Compare the effect of Genetic Algorithm and Functional Measures on feature selection and verify the role of feature selection

Jiawei Du

Research School of Computer Science,
The Australian National University,
Canberra ACT 2601
U7100757@anu.edu.au

Abstract. Data sets may often contain many redundant inputs that are irrelevant or highly similar. These factors will lead to a significant increase in the training time and computational cost of the neural network. The use of feature selection can solve these problems. In this article, we still use the Anger data set to conduct experiments, and use two methods of function measures and genetic algorithm to perform feature selection on the data set. After experiments, we compared the results obtained by using genetic algorithms and functional measures methods, and pruned the unnecessary input features according to the obtained results and retrained the network. Finally, we found that genetic algorithm performs better than functional measures method in feature selection. At the same time, we also confirmed that feature selection on the input can not only effectively improve the accuracy of the classification task, but also reduce the network calculation time.

Keywords: Genetic Algorithm, Functional Measures, Feature selection, prune input

1 Introduction

1.1 Motivation

People nowadays often deliberately hide their emotions. When we are in contact with others, we can only get along with others when we know the true emotions of the other person more truthfully and clearly. Therefore, we hope that we can use some data to judge the authenticity of other's emotions and also this dataset is very precious. For this reason, I chose the anger data set. Similar to the definition of smile in [3], [2] defines anger that is performed or disguised as fake anger, and anger from the heart is real anger.

1.2 Neural Network

We hope to use a neural network to process this data and make predictions on it. The network used in this paper is a two-layer fully connected network with a topology of 6-8-2 and this network is similar to the network used in [2] and [3].

Fig. 1. The network topology used in this experiment

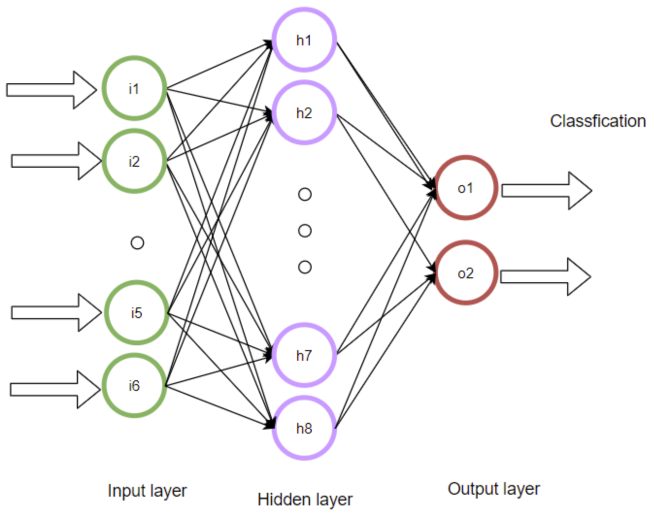


Figure 1 shows the structure of the network in our experiment. Because we judge the authenticity of anger is a two-category classification problem, so the activation function in this network is the sigmoid function. The loss function here is Cross entropy function and the optimizer we used is Adam because it has good performance. Then the error backpropagation method is used to update the parameters. The learning rate here we set is 0.01 and the number of epochs for the network before pruning the input is 8000 and 6000 for the after. There are no horizontal, backward, and multi-layer connections in the network. In addition, due to the dataset only contains 400 rows of data, it is relatively a small dataset. So, our batch size is all the data.

The network uses a random 80% anger data set for training and 20% anger data set for testing. Because we do not use the validation set to design the topology, nor used to select the best from a set of networks, we do not need a validation set. The pupil response data of the anger dataset is used as input, and the output is a prediction of whether the pupil is facing real or false anger. The accuracy rate is obtained by comparing with the label value in the data set. In addition, in order to prove the performance of our network, we used the accuracy of the test set and the confusion matrix to make judgments. With this dataset and network, we can analyze the authenticity of anger from these data.

1.3 Features Selection and Motivation

When we deal with a new data set, these data sets usually contain a large number of input features, many of which are unimportant and redundant. When we put the input features of them into the neural network for training without processing them, these unnecessary features may cause some noise and interference. Moreover, f those unimportant features are also input into the neural network, it will increase the calculation time and computational cost. Therefore, we may not get good results and waste computational cost. From [6], we know that feature selection is a problem of selecting a small part of features, which are necessary and sufficient to describe the target concept under ideal circumstances. Therefore, feature selection can be used to solve the problem of poor results due to redundant features.

