Predict the degree of depression through three-layer neural network with input features selected by GA and explain the results by causal index and characteristic input pattern

Xuehan Liu
Research School of Computer Science
The Australia National University,Canberra
u7094891@anu.edu.au

Abstract: When people observe patients with depression, there are often some physiological manifestations that are difficult to detect. We can use neural networks to analyze this physiological response and identify other people's depressive symptoms. In this paper, a **three-layer neural network(NN)** is used to predict the depression level of the subjects.To optimize the accuracy of the neural network, the **genetic algorithm(GA)** which is using the accuracy of the NN on the test set as the fitness function,selects the optimal input feature set.In order to better understand and predict the output of the neural network, I uses the rules generated by the gradient of the output with respect to the input called **causal index** and the average value of all inputs features selected by GA that produce the same output as the **characteristic input** to predict the neural network results.Finally,we find that GA algorithm can improve the accuracy of neural network, and the accuracy of neural network interpretation method is higher, reaching 80%.

Keywords: Depression detection, physiological signal, neural network,genetic algorithm,characteristic input, causal index,neural network interpretation methods.

1 Introduction

1.1 Motivation

Depression is an internal mental disorder, different from the general mood swings and short-term emotional reactions to daily events, but a serious chronic health disease.Effective diagnosis and treatment is very important. It is the key to improve personal quality of life, mental health, family happiness and reduce social and economic costs.

At present, the diagnosis methods are mainly through doctor consultation,patients fill in self-report questionnaire and other means. The diagnosis of depression level depends on the experience and personal judgment of clinicians, and the willingness and ability of patients to respond to their symptoms. Therefore, the current means of diagnosis of depression is not objective and reliable, so we need more objective data to support the diagnosis method.

With the development of affective computing technology, it is found that the physiological signals of observers can be used to identify the depression level of others. The observer's eye gaze behavior, skin electrical response and skin temperature can reflect the depression level of the observed.

1.2 Data set

The data set used is the physiological signals generated by 12 participants when watching 16 different videos of patients with depression. During the experiment, the changes of pupillary dilation (PD), galvanic skin response(GSR) and skin temperature (st) of the observers were recorded.[1]

1) Galvanic Skin Response (**GSR**): GSR, also known as skin conductance (SC) or electrodermal activity responses (EDA), measures an individual's electricity flow through the skin, which varies due to the amount of sweat on the skin . The GSR is composed of two separate electrodermal activities.

2) Skin Temperature (**ST**): ST fluctuates due to vasodilatation of peripheral blood vessels induced by increased activity of the sympathetic nervous system. It hasbeen found to be negatively correlated with unpleasant emotions such as stress and fear because blood is redirected to vital organs as protection measure.

3) Pupillary Dilation (**PD**): PD provides indications of changes in mental states and the strengths of mental activities. Pupil size was found to constitute a response to emotionally engaging stimuli where pupil is significantly bigger after positively and negatively arousing stimuli than after neutral stimuli.

The corresponding depression level of each patient was generated according to the Beck Depression scale, which is a general method to estimate the depression level currently.Physiological signals are directly recorded by the monitoring equipment, which is not subject to subjective influence, and can provide an objective basis for training to predict the level of depression.

1.3 Neural network model introduction

The main problem to be solved in this paper is to judge the depression level of the patients through the physiological data of volunteers' when observed the videos of the patients with different level of depression. Taking selected physiological data which was extracted by data pre-processing as input, with features such as mean value, maximum value of data, and value obtained by filter, which does not contain time series information.**Three-layer neural network(NN)** is trained with the input to predict patients' depression level.

The fully connected neural network uses a simple idea of **back-propagation**, and the nodes of each layer are related to each other. This kind of network structure is stable and widely used.We do not use the raw data which has the information of time series.For the physiological signals we selected, their continuity is not what our focus. We pay more attention to the features like peak and average of each data, so the RNN network and LSTM network considering context are not our best choice. The fully connected neural network can solve the multi parameter input and multi classification problem, so in this paper, we choose to use three-layer fully connected neural network.

