Comparing Back Propagation Neural Network and LSTM Neural Network on Detecting Angry Emotion

Mingjun Jiang
School of Computing, Australian National University
u7172733@anu.edu.au

Abstract. The issue of this article is how to recognize human anger emotion by observing pupillary response patterns. The first part of this paper creates a classify three layers neural network module with backward propagation and trains the observer's pupillary response dataset and predicts the Genuine/Posed, calculating the progress of loss error and the confusion matrix. The second part of this paper creates a LSTM neural network with three layers and does the same context as the first part: classify the observer's pupillary response dataset. The third part is analyzing the weight matrix including the Garson formula and brute force technique, working out which input is important for the classification. Furthermore, to select the best number of hidden neurons, calculate the different loss errors with a different number of hidden neurons. Finally, find the most important input feature in both bp and LSTM neural network. This paper proves that LSTM neural network have a good result in training pupillary response dataset and the pupillary response can distinguish the angry emotion successfully.

Keywords: detecting angry emotion, pupillary response, machine classification, back propagation neural network, LSTM neural network

1 Introduction

Anger is a significant emotion. How to recognize anger is also a meaningful question. Lu Chen, Tom Gedeon, Md Zakir Hossain, and Sabrina Caldwell researched on how to recognize human's anger emotion by observing pupillary response patterns.[1] The purpose is to examine empirically how well humans are capable of consciously detecting the veracity of anger, and further, their non-conscious ability to do so, as reflected in their pupillary responses. The experiment collected 22 participants' verbal response and their pupillary response in viewing two types of anger stimuli (Genuine and posed). The mean accuracy by neural network training of pupillary response is 95% while the mean accuracy by Verbal response is 60%. It claims that machines can reach high accuracy than observed by human.

The module of neural network effect the accuracy a lot. Back propagation is a basic neural network, which is widely used in machine learning. Besides, the dataset is the changes of pupillary response in the processing of watching the different angry videos, and LSTM neural network is well-suited in classifying, processing and making predictions based on time series data. Therefore, I train the dataset by both bp neural network and LSTM neural network, and then compared them.

Analyzing the pupillary response of seeing angry video to recognize whether it is really angry is benefit for us to know the emotion of human beings, which can use for human robot interaction.

2 Back Propagation Neural Network

2.1 Back Propagation Neural Network description

Backpropagation is the central mechanism by which artificial neural networks learn. It is the messenger telling the neural network whether it made a mistake when it made a prediction. The backpropagation algorithm works by computing the gradient of the loss function with respect to each weight by the chain rule, computing the gradient one layer at a time, iterating backward from the last layer to avoid redundant calculations of intermediate terms in the chain rule.[2]

2.2 Create a bp Neural Network

The input features of the data is the six pre-processing features of the time sequence, which are Mean, Std, Diff1, Diff2, PDAd1, PCAd2. The dataset owns 400 numbers of data in total.

I create a backpropagation neural network to classify the real anger and fake anger, with 1 input layer, 1 hidden layer, and 1 output layer by PyTorch. The layers are fully connected. I marked Genuine as 1, Posed as 0. The input layer has 6 inputs, and the hidden layer has 12 hidden neurons, while the output layer has 2 outputs. And I used the basic sigmoid logistic activation function: $y = (1 + e^{-x})^{-1}$

The dataset has 400 pupillary response data after normalization. I divided 80 percent of data into a training set, and 20 percent of data into a test set. Other parameters are as follows. The learning rate is 0.01 while epochs are 3000. The optimizer is Adam, loss function is CrossEntropyLoss and batch size is 1. After training, the training accuracy is 85%, while test accuracy is 70%.

2.3 Loss Function and Confusion Matrix

The training result is 80%. And I calculate the loss function and confusion matrix to display the training result on other side. The loss function and confusion matrix are as follows.

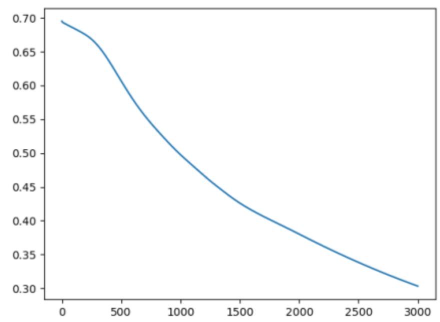

Fig. 1 X axis is the number of epochs and Y axis is loss error. It shows that the loss function of training process. It is clearly that the loss error is decreasing significantly with the progress of training.

Table 1. Confusion Matrix for training set of bp neural network

	Genuine	Posed
Genuine	135	21
Posed	21	135

Table 2. Confusion Matrix for testing set of bp neural network

	Genuine	Posed
Genuine	35	9
Posed	15	29

Both training set and the testing set have a good result, with 85.25% accuracy in training set and 72.32% accuracy in testing set.

