binary classification (pass/fail)
2. Classification with five levels (from 1 - very good to 5 - poor)
3. Regression, with numeric output that range between 0 (0%) and 20 (100%)

I focused on the third model, attempting to predict the numeric output using neural network. This allows us to have a stronger understanding of the correlation between the features and output. This will let us to identify which trait will produce a more educated student. For the other two method, there are too much variance within the output. For example, in binary classification we will not be able to tell the difference between student who pass the course (output = 10) or student who received high distinction (output > 16) as there maybe too many factors that change between these two students.

Before I started training, the data required to undergo pre-processing method so that the neural network can take in some nominal data. For example, student address was changed from Urban and Rural into 0 and 1. This was done using preprocessing method from sklearn library.

2.2 Neural Network

The neural network that was used in this experiment was a fully connected feed-forward neural network. The training was done via backpropagation. Its architecture contained 50 hidden neurons with 1 hidden layer. Each of the input neurons represent each feature from the dataset. The output neuron represents each potential score the student can receive as their final grade mark. We used Cross-Entropy Loss and SGD as the loss function and optimizer respectively with the learning rate of 0.1. Finally, sigmoid was used as activation function and the number of epochs were 500.

2.3 Evolutionary Algorithm and Parameters Tuning

The 4 parameters from evolutionary algorithm that we are tuning in this experiment.

- Population size
- Crossover rate
- Mutation rate
- Number of generations

We used 5 different valued for the population size and number of generations. They are 10, 20, 30, 40, 50. We capped these values at 50 because higher value will require more computation power to achieve the task within sufficient amount of time. I approached these two parameters with a 1:1 ratios to reduce the computation time, so population size of 10 will have 10 generations

I varied the crossover rate and kept mutation rate the same. This is because the function I used to initialize the population is

```
In [174]: # initialise population DNA
pop   np.random.randint(0, 2, (POP_SIZE, DNA_SIZE))#.repeat(POP_SIZE, axis-0)
```

This is will allow the population to all be generate randomly and different from each other population. So, the effect of the input features and overall accuracy of the neural network can be easily determine using crossover rate.

3 Result and Discussion

3.1 Result from Neural Network without Evolutionary Algorithm

The result that was achieve is currently sitting at an average of 45% accuracy. From a high-level stand point this result is extremely poor. However, there is some positive aspect with in the predicted output. Upon further inspection I discovered that both the mean and the variance of predicted values and the actual testing values are very similar to each other, only differ by one in most cases. I decided to view each individual score and found out that the scores that are not the same as predicted only differ in most case by 1 or 2, i.e. if the actual score is 13, the predicted score would be 12. If we look at the score in this perspective the program is fairly accurate. So, under binary classification and 5 level classification our accuracy result will improve significantly.

In my other paper, I listed the top 5 factors that may the neural network determined to be important.

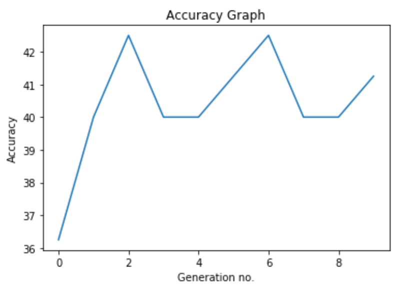
1. G2 – Second period grade
2. G1 – first period grade
3. Mother education
4. Mother occupation
5. Family size

In this result, the neural network priorities past academic performance as the strongest factor for the predicted. The other factors can be draw from another conclusion and dataset. For example, student whose mother education is stronger tend to do well may stemmed from the fact that she can provide academic support to the student who is more likely to approach her over his/her father.

3.2 Result from Neural Network with Evolutionary Algorithm for Features Selection

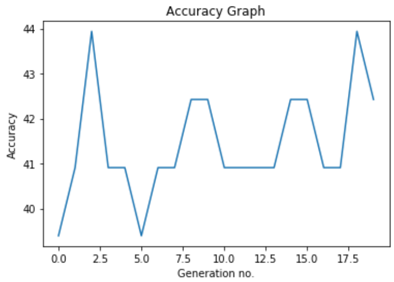
The most fitted DNA will represent the input features from the questionnaires as a binary representation, where 0 is the input that has been disabled and 1 is the input that was pass into the neural network.

3.2.1 Change in population size and number of generations



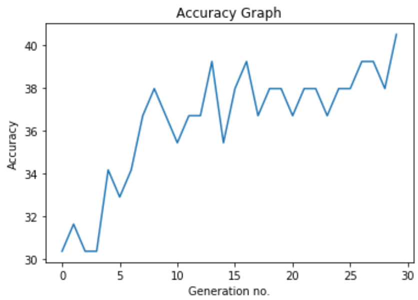
Most fitted DNA: [1 1 1 0 0 1 1 0 0 0 0 0 1 1 1 1 1 1 1 1
0 1 1 0 1 1 0 1 0 1 0 1]
Highest Accuracy: 41.25

Figure 3: Population Size and Number of Generations 10



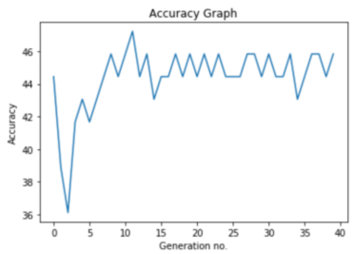
Most fitted DNA: [1 1 1 1 1 0 0 0 1 0 1 0 1 0 1 1 0 0 0 0
1 1 1 0 0 1 0 0 0 0 0 1]
Highest Accuracy: 42.42424242424242

