Which Random Seed Is Best for Random Seeds?

Using artificial neural networks to aid seed banks and maintain plant diversity

Catherine Hooper
School of Computer Science
University of Reading
zz001776@live.reading.ac.uk

ABSTRACT

The aim of this poster is to show how a neural network was trained to categorise wheat seeds into three different varieties, analyse the accuracy of the results, and investigate how this could be used in seed banks.

INTRODUCTION

Artificial neural networks are biologically inspired simulations involving artificial neurons, interconnected to work together and perform a specific task. In this case the task is categorisation. These networks resemble the human brain, as they acquire knowledge through learning. Any knowledge in the network is stored using weights between the neurons, known as synaptic weights and these weights are adjusted as the network learns, in this instance, through supervised learning [1].

The data set used in this experiment was retrieved from a data repository of data for use in neural networks [2]. The inputs used to train the network were area, perimeter, and compactness of the seed, length and width of the kernel, asymmetry coefficient, and length of the kernel groove. All the data applied to this neural network is real, continuous, and was obtained from seeds using soft X-ray. The images from which the data was gathered can be seen in Figure 1. Neural networks are preferred in situations when it is unknown what features can distinguish the seeds.

The output of the neural network in this case is 1, 2, or 3; corresponding to the variety of wheat from which the seed has been taken, Kama, Rosa, and Canadian. The data set provides the input values for 210 samples of wheat seed, with 70 samples of each variety [2]. Of this data, 24 pieces of data were used as the unseen data set, 10 for the validation data, and the remaining 156 samples were used as the training data.

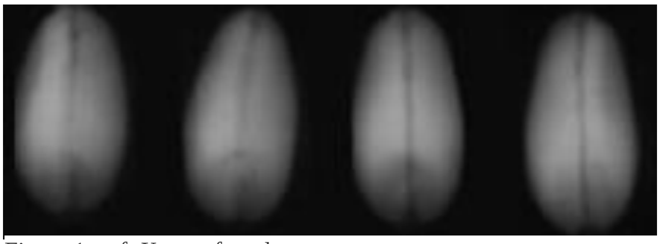

Figure 1: soft X-ray of seeds

TRAINING OF NEURAL NETWORK

To find the best parameters to use on the neural network, the data was presented to the network with different values for learning rate, momentum, number of hidden neurons, and random seed. The learning rate of a network controls the changes of the weights and bias in the algorithm and momentum adds a fraction of the previous weights change to the current one. The number of hidden neurons is the number of neurons that are neither in the input or output layers of the network, and finally, random seed determines the initial weights between the neurons in the network [3].

The parameters which give the lowest sum of squared errors (SSE) are deemed as optimal. Since new data is used for testing the data, a lower SSE should not occur in the case of the model over-fitting the data.

Parameters				% learned	SSE	No. of epochs
a	b	c	d			
0.1	0.6	10	0	69	0.0515	14810
0.1	0.6	10	1000	73	0.1085	1980
0.1	0.6	10	25	73	0.0989	2150
0.1	0.6	15	0	82	0.0694	3010
0.1	0.6	15	1000	86	0.0134	2640
0.1	0.4	15	0	82	0.0732	2990
0.2	0.4	15	0	78	0.0788	2760
0.2	0.4	7	0	91	0.0512	4950
0.2	0.4	7	1000	78	0.1065	35240
0.1	0.4	7	1000	69	0.0563	39810

Table 1: a = learning rate, b = momentum, c = number of hidden neurons, d = random seed.

RESULTS OF NEURAL NETWORK

As can be seen from Table 1, the parameters which produced the lowest SSE, 0.0132, for the unseen data were: learning rate 0.1, momentum 0.6, 15 hidden neurons, and random seed 1000. This produced a percentage learned of 86%, which is not the highest percentage classified out of these results.

DISCUSSION

Overall, the neural network used in this experiment was successfully able to classify the majority of data in the unseen data set, even with the small amount of data provided. However, it can be assumed that if there was more data on the seeds in the training set, the network would be even more successful in learning and classifying the data. Also, if a larger range of parameter combinations are tested, then the neural network is likely to be more successful in learning the data.

In the future, this could be used to classify a wider range of seed varieties. This could be useful for seed banks as the faster and more accurate categorisation of these seeds will help in determining the different seeds in the bank.

This could also be developed to help the environment as a whole. These seed banks form repositories of genetic variation, which can be used as a source of genes for improving crops and preserving, if not enhancing, crop diversity [4]. However, neural network technologies can be applied to other areas, such as modelling the effects of environmental variables on living species.

REFERENCES

[1] C. C. Aggarwal, in Neural Networks and Deep Learning: A Textbook, Springer, 2018.

[2] Systems Research Institute, "Seeds Data Set," Cracow, 2010.

[3] Royal Astronomical Society, "Estimating photometric redshifts with artificial neural networks," Oxford Academic, 2003.

[4] S. D. Tanksley and S. R. McCouch, "Seed Banks and Molecular Maps: Unlocking Genetic Potential from the Wild," Science, 1997.