Predicted the authenticity of anger through LSTMs and three -layer neural network and explain result by causal index and characteristic input pattern

Quning Zhu
Research School of Computer Science,
The Australian National University,
Canberra ACT 2601
u6693699@anu.edu.au

Abstract. In this paper, **LSTMs** and **three-layer neural network** are trained to predict the authenticity of the anger in the video. The pupil response when people watching some videos is used as physiological signal to predict whether the anger they see is real or performed. The accuracy shows that the result of training LSTMs with the original time series data of pupil change is obviously better than that of the three-layer neural network trained with the statistical variables of original dataset. In addition, I calculated the gradient of the output with respect to the input as ***causal index*** to explain the output from neural networks and the mean value of all inputs that produce the same output as ***characteristic input*** for this output to predict the results from the neural network. Finally the adaptability of these neural network interpretation methods on LSTMs is checked in this paper.

Keywords: Analysis of emotional authenticity, LSTMs, Three-layer-neural network, causal index, characteristic input, neural network interpretation methods.

1 Introduction

1.1 Dataset

The purpose of the experiment to obtain this data set is to test how keen humans are in the ability to consciously detect the authenticity of anger and to further test their ability to unconsciously detect anger through pupil response (a physiological signal of humans). [1] The data set used in the experiment was the pupil changes of 22 participants when they watched 20 angry videos. Among those videos, 10 of them express real anger, and the other 10 videos express performed anger. The new dataset contains three tables, which provide the changes of left and right pupil with time for the 22 participants when each of them watch each video, and the average changes of pupil for all participants with time during each video. The first data set received was statistical data (such as the mean value) on pupil changes of each participant for each video; this dataset is still used in traditional neural network training. The physiological signals are directly recorded from the participants, not controlled by human consciousness, so they are less affected by subjective bias.[1] Therefore, the neural network trained by the dataset above can be used to represent the ability of human beings to perceive the truth of anger unconsciously.

1.2 Problem description Model introduction

The problem that this paper solves is to predict whether the anger in this video is performed or real according to the pupil changes of human beings when watching a video which expressed anger. Using the data sequence of pupil size with time when a person watching a video showing anger as input to train LSTMs and three-layer neural network. The purpose is to find out whether the anger expressed in the video watched by this person is real or performed. First of all, this paper uses the traditional neural network method, I use the statistical data of pupil size (such as the mean value, etc.) to train the three-layer neural network and get the result with the accuracy of 77%. As there is no direct connection between the neurons of the traditional neural network, it is almost impossible to use the information in front to predict the events in the back. However, the new dataset introduced in this paper is the original data of pupil size with time during the whole process of watching the video, which is time series information. Therefore, in order to use the data series for prediction, this paper further introduces the ***Recurrent Neural Network (RNN)*** which can allow the information to persist. Traditional RNN can connect the information between input sequences, which is a recursive process. After error retransmission in each layer, a multiplier of the derivative of the activation function will be introduced. As a result, after multiple steps, the multiplier's multiplication will cause a series of troubles, making the error approach zero or increase exponentially. This is the so-called problem of ***gradient disappearance*** and ***gradient eruption.*** [7] So the traditional RNN can't deal with long-distance dependence. In this paper, the length of data sequence is 186, which is a long time series，to fit this situation, this paper adopts ***Long Short-Term Memory Networks(LSTMs)***, according to [8]，“Long Short-term Memory (LSTM) is an RNN architecture designed to be better at storing and accessing information thanstandard RNNs.”LSTMs is mainly used for time issues. It can analyze audio and video data etc [9]. This network can bring the addition operation into the network through the delicate gate control.

Gate is a node which can control information flow. They are composed of sigma neural network layer and point by dot product. LSTMs has three gates to protect and control the information flow vector state. Gates let LSTMs have ability to remove or add information to nodes to change the state of information flow, and solves the problem of gradient disappearance caused by long-distance dependence. Therefore, LSTMs is more in line with the requirements of my question.