1.4 Genetic algorithm(GA) introduction

Genetic algorithm is a kind of evolutionary algorithm, which is used to solve the optimization problem in computational mathematics. It is based on some of the phenomena in biochemistry, such as heredity, mutation, natural selection and hybridization.

Genetic algorithm is an evolutionary theory that simulates biological species. Species develop in the direction of their own benefit, which is shown in the direction of optimization in genetic algorithm. In the process of evolution, genetic algorithm simulates the behavior of genes. Firstly, it selects the dominant genes, matches the genes, exchanges the alleles, and mutates the genes with a certain probability, which leads to the generation of next generation genes and new individuals.A complete genetic algorithm mainly includes several steps: gene coding, population initialization, selection operation, crossover operation, mutation operation, end condition judgment and so on.

Because the input features of the original data set in this paper have 85 dimensions, some of them will affect the training effect of neural network. Therefore, GA is used to **select input features**. In this paper, the gene in GA is a **combination of input features**, which is represented by binary coding of each feature. If the code of a position is 1, the feature is selected; otherwise, it is 0.The binary sequence generated immediately is used to initialize the population. Because our goal is to find the best combination of input features to make the model perform well, the fitness function of GA algorithm is defined as the **accuracy of three-layer NN** on the test set. The end of GA algorithm is to achieve a fixed number of **iterations**.

1.5 Decision tree and feature input

We can train the neural network to get the model that meets the requirements to solve the specific prediction or classification problems. However, the work of neural network is still a black box model for us, and we can not intuitively understand how neural network predicts. To explain neural networks, it is necessary to find a set of rules to express the learning process of neural networks.

The most simple and intuitive way is to prune the neural network to the minimum, which is easy to extract the symbolic pattern inside the neural network, but it will reduce the stability and adaptability of the network. Another method is to extract the causal index of input relative to output, which means that the pattern of each class is created as the feature pattern of the class, and the feature pattern is extracted as the rule interpretation neural network.

In this study, decision tree is used to learn the training data. Decision tree is a prediction model, which represents a mapping relationship between object attributes and object values. Data in each branch is divided into different categories according to different conditions. Through the rules generated by decision tree, we can explain the judgment process of neural network. In addition, we can use the average value of all the data whose neural network prediction result is on as the feature mode, and select the closest on mode according to the distance between the input data and different feature modes, which is judged as this kind of data, so as to explain the process of neural network prediction.

2 Method

2.1 Data preprocessing

The data set used in this paper is the preprocessed data of the original data set. The original data is to record the physiological condition of the observer during watching the video. Among them, GSR, St are sampled at 4Hz frequency, PD are sampled at 60Hz frequency.The preprocessing of the original data includes filling the missing pupil data due to blinking. In order to reduce the difference between individuals and the numerical difference of different data, we standardize and normalize the data of different observers.

Then, digital features such as maximum, minimum and average are extracted from the three physiological data, and different filters are used to filter the data. Finally, 23 GSR features, 39 PD features, and 23 ST features were obtained[1].The characteristics of each row of input data are plotted with scatter diagram. It is found that the difference between different types of data is small, and the training result of the model may not be good.