Figure 4: Population Size and Number of Generations 20



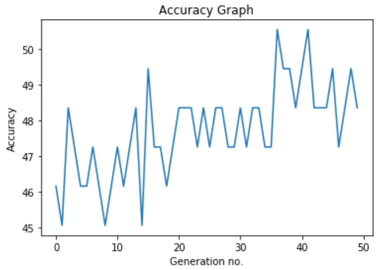
Most fitted DNA: [1 0 1 1 1 1 0 1 1 0 1 1 1 0 0 0 0 1 1 1
0 1 1 0 0 1 1 1 1 1 1 1]
Highest Accuracy: 40.50632911392405

Figure 5: Population Size and Number of Generations 30



Most fitted DNA: [0 1 1 1 0 0 0 1 1 1 0 1 0 0 1 1 0 1 0 0
0 1 0 1 1 0 1 1 1 1 1 1]
Highest Accuracy: 45.833333333333336

Figure 6: Population Size and Number of Generations 40



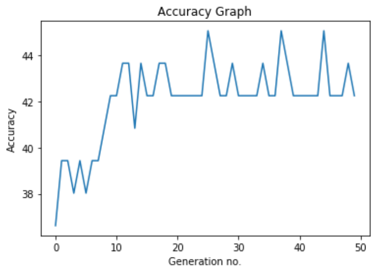
Most fitted DNA: [0 0 1 0 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 1
0 0 0 1 0 0 1 1 1 1 1 1]
Highest Accuracy: 49.45054945054945

Figure 7: Population Size and Number of Generations 50

From the result, we can see that as we increased the number of generations and population size, the accuracy slowing increasing as well. The result from figures 6 showed that the results from figures 3, 4, 5 has not converge. Figure 6 on the other hand appear to be consistent within the later generation. However, as we experiment for population size and number of generations of 50, we discovered that there is still an increase in accuracy. This can mean that population and generation 40 may stumble into local minima and stuck there.

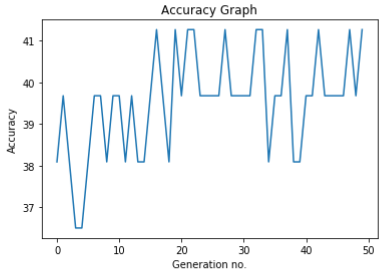
3.2.2 Change in Crossover Rate

I picked 3 different crossover rates for this experiment. The default rate is 0.8, so I decided to perform additional experiments with the rate of 0.2 and 0.5. The result frim 0.8 can be observed at Figure 7 from the previous section. This is to see how low, medium, and large crossover rate impact the neural network. The other parameters for this part are set to default, with the population size and number of generation is set to 50, and mutation rate at 0.002.



Most fitted DNA: [0 1 1 1 1 0 1 0 0 0 1 0 0 0 0 0 1 0 1 1
1 1 1 0 1 0 0 1 1 0 1 1]
Highest Accuracy: 42.25352112676056

Figure 8: Crossover Rate of 0.2




```
Most fitted DNA: [0 0 1 1 0 1 0 1 0 1 0 0 1 0 1 1 0 0 0 1
1 0 0 1 1 1 0 0 1 1 0 1]
Highest Accuracy:  41.26984126984127
```

Figure 9: Crossover Rate of 0.5

The result from crossover rate showed that a large crossover rate will produced a better result, evidently from the Figure 7 with the accuracy of 49%. This could mean because the amount of generation we ran is rather limited and thus a large crossover rate will be more efficiency as it prioritises exploration space.

4 Conclusion/Future Work

4.1 Future Work

As in currently stand, the evolutionary algorithm we used in this experiment is a vanilla native evolutionary algorithm. There are other several evolutionary algorithms that can be utilized to increase the calculation of the algorithm. In term of computational performance, we could also improve the training process of both neural network and evolutionary algorithm, allowing us to further increase the number of generations and population size for more accurate result.

Finally, the number of data can be larger to help further train the neural network, as it currently stands the amount of data is only sufficient. Further information and different questionnaire could also provide a much more robust system that could give a better concrete conclusion. We could also remove other academic related data so any output we get it will be solely based on the external factors which can provide us with a better descriptive conclusion.

4.2 Conclusion

Finally, with the introduction of evolutionary algorithm we can also eliminate some of the data that seem irrelevant to the topic. For example, we can remove sex entirely as its contribution is not important comparing to the other factors. This will allow us to speed up the computation time for the neural network, as we significantly reduce the size of the dataset.

5 Reference

1. P. Cortez and A. Silva. Using Data Mining to Predict Secondary School Student Performance. In A. Brito and J. Teixeira Eds., Proceedings of 5th

FUture BUsiness TEChnology Conference (FUBUTEC 2008) pp. 5-12, Porto, Portugal, April, 2008, EUROSIS, ISBN 978-9077381-39-7

2. T. D. Gedeon and H. S. Turner. Explaining student grades predicted by a neural network., Proceedings of 1993 International Joint Conference on Neural Networks, 1993
3. William T. Kearney. 2016. *Using Genetic Algorithms to Evolve Artificial Neural Networks*. [ONLINE] Available at: https://digitalcommons.colby.edu/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1836&context=honorstheses. [Accessed 28 May 2018].
4. Nadia Abd-Alsabour. 2014. *A Review on Evolutionary Feature Selection*. [ONLINE] Available at: https://pdfs.semanticscholar.org/ec85/bee7c12e206a423886d4d7915af440aff410.pdf. [Accessed 27 May 2018]
5. Fernando Gómez and Alberto Quesada. 2018. *Genetic algorithms for feature selection in Data Analytics*. [ONLINE] Available at: https://www.neuraldesigner.com/blog/genetic_algorithms_for_feature_selection. [Accessed 27 May 2018].