3 Analyzing Back Propagation Neural Network

3.1 Analyzing techniques

Garson (1991) used G_{ik} to measure the contribution. G_{ik} is the contribution of input i to output k. The disadvantage is that some positive weights and negative weights may cancel in the processing of calculating. [3] W_{ij} means that the weight value between input i and hidden neuron j (as the same as W_{pj}, W_{qj}), and W_{jk} means that the weight value between hidden neuron j and output k. The formula is as follows.

$$G_{ik} = \frac{\sum_{j=1}^{nh} \frac{W_{ij}}{\sum_{p=1}^{ni} Wpj} \cdot W_{jk}}{\sum_{q=1}^{ni} \left(\sum_{j=1}^{nh} \frac{W_{qj}}{\sum_{p=1}^{ni} W_{pj}} \cdot W_{qj} \right)}$$

Milne (1995) used M_{ik} to measure the contribution. M_{ik} is the contribution of input i to output k. [4] Comparing to G_{ik}, M_{ik} considers the contribution by calculating absolute of weights. The other parameters meaning is as same as G_{ik}. M_{ik} eliminates the affection of positive and negative weights cancel, but the meaning of divisor is not clearly. The formula is as follows.

$$M_{ik} = \frac{\sum_{j=1}^{nh} \frac{W_{ij}}{\sum_{p=1}^{ni} |Wpj|} \cdot W_{jk}}{\sum_{q=1}^{ni} \left(\sum_{j=1}^{nh} \left| \frac{W_{qj}}{\sum_{p=1}^{ni} |W_{pj}|} \cdot W_{qj} \right| \right)}$$

Wong, Gedeon and Taggart (1995) used P_{ij} to measure the contribution. The other parameters meaning is as same as G_{ik}. P_{ij} is the absolute of value contribution of input i to hidden neuron j. The formula is as follows.

$$P_{ij} = \frac{|W_{ij}|}{\sum_{p=1}^{ni} |W_{pj}|}$$

Q_{ik} is the extend for input i to output k, which sign is clearly. The formula is as follows.

$$Q_{ik} = \sum_{r=1}^{nh} (P_{ir} \times P_{rk})$$

Therefore, G_{ik}, M_{ik}, Q_{ik} are three important parameters to measure the contribution of input i (six inputs in total) to output k (two inputs in total). Each of them has their own advantage and disadvantage.

3.2 Result

After training, I calculate the G, M, Q of bp the neural network separately, and shows curves of G, M and Q. the answer is as follows.

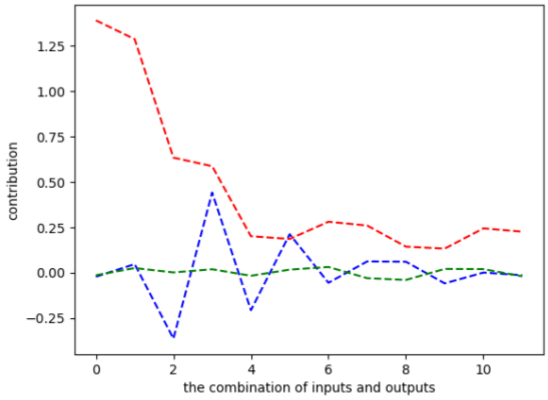

Fig. 2. X axis is the combination of inputs and outputs while Y axis is the corresponding contribution. The blue line is G, the green line is M, and the red line is Q.

The input features of human's pupillary response patterns are Mean, Standard, Diff1, Diff2, PCAd1, PCAd2 in order. The first half of the combinations is these six features with output1, while others is these six features with output2. The absolute of contribution is bigger, the weight of this input is higher.

There are some combinations of input and output contribute high weights in the neural network by all these three calculate methods. Compare to G_{ik} and Q_{ik}, M_{ik} does not exist significant variation in the calculation, While the analyzing the contribution by G_{ik} and Q_{ik} will get similar conclusion. Therefore, I will take G_{ik} and Q_{ik} as the main parameters to analyze this neural network.

In detail, the combination of Mean and output1, the combination of Standard and output1, the combination of Diff1 and output1 contribute most weights in this neural network. The combination of Diff2 and output1 also contribute important part in the neural network.

In the combination of features and output2, Mean, Standard are also the most important feature and influence the result most, which can discover in Q_{ik}. However, in M_{ik} and G_{ik}, the contribution of features is not obviously.
Above all, feature Mean and feature Std are the significant input in the neural network, which influence the classification most. They contribute a lot and have high weight in the connections.