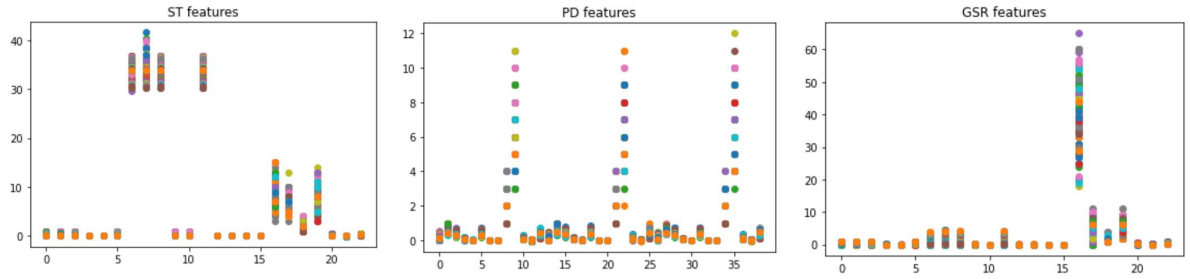


Fig1 scatter plot of input features

After reading the extracted data, the data is divided into training set and test set according to the ratio of 6:4. The feature is of *float* type and the label is of *long* type, which is convenient for NN input.

2.2 Classification model based on Neural Network(NN)

This paper uses three-layer neural network, including input layer, hidden layer and output layer. The number of neurons in the input layer is set as the characteristic number in the data. The hidden layer is composed of 20 neurons. I tested the number of different neurons in the range of 5-100, and **20** neurons performed better. The output layer consists of four neurons, corresponding to the classification of four depression degrees. In the input and hidden layer, sigmoid is used as the activation function, which is a nonlinear activation function and can map data between (0,1). Using softmax function in the output layer is more suitable for the output of multi classification problems(Fig 1). the training times of the model is set to **2000** times.

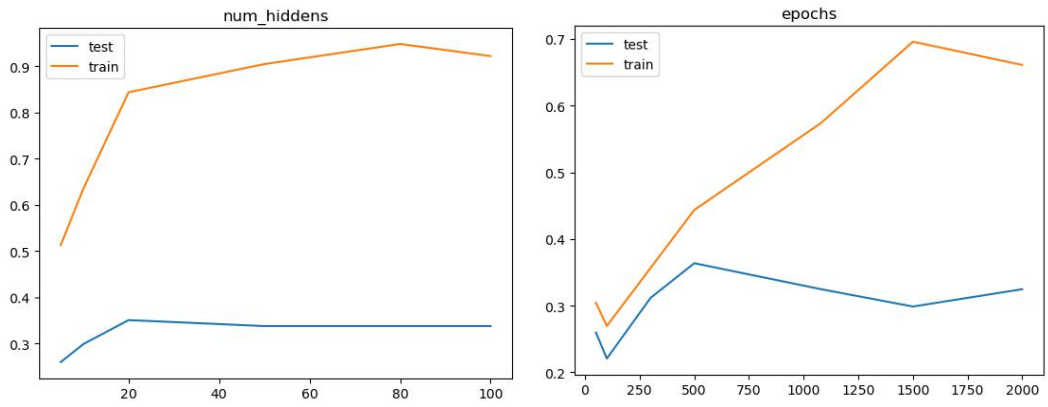


Fig2 accuracy for different parameters

The performance of the model on the dataset is not good. In order to improve the model, I tested the selection of a variety of optimizers and activation functions. Finally, the model performs better when the optimizer uses **Adam** and the loss function **cross entropy**. At the same time, in order to avoid over fitting, a dropout of 0.2 is added to test the accuracy of testing.

Table1 accuracy for optims and loss

	Adam	SGD
Cross_Entropy_Loss	0.377	0.259
NLL_Loss	0.234	0.247

For the results of neural network, this paper does not set cross validation and other verification methods. The main reason is that the results of neural network are mainly used to provide the fitness of GA. for each generation containing 100 combinations, the deviation of the model can be ignored because there are enough test data. Moreover, it can greatly reduce the running time of GA and accelerate the convergence.

2.3 Genetic algorithm

In this paper, the purpose of GA is to find the best combination of input features for neural network accuracy. The fitness function returns a single value, which belongs to the single objective optimization problem. When creating the single objective optimization problem, weights is used to indicate maximization and minimization. Our purpose is to select the maximum value, so weight is set to 1.0

We use binary coding for different feature combinations. In binary coding, two kinds of bases in human chromosome are simulated by 01, and a certain length of 01 string is used to describe variables. Because there is no built-in *binary encoding* in *deap* library, we use Bernoulli distribution in *SciPy* module to generate a binary sequence. Then we use *population* to initialize a common population, which has no special order or sub population.