1.3 Causal Index and Characteristic Input

Although the neural network can be trained to build a model that meets the requirements, and get a higher accuracy, it still has a disadvantage: people cannot intuitively get how the neural network makes a certain kind of prediction. Nowadays, many methods have been used to explain the conclusions of neural network. "These conclusions need to generate at least a set of rules to express the knowledge learned by neural network" [3]. One way to explain neural network depends on pruning the network to a minimal size. This method forces the internal representation of neural network to approximate a symbolic and readily extractable mode. However it will reduce the robustness of the neural network. [3] So I used the technology in [2] to get the explanation of the model, that is, to use the ***causal index*** to explain the decision made by neural network. ***Causal index*** depends on the fact that the networks activation functions are differentiable [3]. The gradient change between output and input of a neural network will be used to find rules. We can uses mathematical means to find the rate of change of output neuron with respect to input neuron [3], then the rate we calculate is ***causal index***. Its change can reflect how the neural network gets their decision. In addition, it is mentioned in [2] that using the mean or median of each input value, a pattern representing each input class is created as a ***Characteristic pattern*** for that class. Those input patterns which turn an output on will be used to create ***characteristic ON pattern.*** Similarly a ***characteristic OFF pattern*** can also be created. As for the problem I want to solve, since the video I need to predict only have two categories (Posed or Genuine), I can directly regard the input pattern makes the result as Genuine as output ON and others(result as Posed) as output OFF. This method can be used to predict the output of the neural network according to the input. In this paper, in addition to implement these interpretation methods on three-layer neural network, I also test their adaptability on LSTMs network.

2 Method

The hardware basis of this paper is: Memory: 8.00GB; CPU: Intel(R) Core(TM) i7-8750H CPU @2.20GHz 2.21 GHz; Operating system: Windows 10; Software: python3

2.1 Data preprocess

The original data of all the pupil size of left and right eyes of each participant for each video is used in the LSTMs training. First of all, I integrated everyone's left eye and right eye data, and removed all empty data columns. Then I pad the dataset by filling in all null values with 0. Finally, all the data are normalized and labeled according to the video name, indicating that the video expresses real anger or performed anger. Real anger will be labeled as 1 and performed anger will be labeled as 0. Since the data sent to the LSTMs each time is a pupil size data sequence, it is reflected as a column in the dataset in the original data. For the convenience of operation, I transpose the data set. Since then, each row in the dataset is an input to the LSTMs each time. I store the processed data in "LSTM_ preprocess.csv". Finally the processed data is randomly divided into training set (75%); verification set (12.5%) and test set (12.5%).

Traditional neural network used data from statistical analysis of pupil size sequence in participants. For the original dataset, I deleted the video number and the identifier of each participant, and trained the three-layer neural network with the remaining six features. In the training, the angry in the video is "real(Genuine) or not(Posed)" is used as the prediction target, then label "Posted" is represented by 0, and "Genuine" is represented by 1 just like I do above for the data using by LSTMs. The data set is randomly divided into two parts according to the proportion of 8:2, that is, 80% of the data is used for training and 20% for testing. In addition, the normalization of data sets is also implemented to further improve the accuracy of the model.

2.2 LSTMs

The input layer of LSTMs requires a three-dimensional input, including the batch size for training, input dimension and time step of each input data. In this paper, for each training epoch, the input is a row of pupil size data series which changes with time. Only one data representing a pupil size is put into LSTMs each time. Therefore, the input dimension is 1, and the time step is the length of the data sequence, which are 186 in this paper. In the experiment, I tried the mini batch method, which will have a good effect combined with SGD (random gradient descent optimization algorithm). In this method, parameters are updated by small batch of data, which reduces the calculation burden. However, after debugging, I finally chose the batch size = 1 with the ***Adam optimizer***. In the following traditional three-layer neural network training, I also continued to choose this optimizer. Adam optimizer combines the advantages of optimizer AdaGrad and RMSProp. In this optimizer, the first moment estimation and the second moment estimation of the

gradient are considered synthetically to calculate the update step [4]. This optimizer is simple to implement, efficient in calculation and requires less memory. Then its parameters are very explanatory, and usually do not need to be adjusted or only need a few fine-tuning, what is more, it can also realize step annealing process naturally (automatically adjust learning rate).

The loss function I chose is cross entropy loss, which is widely used in classification problems. It describes the distance between two probability distributions. The closer the cross entropy is, the closer the two distributions are. This loss function could measure subtle differences. Therefore, this method is also adopted in both LSTMs and traditional neural network.