Thus, in order to make the network have a good result and reduce the computational cost, we use feature selection to select the vital features. In this report, we use two different methods to do the feature selection. One is the functional measures method mentioned in [1] to analyze the importance of the input to the network and then to choose the significant features. This method was proposed by Gedeon in 1996. Second is the genetic algorithm which is a very popular feature selection algorithm.

We know that genetic algorithm is one of the most advanced feature selection algorithms, so we want to compare the performance of these two methods. After we use above two methods, we use the result and prune the unimportant inputs and then retrain the network. Finally, we compare the accuracy obtained after using these two methods and verified our assumptions. The following second and third parts introduce the methods used and the comparison results in detail.

2 Method

2.1 Dataset

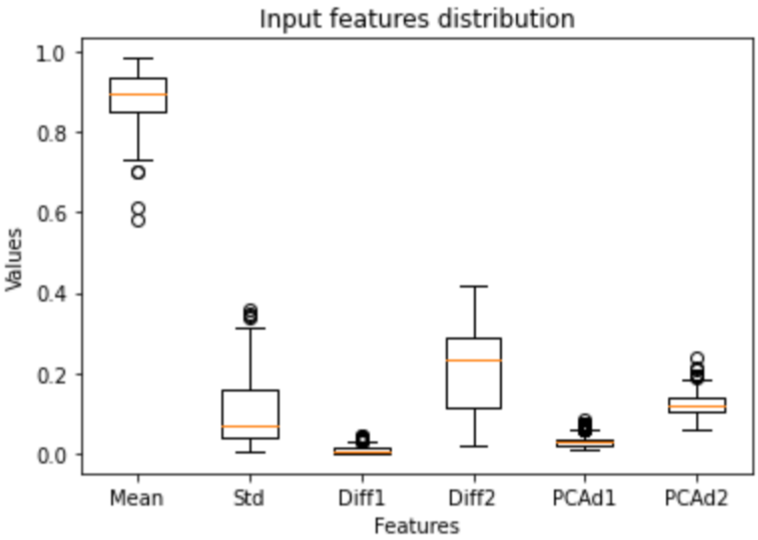
This data set contains some data related to pupil changes when people face anger and labels of real or false anger emotions. These data were obtained by the experimenter in [2] by measuring the pupil response of experimental participants facing acted and true anger. Since we are predicting the authenticity of anger, we do not need a video encoding column, so we delete this column.

Table 1. The explanation of the dataset features

Input features	Explanation
Mean	The average value of pupil response
Std	The standard deviation of pupil response.
Diff1	The change in the size of the left pupil after watching the video.
Diff2	The change in the size of the right pupil after watching the video.
PCAd1	The orthogonal linear transformation with the first principal component
PCAd2	The orthogonal linear transformation with second principal component

This dataset contains 400 rows of data and 6 features of the pupil actions. The features are shown in the Table 1.

Fig. 2. The distribution of input features



In addition, from the Figure 2, we can find that the data distribution intervals of these six features are quite different. Input 1 (Mean) has the highest average value, while inputs 3 (Diff1) and input 5 (PCAd1) have lower average values. Because the input distribution of the data set is unbalanced, so we use the Z-score standardization method to normalize the data except for the label column. The label column of this data set has two different labels, which shows that our prediction of the authenticity of anger is a two-classification problem and then we use 0, 1 to replace the data in the label column. 0 represents faker and 1 represents true.

2.2 Genetic Algorithm

Genetic Algorithm (GA) is a computational model that simulates the natural selection and genetic mechanism of Darwin's biological evolution theory. The algorithm searches for the optimal solution by simulating the natural evolution process.

Genetic algorithm can be divided into five stages. The first stage is to generate the initial population. An individual is characterized by a set of parameters (variables) known as Genes. Genes are joined into a string to form a Chromosome (solution). In the genetic algorithm, each individual is a solution to the problem we seek. In our experiment, we use binary encoding. The selected input feature is coded as 1, and the unselected input feature is coded as 0. Besides, because we have 6 input features, it has 32 combinations, so we set the population size to 32.

The second stage is the fitness function. The fitness function is to calculate the fitness score of each individual. The higher the score, the more adapted the individual is to the environment. The probability of an individual being selected for reproduction is based on its fitness score. For our experiment, we select the characteristics that each individual represents in turn. Then put the number of input neurons and the number of selected features to be consistent, and the hidden layer and output layer topology remain unchanged for training. The fitness score of each individual is the accuracy of the final test set. We calculate the fitness function based on the following formula (1).

$$Fitness(x) = Accuracy(x). \tag{1}$$

X is each individual of the population. When the accuracy is higher, the fitness is higher.