Therefore, in human's pupillary response pattern, mean and standard mainly used mean and std of the pupil shifting parameter to reflect whether it is a really angry expression or not.

3.3 Brute Force Analysis

I used brute force analysis to eliminate pairs of inputs. Calculate the loss of the new neural network after eliminating. It can also reflect the contribution of each input features. There are 15 combinations in total. Additionally, I run 4 times in case some occasional cases.

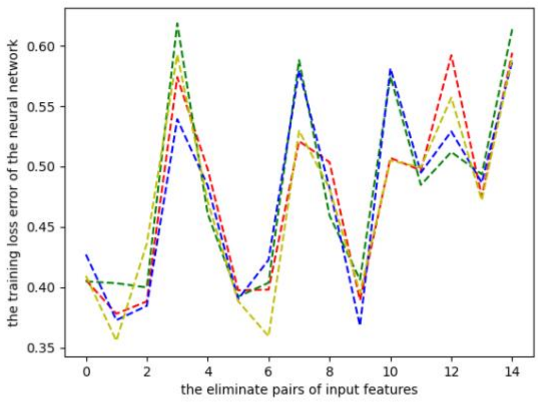

Fig. 3. X axis is the eliminate pairs of input features, and the Y axis is the training loss error of the neural network with corresponding hidden neurons.

In this analyzing, the training loss error fluctuate significantly with the different pairs of input features eliminated. The sequence of eliminated pairs are feature1 and feature2, feature1 and feature3, …, feature1 and feature6, feature2 and feature3, feature2 and feature4, …, etc. The input features of human's pupillary response patterns are Mean, Standard, Diff1, Diff2, PCAd1, PCAd2 in order.

To be specific, the fourth, eighth, eleventh, thirteenth, fifteenth pairs of input features contribute important weights in the neural network. Without of these pairs, the loss error of the neural network will increase significantly. These pairs of input features corresponding the specifically features as follows. Mean and PCAd1, Standard and PCAd1, Diff1 and PCAd1, Diff2 and PCAd1, PCAd1 and PCAd2.

It is clearly that all of these pairs have the feature PCAd1. Hence, PCAd1 is an important input feature in this neural network. In other word, without PCAd1, the loss error of neural network will increase significantly.

3.4 Number of hidden units

To select the best hidden units, I modify the number of hidden units, and use brute analysis to train the neural network with different hidden units. The loss error of each hidden units are as follows.

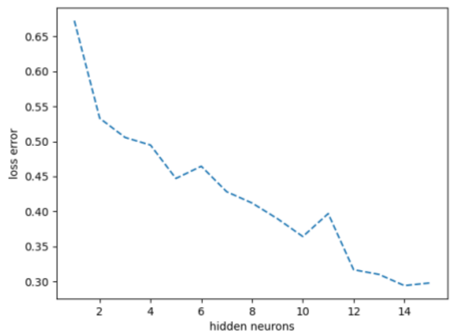

Fig. 4. X axis is the number of hidden neurons, and the Y axis is the training loss error of the neural network with corresponding hidden neurons.

In fig. 4, it is obvious that the loss error of neural network is decreasing significantly at first, reaching the bottom when the hidden neurons is 12. After that, with the increasing of the hidden neurons, the loss error is beginning to fluctuate slightly. Therefore, I select 12 as the best hidden neurons.

4 LSTM Neural Network

4.1 LSTM Neural Network introduction

Long short-term memory (LSTM) is a special kind of RNN, mainly to solve the problem of gradient disappearance and gradient explosion during long sequence training. Simply put, compared to ordinary RNNs, LSTM can perform better in longer sequences. [7] LSTMS follow the same basic structure as RNNS except that the nonlinear units in the hidden layer are replaced by memory blocks Each memory block contains one or more memory cells along with input, output and forget gates which control flow of information into and out of the memory cell

LSTM networks are well-suited to classifying, processing and making predictions based on time series data, since there can be lags of unknown duration between important events in a time series. LSTMs were developed to deal with the vanishing gradient problem that can be encountered when training traditional RNNs. Relative insensitivity to gap length is an advantage of LSTM over RNNs, hidden Markov models and other sequence learning methods in numerous applications.

4.2 Pre-processing

The dataset contains twenty videos which were displayed in twenty different sequences. For each video (T1-T10, F1-F10), recording twenty different observer's pupillary response by time sequence. I use the data of pupillary response of left eye and select the top 100 data in each time sequence, using the interpolate method to remove zeros from eye blinks. Then, I delete the empty column and get 373 time sequences in the end.