For LSTMs，learning rate is the most important parameter, followed by the size of hidden layer. Changing the size of the hidden layer will not change the optimal learning rate's region [9]. Therefore, I pay great attention to the choice of learning rate and hidden layer size in the process of practice. For LSTMs and this dataset, according to the debugging, finally 0.0001 is selected as the learning rate. In my experiment, I tried learning rate decay method to let the learning rate decreasing with training epochs, but the effect was not as good as fixed learning rate, so I didn't use this method at last. In addition, generally speaking, the more layers, the smaller the error of the whole network, but it will make the whole network more complex, increase the training time of the network, and may also lead to "over fitting". Considering the size of the data set in this paper, the LSTMs with hidden layer 2 is adopted in this paper. It takes the multidimensional hidden output of the three time steps in the first layer as the input of the three time steps in the second layer. Besides that, the number of hidden nodes in each layer is the direct cause of over fitting. Therefore, it is necessary to adopt as compact structure as possible under the condition of satisfying the accuracy. After debugging, I chose 32 as the number of hidden units. Since this is a binary classification problem, the number of neurons in the output layer is set to 2, it contains the probability that this input belongs to our two categories; the one with larger probability is the final result. A linear activation function is used to connect the hidden layer and the output layer. The LSTMs was trained 300 times.

To sum up, the model parameters of LSTM are: learning rate is 0.0001, number of hidden layers is 2, number of hidden layer neurons is 32, size of input sequence (batch size, time step, input size) is (1,186,1), output layer has two neurons and training epochs is 300. During the training process, the loss value and accuracy calculate by validation dataset is output every 10 times, and the best model is saved according to the accuracy of the output. In the final test, the model with the highest accuracy obtained in the training process will be used. Finally, in order to improve the training speed, all the training in the LSTMs part of this paper is completed on GPU.

2.3 Neural Network

In order to cooperate with the explain work, this paper uses three-layer neural network which includes input layer, output layer and one hidden layer. The number of neurons in the input layer is the number of features in processed data, and there are two neurons in the output layer. ***Sigmoid*** function is chosen to be the activation function for both hidden layer and output layer. This function is a commonly used nonlinear activation function, which can map all real numbers to (0, 1) intervals, and normalize the data by nonlinear method. It is usually used in the output layer of regression prediction and binary classification which is in line with our requirements. According to the debugging results, the activation function can get the highest accuracy. The loss function and model optimizer are just the same as LSTMs which showed above.

In the process of model parameter adjustment, I adopted the 5-fold cross validation method. In this way, the original data is divided into five groups on average, each subset data is used as a validation set one time, and the remaining four sets of subset data are used as train sets, so that five models can be obtained, and the average of the accuracy on the validation set of these five models is used as the final accuracy under this 5-fold cross validation. This method makes full use of all samples. It can effectively avoid over-learning and under-learning, and the final result is more convincing.

The final parameters are determined as follows: The number of neurons hidden layer is 64, the learning rate is 0.03, and the model was trained 400 epochs. During the training process, all the loss values and partial accuracy is recorded, and the loss value and accuracy are printed out every 100 times and 20% of the data set is used for model test.

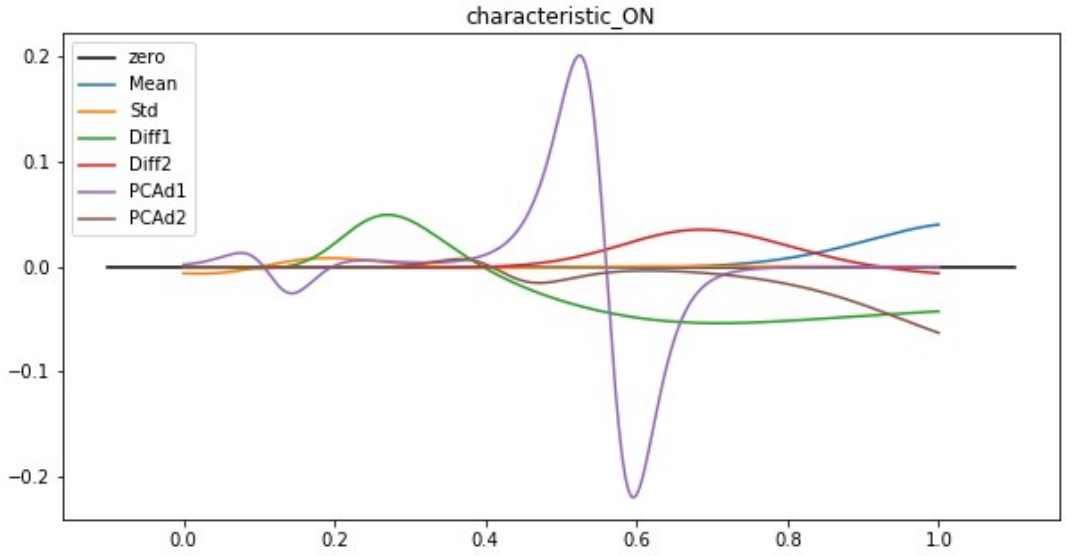
Finally, the random number seed is set to 1 in the running process of the two models, which is convenient for replication and comparison.