The third stage is natural selection. The purpose of this stage is to select individuals with the highest fitness scores as parents and pass on their genotypes to the next generation. Of course, the higher the fitness, the higher the probability of being selected as a parent.

The fourth stage is crossover. At this stage, we randomly select a point in the parents' genes as the crossover point, and then exchange the genotypes of the parents. In this way we have produced a new offspring.

The fifth stage is mutation. For our offspring, there is a certain probability that genetic mutations will occur. This means that the value of one of their genes will be reversed.

Table 2. The Hyperparameters for the Genetic Algorithm

Parameters_Name	Value	Explanation
DNA_SIZE	6	Equal to the number of input features
POP_SIZE	32	The population size
CROSS_RATE	0.8	Probability of crossover
MUTATION_RATE	0.002	Probability of mutation
N_GENERATIONS	25	Number of interation (Generation)

Table 2 shows the hyperparameter settings of the genetic algorithm in our experiment.

2.3 Funtional Measures

The functional measures technique proposed in Gedeon's work [1] allows us to analyze the importance of neural network inputs. We first use the previously mentioned two-layer fully connected network to train the data. After training, we use the weight vector input to the hidden layer neuron to represent each input, that is, each column of the hidden layer weight matrix represents an input.

From the [4] we know if the angle between the two vectors (vectors here represents different hidden neurons) approaches 0 degrees, then the two inputs are very similar. If the angle between the two vectors approaches 180 degrees, then the two vectors are complementary. If the angle between the vectors approaches 90 degrees, it indicates that the two vectors are not similar. Therefore, [1] propose a new technique that judge the difference of different inputs by comparing the angles between the weight vectors. Hence, if the angle between the weight vector (representing one input) and other weight vector is very close to 90 degrees, then this input is irreplaceable. Conversely, if the included angle approaches 0 degrees or 180 degrees, then the input is not so important.

In our experiment, we define a judgment range, the range is 0-30 degrees and 150-180 degrees. That is, if the angle between the vectors is around 0-30 degrees, then our two vectors are similar. The angle between the vectors is around 150-180 degrees, so our two vectors are complementary. If the angle between an input and other inputs often appears within this range, then we consider that input to be the least important input. On the contrary, it is the most important input. We calculate the angle using the formula (2).

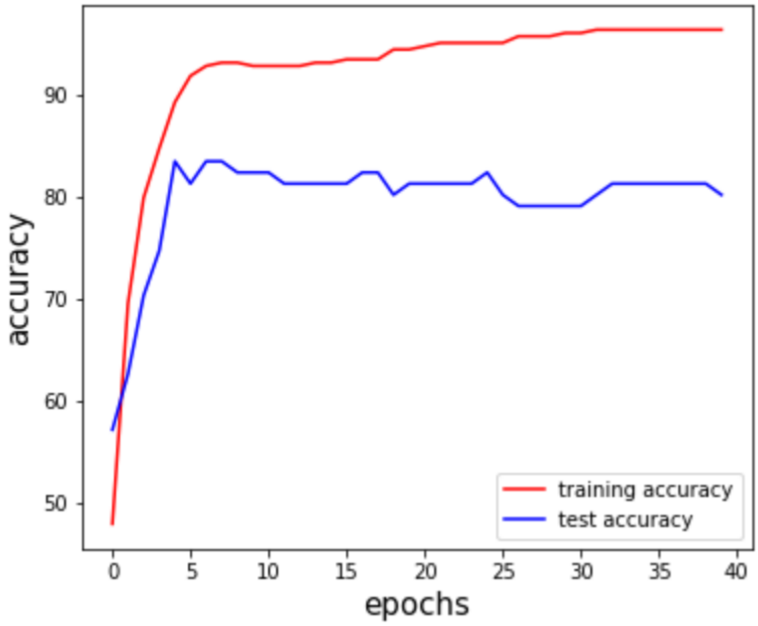
$$cos\theta = \frac{AB}{\|A\|\|B\|} . \tag{2}$$

After we pruning the least important input, the rest of inputs are we selected. So far, we have used functional measures method to finish the feature selection.