In genetic algorithm, the fitness of individuals is used to evaluate the quality of each individual, so as to determine the size of its genetic opportunity. In this example, the objective function is to calculate the accuracy of the three-layer neural network, and the maximum value of the function is taken as the optimization objective. The value of the objective function can be directly used as the individual fitness.

Selection operation (or replication operation) inherits the individuals with higher fitness in the current population to the next generation according to certain rules or models. Generally, individuals with higher fitness will have more chances to pass on to the next generation. There are two proper methods for this paper.

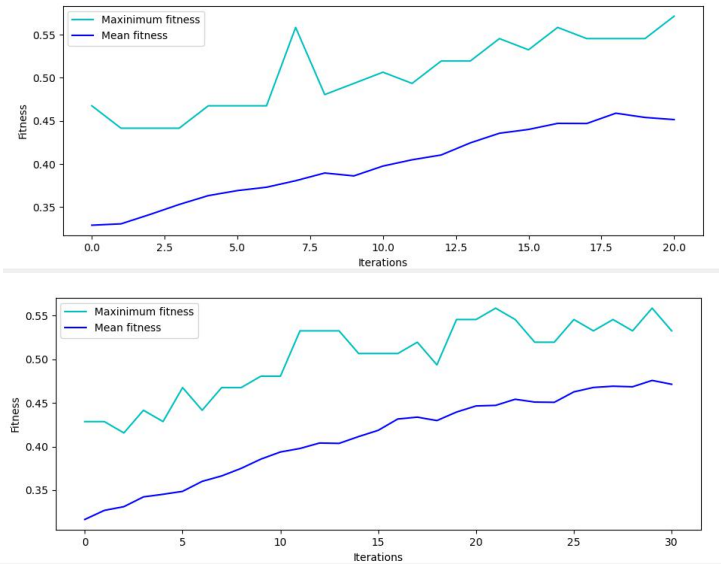
- Tournament selection is a way to simulate tournament. Firstly, tournsize individuals are randomly selected from the population, and then the individuals with the best fitness are selected from them. This process is repeated K times to obtain breeding population. The larger the tournsize, the higher the selection intensity, and the higher the average fitness of the breeding population after selection. The tournsize used in this paper is 2.
- Roulette selection is the most common choice strategy, which can be regarded as random sampling with put back. In roulette selection, the probability of each individual being selected is proportional to its fitness function.

Tournament selection usually has faster convergence speed than roulette selection, so this paper chooses tournament selection method.

Crossover operation is the main operation process of generating new individuals in genetic algorithm. It exchanges some chromosomes between two individuals with a certain probability. In this paper, the method of single point crossover is used, and the specific operation process is to exchange the variables of one of the two genes.

Mutation operation is to calculate the gene value of one or some loci according to a small probability Line change is also an operation method to generate new individuals. I use the basic bit mutation method for mutation operation, because this mutation method is suitable for binary sequences.

A new generation of population P (t+1) can be obtained after a round of selection, crossover and mutation operations on population P (t), and do the cycle until reach the specified number of generations.When selecting generations number, I find that the GA population in this paper can get better results when the generation is 50, but the running time is too long. When the number of generations is 20, the GA population does not converge completely and the optimization result is not good. So we chose 30 generations.