2.4 Characteristic input and Causal index

For comparison, both models calculate characteristic input in the same way. 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. Finally, I compare the prediction results with the actual results of neural network and calculate the accuracy.

For three-layer neural network, according to the formula mentioned in [2], causal index is the gradient of the output with respect to the input. To get the causal index， I build new datasets for each feature. In each datasets, except for the feature need to calculate causal index will be set as 1000 random numbers from 0 to 1, other features are all replaced by the characteristic value in the characteristic ON or characteristic OFF group. That means each dataset will have 1000 data. These data are put into the trained model to get the result. After that, I calculate the causal index of each feature in the ON and OFF states and draw the curve. In this paper, two kinds of graphs are visualized, one is the causal index curves of all features in the whole ON group and OFF group which show in ***Figure 1*****,** and another is the comparison of causal index curve of each feature in the ON or OFF state partly shows in ***Figure 2.*** For ***Figure 2,*** the complete pictures can be viewed in the code file. Only one picture of the feature (PCAd1) is shown.



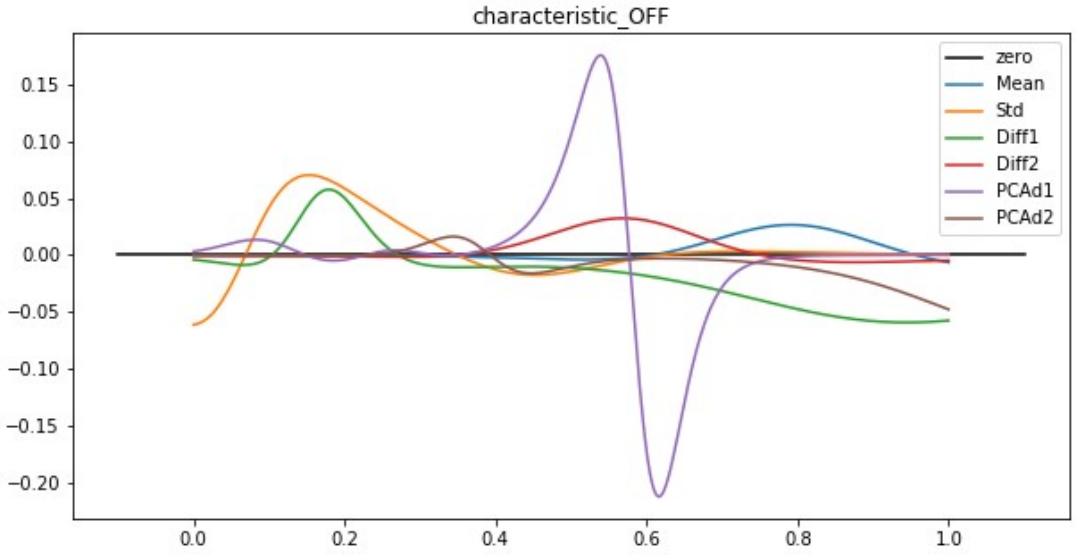


Fig. 1. The casual index for all features characteristic ON pattern and characteristic OFF pattern