3 Result and Discussion

3.1 Original input network

Fig. 1. The accuracy of the original network training set and test set



The data shown below all represent average data. From the Figure 3 we can find that our prediction of the authenticity of anger can reach more than 95% on the training set and about 82% on the test set. Compared with the results in [2], our accuracy rate is lower. In [2], their accuracy rate can reach about 95%, while our accuracy is only about 82%. This shows that the performance of our network is not as good as theirs. However, this will not affect our experiment.

3.2 Features selected by Genetic Algorithm

We first use genetic algorithm to do the feature selection. In the experiment, we found that the algorithm had already converged in the 9th generation. The final selection results of the genetic algorithm show below.

Table 3. The features input selected by Genetic Algorithm

Selected Chromosome type	Corresponding select input features
[0 0 0 0 1 1]	[PCAd1, PCAd2]

From the Table 3, we find that the genetic algorithm selected two features, namely PCAd1 and PCAd2. From math perspectival we know that PCA uses orthogonal transformation to transform a series of potentially linearly related variables into a set of linearly unrelated new variables and these new variables contain most of the features of the previous data. From the description of the dataset, we know that PCAd1 and PCAd2 are the new features obtained by using PCA. Therefore, PCAd1and PCAd2 are the two most important input features and that matches the features selected by the genetic algorithm.

3.3 Feature selected by Functional Measures

Next, we use functional measures method to implement feature selection. We calculate the angle between each input from the original neural network that has been trained. Because the neural network we used has only one hidden layer, we take out the weight matrix of the hidden layer of the original neural network, and then use each column to represent an input. The angle value between each inputs shows below.

Table 4. Angle between each input from the orginal network

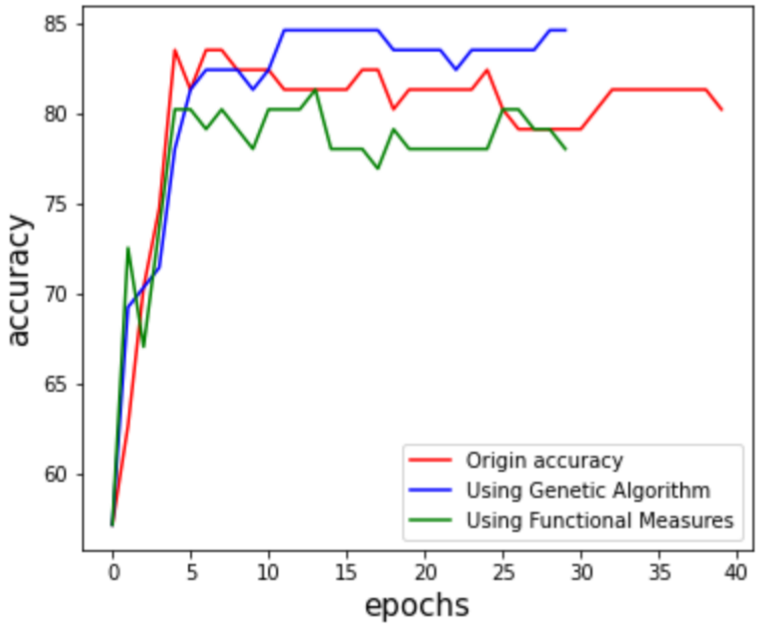
Angle(degree)	Input 1(Mean)	Input 2 (Std)	Input 3 (Diff1)	Input 4 (Diff2)	Input 5 (PCAd1)	Input 6 (PCAd2)
Input 1(Mean)	0.0	94.07	85.81	29.46	119.51	79.69
Input 2 (Std)	94.07	0.0	114.21	80.50	75.88	62.87
Input 3 (Diff1)	85.81	114.21	0.0	113.28	99.17	100.42
Input 4 (Diff2)	29.46	80.50	113.28	0.0	110.60	74.62
Input 5 (PCAd1)	119.51	75.88	99.17	110.60	0.0	45.68
Input 6 (PCAd2)	79.69	62.87	100.42	74.62	45.68	0.0

From the table 4, we can find that input 1(Mean) and input 4 (Diff2) has the smallest angle (29.46). This shows that input 1 and input 4 are relatively similar, so we choose to prune input 4 (Diff2). From the dataset, we know that input 4 is the change in the size of the right pupil. This is probably because the change in the size of the right pupil is not very obvious, and its contribution to distinguishing the true and false of anger is not as good as other features.