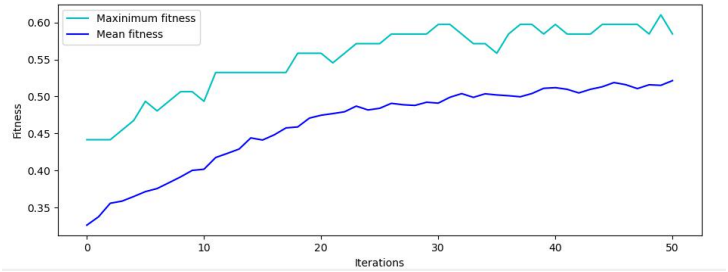


Fig3 fitness of GA with different generations

Table2 implementations settings for GA

GA parameter	**value**
Population size	100
generations	30
Crossover rate	0.8
Mutation rate	0.2(1/(length of chromosome))
Crossover type	single point crossover
Mutation type	bit mutation
Selection type	tournament selection

2.3 Characteristic input and Decision tree

This paper solves the problem of multi classification, so there are four groups of feature inputs. When a group of data is predicted as a label by neural network, the group of data is set as the on output of a label and the off output of other labels.

I use test set to determine the characteristic input and predict the results of the neural network. The specific method is to group the data according to the prediction results of neural network. Those results are 1 are divided into ON group and 0 are into OFF group. Then, the average value of each group of data is used as the characteristic input for the corresponding output. Next, the output of neural network is predicted according to the variance between characteristic input and each input. The output corresponding to the characteristic input which has a smaller variance between each input is used as the prediction of the neural network results of this input.

Then decision tree method is to train the decision tree by using the data predicted on the training set after neural network training, and test the prediction performance of the decision tree and the fitting degree of the neural network on the test set. The decision tree trained by this method fits the decision rules of neural network. We can explain the prediction strategy of neural network by observing the decision process of decision tree.

Each node of the decision tree represents the condition judgment of the current branch, including whether the value of a feature meets the condition and whether the entropy meets the condition. If yes, it will enter the left branch until it reaches the leaf node, and then it will be divided into a certain class.

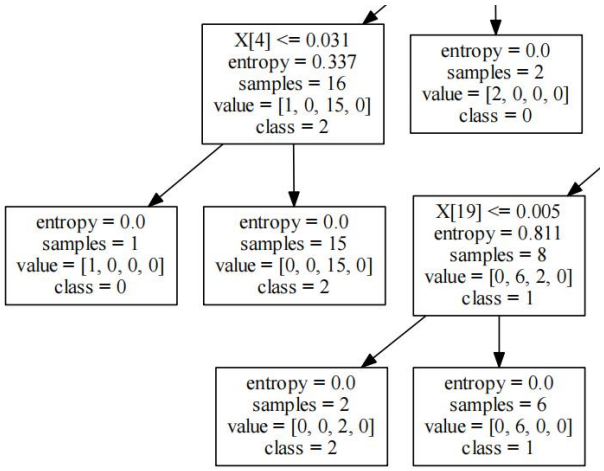


Fig4 Partial graph of decision tree(full graph in appendix)

3 Results and discussion

3.1 Results

NN model and NN + GA model are used to train on all the features. The classification performance was calculated according to the average results of 10 runs. Accuracy, precision, recall and F1 were used as the evaluation criteria.

Table3 result of depression prediction with NN and GA+NN

	NN	**GA+NN**
accuracy	0.38	0.49
precision	0.40	0.49
recall	0.39	0.50
F1	0.37	0.47

The optimal feature combination selected by GA algorithm is used to predict and interpret the neural network with feature pattern and decision tree respectively, and the average result is obtained by running 10 times.

Table4 result of characteristic pattern and decision tree

	characteristic pattern	**decision tree**
Predict accuracy	0.34	0.39
Fitness on NN	0.58	0.44

3.2 Discussion

According to the observation of the results, it can be found that the realization of the three-layer neural network in this data set is not good, and it can not achieve the accuracy mentioned in the paper after repeated debugging. Combined with the analysis of the data set, it is found that the data difference of each category is relatively small, so the neural network can not effectively distinguish. We use GA algorithm to remove some interference features, the accuracy of the model has been significantly improved.