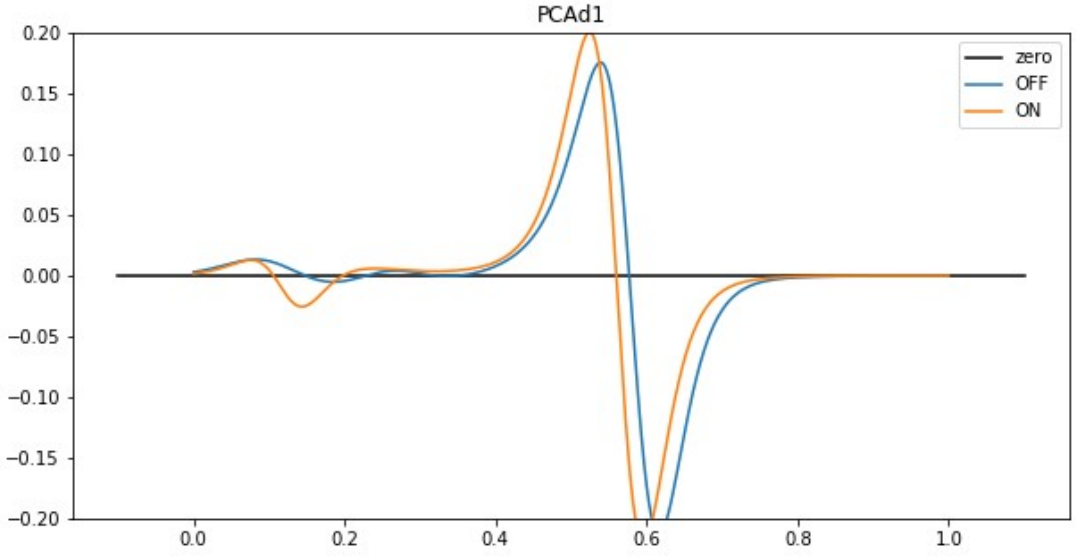


Fig. 2. The casual index for feature PCAd1 in characteristic ON pattern and characteristic OFF pattern

Because the input of LSTMs is a sequence rather than several features of traditional neural network, the method of calculating the traditional neural network's causal index cannot be used. So this paper only implements characteristic input on LSTMs and calculates the accuracy. The way to get characteristic input on LSTMs is the same as that on three-layer neural network just like I mentioned above.

For the traditional neural network, I get the rules of each output of neural network according to the curve of causal index. According to the image obtained above and the relationship between the causal index and rule described in [2], I find all the characteristic values that can make the result reverse, record the points where the causal index tends to be gentle, and finally I generate the rules about ON and OFF. For example in ***Figure 2***, It is obvious from the curve that PCAd1 causes the result to flip, and tends to be gentle around 0.8, and this trend appears in both ON and OFF pattern. So PCAd1 < 0.8 will be add in the rule for characteristic ON pattern and characteristic OFF pattern. “PCAd1<0.8” is called a sub rule. And the total rules of ON pattern and OFF pattern are consisted of several sub-rules like this. The next work is to analyze the causality index curve of other characteristics and get all the sub-rules. All the sub-rules that affect ON pattern are connected with join, and all the sub-rules for OFF pattern are connected with disjunction.

After the rule is generated, instances are extracted from the results of the test set to verify the accuracy of the rule, and the instances are brought into the rule to verify the accuracy of the rule. Because our model only needs to divide video into two categories, there are two rules, which are rules leads to “Posed” result and “Genuine” result. Since the result must be “Posted” or “Genuine”, the rules for generating all results can be summarized directly through the ON and OFF pattern’s causal index. In this article, ON means “Genuine” and OFF means “Posted”.

Since we only have two prediction results "Posted" and "Genuine", it is not meaningful for this case to analyze the next most likely output, so this step is finally omitted in my results. The output of interpretation decision is divided into five parts:

1. Network output,
2. the most likely output of the network according to the rules and characteristic input,
3. Characteristic value of inputs which are important on the current output.
4. The rule corresponding to the current output,
5. Input of the current case (for comparison).

Finally, in order to check the accuracy of causal index to the rules generated by traditional neural network, this paper sets all the features that do not appear in rules to 0 to generate a new data set, and trains the model again with the new data set. The parameters of the model and the partition scale of the dataset are all kept unchanged. Then I test the accuracy of the new training model, and compare it with the accuracy of the model trained with all features, so as to reflect the accuracy of the important input selected by our rules. If there is no significant difference between the two accuracy rates, then it can be said that my rules are accurate for finding important input.