3.4 Re-train and Compare the performance

According to the results obtained by using genetic algorithms and functional measures methods, we retrain the network, and the results obtained are as follows.

Fig. 2. The final accuracy obtained by different selected features



We can clearly find from Figure 4 that after using the genetic algorithm for feature selection, the highest accuracy rate (84%) was obtained in the classification task, while after using the functional measures method for feature selection, the accuracy rate was only 78%. This shows that after using the functional measures method for feature selection, there are still unimportant features in the selected features, and the noise is not completely eliminated. This is because when using functional measures for feature selection, we only deleted the least important features, and did not prune the less important features. Because the angle between these less important features does not satisfy our condition for trimming them. Therefore, the effect of using functional measures for feature selection is not as good as using genetic algorithms. The accuracy of feature selection using functional measurement methods is lower than that of the original network, which also shows that using genetic algorithms for feature selection is a better way. This is also confirmed by the results in Figure 4.

Furthermore, we also tested the running time before and after deleting the input. We let the network keep the same number of epochs and train separately in these two cases. The results are shown below.

Table 5. The time costed for network calculation before and after feature selection

Epochs\ Cases	Before prune input (second)	After prune input (second)
100,000	69.45	69.06
200,000	150.77	143.73
500,000	403.97	381.71

From Table 5, We found that after using feature selection, the time it takes to run the network is reduced. Although in our experiments, the time reduction is not very obvious, because our data set is relatively small, and the network is also a very simple two-layer neural network. But for those cases where complex and deep neural networks are used to process huge data sets, this time reduction is very obvious and very important.

4 Conclusion and Future work

Our experiment is to use the anger data set to judge the authenticity of anger. We also compare the performance of genetic algorithm and functional measures in feature selection and the impact of feature selection on classification tasks. Through experiments, we find that using the input features selected by the genetic algorithm, we can obtain a higher accuracy rate than using the functional measures method. The reason is because the functional measures method does not make a thorough selection of the input, but only a part of it. Therefore, the effect of using the function measures method is not as good as using the genetic algorithm for feature selection. In addition, the use of feature selection to process input features reduces the time and cost required for calculations. Although it is not obvious in our experiments, it will be very obvious when the data set is very large and the input features are very large. And feature selection can improve accuracy and reduce computational cost. Finally, based on the above results and analysis, we prove that genetic algorithm performs better than functional measures in feature selection. Besides, we also proved that feature selection can improve the accuracy of classification tasks and obtain better results.

Meanwhile, there are still some limitations and improvement directions in this experiment. First of all, our network performance is not good enough compared to [1]. The reason maybe the hyper-parameter settings of our network are not good enough or the data set we used is not complete compared with [1], and the amount of data and features that can be used are not enough. Moreover, maybe we need to do more processing on the data, such as using L1 regularization and other methods. In the next step, we will re-adjust better parameters, use the complete dataset and perform better preprocessing of the data.

Secondly, the accuracy of our test set fluctuates greatly, ranging from 75% to 90%. This may be because the data distribution of the randomly allocated test set is not very balanced. In the next step, we will impose more restrictions on the operation of random allocation, such as distributing in proportion to the data type of the original data set.

Thirdly, we will compare the genetic algorithm and functional measures on more different neural network models to increase the credibility of our conclusions.

5 Reference

[1] Gedeon, Tamás D. "Data mining of inputs: analysing magnitude and functional measures." *International Journal of Neural Systems* 8.02 (1997): 209-218.

[2] Chen, Lu, et al. "Are you really angry? Detecting emotion veracity as a proposed tool for interaction." *Proceedings of the 29th Australian Conference on Computer-Human Interaction*. 2017.

[3] Hossain, Md Zakir, and Tom Gedeon. "Classifying posed and real smiles from observers' peripheral physiology." *Proceedings of the 11th EAI International Conference on Pervasive Computing Technologies for Healthcare*. 2017.

[4] Gedeon, T. D., and D. Harris. "Network reduction techniques." Proceedings International Conference on Neural Networks Methodologies and Applications. Vol. 1. 1991.

[5] Gedeon, Tamás D. "Indicators of hidden neuron functionality: the weight matrix versus neuron behaviour." *Proceedings 1995 Second New Zealand International Two-Stream Conference on Artificial Neural Networks and Expert Systems*. IEEE, 1995.

[6] Kira, Kenji, and Larry A. Rendell. "The feature selection problem: Traditional methods and a new algorithm." *Aaai*. Vol. 2. 1992.