As the parameters and selection, crossover and mutation mode of GA algorithm used in this paper are modified to adapt to the data set, the improvement of the accuracy of the model by GA algorithm in this paper is about 0.1, which is

slightly larger than the growth rate in reference paper[1]. But it still can not reach the ideal level, which also shows that the neural network can not greatly improve the accuracy of neural network, so it is suitable to be used as one of the methods to adjust the neural network.

Because the prediction data of neural network is used to train decision tree and generate characteristic pattern, and the structure of these two methods is simple, the prediction accuracy of decision tree and characteristic pattern in Table 2 is low, about 0.33 and 0.34, which is lower than the prediction result of neural network.

The prediction accuracy of characteristic model for neural network can reach about 0.6, which indicates that this method can successfully predict some outputs of neural network, but the prediction effect is not accurate enough. Only by calculating the distance between feature input pattern and input data to predict the result of neural network, the output of neural network is still a method to be improved. Neural network analysis should use more complex and accurate methods, such as using median, variance as characteristic pattern input, or weighting some key features in prediction and classification to reflect the importance of features.

Decision tree is mainly used to explain the prediction strategy of neural network and extract the rules of neural network judgment. I use the data predicted by the neural network to train the decision tree and make the decision tree fit the neural network. However, because the underlying judgment strategy of the decision tree is not different from that of the neural network, the accuracy of the decision tree interpretation of the neural network is low. But it can be used as a reference to help us understand the work of neural network.

4 Conclusion and future work

In order to solve the problem proposed in this paper, based on the physiological signals of the observer, I use **three-layer fully connected neural network** to classify the depression level, and use **genetic algorithm** to select the optimal combination of input features, which has achieved good results on the existing data sets. Compared with the **decision tree** method and the **characteristic pattern** method, it is found that the characteristic pattern can better predict the neural network, and the decision tree can explain the prediction strategy of the neural network, so that people can better understand the reasons of the neural network decision-making, and adjust the importance of parameters in the diagnosis and decision-making of depression. The results show that for the diagnosis of depression, more complex neural network structure and more accurate interpretation rules are needed, but the current methods have some limitations.

In the future, it can improve the quality of data sets, and it may be more meaningful to refer to the physiological signals of more professionals than ordinary people when collecting data sets. In the aspect of classification model, we can choose to use the original data, such as using time series and **LSTM** model to classify the depression level. In the prediction and interpretation of neural network, more accurate parameters such as **variance**, **median** and causal index can also be used to explain neural network. The technology mentioned in this paper can improve the effective diagnosis of depression, make patients get timely and effective treatment and better quality of life.

Reference

[1] Zhu, X., Gedeon, T., Caldwell, S., & Jones, R. (2019). Detecting emotional reactions to videos of depression. In INES'19: IEEE 23rd International Conference on Intelligent Engineering Systems (6 pp).

[2] Gedeon T. D. and Turner H. S.:Explaining student grades predicted by a neural network. In Proceedings of 1993 International Joint Conference on Neural Networks(1993)

[3] Harry,S. T., Tamás,D.G.: Extracting Meaning from Neural Networks

[4] Zhu, Q. (2020) "Predicted the authenticity of anger through LSTMs and three-layer neural network and explain result by causal index and characteristic input pattern," 3rd ANU Bio-inspired Computing conference (ABCs 2020), paper 83, 7 pages, Canberra.

[5] Tameru Hailesilassie,Rule Extraction Algorithm for Deep Neural Networks: A Review,(IJCSIS) International

[6] N. Cummins, S. Scherer, J. Krajewski, S. Schnieder, J. Epps, and T. F. Quatieri, “A review of depression and suicide risk assessment using speech analysis,” Speech Commun., vol. 71, pp. 10–49, 2015.

[7] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” arXiv Prepr. arXiv1412.6980, 2014.

Appendix

1. Full graph of Decisiontree
2.