3 Results and Discussion

Judgment method	Predicted accuracy	Explain accuracy
Human consciousness	60%	-
3-layer neural network	77%	67%
LSTMs	94%	52%

Table 1. The accuracy of three methods and the accuracy of result prediction by characteristic pattern

3.1 Predicted result

The second column of ***Table 1*** shows that the accuracy of the LSTMs is 94%, which is nearly 20% higher than that of the traditional neural network (77%). That is to say, LSTMs is more suitable to analyze the authenticity of anger in video according to the time series of pupil size. And can gain very accurate results. The LSTMs methods can reach the same level of accuracy like 94% achieved in [1]. According to the different conditions of randomly divided data sets, the accuracy can even reach 97%. The results of many experiments are about 95%. It has been mentioned in part 3 that in order to reproduce the model effect and compare the two models, I set the random number seed in the operation of the two models to 1, so the accuracy of LSTMs is fixed at 94%.

From the results obtained, for analyzing the authenticity of the anger expressed in the video, the results using the time series of the original pupil size data are much better than the results using the statistical values (such as the mean value) between the pupil data. For the traditional neural network, neurons do not affect each other, so it is impossible to adjust the output according to the previous input or learn the time series data. RNN solves this problem. It allows information to persist. However, the traditional RNN has the problems of "vanishing gradient" and "exploding gradient" introduced in 1, so it cannot deal with long-distance data series in practical application. The length of the data sequence in this paper is 186, which could be considered as a long sequence. The performance of adopting the traditional RNN is not ideal. LSTMs is more suitable for this situation. It can learn about long-term dependency. The key point of LSTM is the state transmission of its cell. The vector passes through the whole cell, only a few linear operations are done to achieve long-term memory retention [10]. In addition, in order to add or delete information and make the network contain useful information as much as possible, LSTM adopts the three gates structure: forget gate layer, input gate layer and output gate layer to protect and control information [10]. These three gates determine what information the cell needs to discard, what new information it stores, and what information it ultimately outputs. These structures make the network not lose long-term memory caused by vanishing gradient. All the input time series data can be used to affect and adjust the network.

According to ***Table 1***, the accuracy of human conscious recognition is 60% [1].The pupil response to physiological signals is less affected by subjective bias [1], and both of my networks are trained based on human physiological signals and get higher accuracy (77% and 94%). From the results, the accuracy of using deep learning method to judge the authenticity of human emotions according to human physiological signals is higher than that of human self-awareness. This shows that the ability from human unconscious to distinguish the authenticity of emotions is higher than that of their consciousness. That is to say, the machine can use the mathematical analysis method that the human brain cannot use at present, combine with the physiological signals from human body, to get more accurate results in the authenticity analysis of emotions than the conscious analysis of human being. This result shows the broad application prospect of machine learning in human emotion analysis, as it can recognize the real anger and the displayed anger with high precision by using the physiological signals of the emotional perceiver. Besides, this is the result of the analysis of human physiological signals by the machine, rather than the analysis of video directly by the machine. Using machine learning directly on video is unlikely to be useful [5, 6]. Therefore, for the detection of human emotional authenticity, it is a good way to use neural network to analyze the physiological signals of the stimulated subjects rather than analyze the video directly.

3.2 Explain result

According to the accuracy of the prediction of the neural networks' result predicted by characteristic input pattern shows in the third column of ***Table 1***, the accuracy of this method on the 3-layer NN is 67%, which is higher than 52% on the LSTMs. This method can successfully predict some output correctly from neural networks, but the effect is not very accurate, what is more, although this method has certain accuracy in LSTMs, it is not as good as in 3-layer NN. This result shows that the adaptability of characteristic input pattern on LSTMs is not as good as that of NN, which indicates the limitation of this method to a certain extent. Besides that, even on the 3-layer NN, the accuracy of this method is not very high. Only by calculating the variance between the characteristic input pattern and the input data to predict the results of the neural network and get the specific output pattern of the neural network is still a method which needs to be improved. That is to say, it is not accurate enough to predict the output of the neural network only by analyzing the input of the neural network. It is obviously not excellent enough to use the average or median of the input group which get the same output of the neural network to gain the characteristic input pattern of that output. The analysis of neural network should have more accurate and complex standards or methods.

In this paper, I calculate the causal index of all features on the 3-layer NN and get the interpretation rules according to their curves. The accuracy of interpretation is about 58%, which is lower than the result in [2]. From the obtained causal index curve, PCAd1 has the best effect ***(Figure 2***) and can be used to distinguish rules very clearly, while the gradient curve between some features and output is difficult to describe. These features are judged as unimportant inputs. Although they improve the effect of the neural network to a certain extent, they are not as influential as the features with causal index curve which can gain distinguishable rules. To test the accuracy of the important input from the detection rules, I set two features which never show in both rules to 0, that is, 1 / 3 of the data is discarded, and the accuracy of the model is reduced from ***77% to 74%***. This proves that the important input generated by the rules is quite accurate. To a certain extent, this method makes up for the shortcomings that the results of neural network cannot be explained and makes it more convenient for users to understand which features play an important role in training. At the same time, the method of pruning the neurons in the neural network is not used, which protects the robustness of the neural network. Because the LSTMs learning object is long-distance time data, which is different from the traditional NN learning object (signal features), it is difficult to calculate the causal index. Therefore, the way to determine rule based on causal index cannot be promoted on LSTMs. That is to say the way of calculate causal index to gain the rules to explain the output of neural network has poor adaptability on LSTMs.

4 Conclusion and Future Work

For the problem of this paper, the accuracy of training LSTMs with original time series data is significantly higher than that of training traditional NN with statistical data. In other words, in terms of analyzing the authenticity of emotions, the changes sequence of all physiological signals when human perceives an emotion can better reflect the authenticity of an emotion than the statistical data such as the average value of this sequence. In addition, Deep Learning method shows a high degree of accuracy in using human physiological signals to distinguish whether an angry mood is a performance, especially LSTMs which can accurately distinguish the authenticity of an emotion according to the sequence of human physiological signals. Those methods show that they can be widely used to identify human emotional authenticity through human physiological signals.

Secondly, the method of using causal index to explain the neural network is also successfully realized on the 3-layer neural network, and shows high accuracy in determining the important input which affects the performance of neural network. This method can explain the decision-making of 3-layer neural network to a certain extent, so that people can better understand the reasons why neural networks make decisions. However, the method of using causal index to generate interpretation rules has not been practiced on LSTMs, which shows that it has certain limitations and widely improvement space for different input objects' networks. In addition, The accuracy of predicting the result from neural network by using characteristic input pattern is not very ideal, the accuracy on LSTMs is lower than that on 3-layer NN，therefore, the method that explains the result of neural network by characteristic input can be further improved and developed to obtain higher accuracy and wider applicability.

In the future, we can adopt the variants of LSTMs, such as GRU model to further improve the result and performance of predicting the authenticity of anger according to human physiological signals, make the results more accurate, and further simplify the model construction. In addition, we can study the adaptability of causal index to other networks except that on RNN. And improve the way to summarize the information displayed by the causal index, so as to make the interpretation more accurate and clear. Finally, we can improve the way to find the characteristic input pattern of neural network output, which makes the input pattern more accurate and can better explain the output.

References

[1] Lu, C., Tom,G. , Md,Z.H., Sabrina,C.:Are you really angry? Detecting emotion veracity as a proposed tool for interaction. In Proceedings of the 29th Australian Conference on Human-Computer Interaction (OzCHI '17), Brisbane, QLD, 5(2017).

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

[3] 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)

[4] Kingma, Diederik P. and Jimmy Ba. "Adam: A Method for Stochastic Optimization." *CoRR* abs/1412.6980 (2015):

[5] Akshay A., Conrad S., Tamás D. G., and Roland G.: Learning-based face synthesis for pose-robust recognition from single image. In Proc. Brit. Mach. Vis. Conf. (BMVC '09). British Machine Vision Association and Society for Pattern Recognition, London, U.K. 1-10. DOI: https://doi.org/10.5244/C.23.31(2009)

[6] Ian G., Nicolas P., Sandy H., Yan D., Pieter A., Jack C.: Attacking Machine Learning with Adversarial Examples. Retrieved from https://blog.openai.com/adversarial-example-research(2017.)

[7] Hochreiter S.The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems. 6. 107-116. 10.1142/S0218488598000094. (1998).

[8] Kawakami K. Supervised sequence labeling with recurrent neural networks. Ph. D. dissertation, PhD thesis. Ph. D. thesis. (2008).

[9] Greff K, Srivastava RK, Koutník J, Steunebrink BR, Schmidhuber J. LSTM: A search space odyssey. IEEE transactions on neural networks and learning systems. Jul 8;28(10):2222-32(2016).

[10] Olah, C. Understanding lstm networks. (2015).