4.3 Create a LSTM Neural Network

I create a LSTM neural network to classify the real anger (Genuine) and fake anger (Posed) by PyTorch. The relative parameters are 100 hidden neurons and 1 LSTM layer. An output layers which has 2 features are fully linear connected after LSTM layer. I marked Genuine as 1, Posed as 0. Each input is a sequence which length is 100.

The dataset has 373 pupillary response data after normalization, which delete the empty values. I divided 80 percent of data into a training set, and 20 percent of data into a test set. Other parameters are as follows. The learning rate is 0.001 while epochs are 100. The optimizer is Adam, loss function is CrossEntropyLoss and batch size is 1. After training, the training accuracy is 99%, while test accuracy is 98%.

4.4 Loss Function and Confusion Matrix

The training result is 100.00%. And I calculate the loss function and confusion matrix to display the training result on other side. The loss function and confusion matrix are as follows.

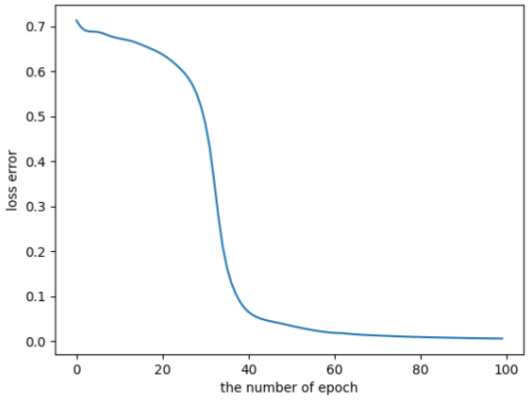

Fig. 5 X axis is the number of epochs and Y axis is loss error. It shows that the loss function of training process. It is obvious that at the beginning of the training, the loss error decrease slightly, after that, the loss error is decreasing significantly with the progress of training until the accuracy is high.

Table 3. Confusion Matrix for training set of LSTM neural network

	Genuine	**Posed**
Genuine	142	0
Posed	0	145

Table 4. Confusion Matrix for testing set of LSTM neural network

	Genuine	**Posed**
Genuine	43	1
Posed	0	42

Both training set and the testing set have a good result, with 99.67% accuracy in training set and 98.84% accuracy in testing set. Comparing with bp neural network, the result of LSTM neural network is much better.

4.5 Analyzing LSTM neural network

I used brute force analysis to eliminate inputs. I divide the data sequence into 10 parts by time sequence. Calculate the test accuracy and loss error of the new neural network after eliminating. It can reflect the contribution of each input features.

Table 5. loss error of train set and accuracy of test set which eliminate inputs (1-10, 2-20 ….)

Eliminate inputs	Loss error	Accuracy
1-10	0.0103	98.61
11-20	0.0086	98.61
21-30	0.0059	98.61
31-40	0.0090	98.61
41-50	0.0079	98.61
51-60	0.0056	98.61
61-70	0.0062	98.61
71-80	0.0091	98.61
81-90	0.0113	98.61
91-100	0.0187	77.78

It is obvious that the lack of the former input will not have a great impact on the result. The lack of the following 80-100 input will cause a significant change in the loss error of the result, while the lack of 90-100 input will cause a significant decrease in accuracy of test set. Hence, 90-100 input make more contribution in LSTM neural network.

5 Conclusion

For bp neural network, by magnitude analysis, after training, the input Mean and Standard occupy largest weight and contribute most to the neural network. By brute force analysis, input PCAd1 influence the neural network most. Without feature PCAd1, the loss error will increase significantly.

For LSTM neural network, it shows that the 90-100 input of data influence the neural network most. Therefore, the pupillary response of observers at that time, can reflect whether the video is really angry most.

Compared bp neural network and LSTM neural network, the LSTM neural network has a better result than np neural network. It proves that LSTM neural network performs well in training sequentially. Finally, it can reach 99% accuracy in training set and 98% accuracy in test set. Therefore, the pupillary response can find whether the video is really angry.

6 Future

Analyzing the pupillary response of seeing angry video to recognize whether it is really angry is benefit for us to know the emotion of human beings, which can use for human robot interaction. Furthermore, it is a meaningful issue to research the other emotions' pupillary response to find whether the pupillary response can reflect the other emotions. In the future, we can use this neural network to make the robots feel the emotion of humans.

References

1. Goodfellow, Bengio & Courville 2016, p. 200.
2. Lu Chen, Tom Gedeon, Md Zakir Hossain, and Sabrina Caldwell: 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, Australia, (2017).
3. Garson, GD: Interpreting Neural Network Connection Weights. In: AI Expert, pp. 47-51, April, (1991).
4. Milne, LK: Feature Selection Using Neural Networks with Contribution Measures. In: Proceedings Australian Conference on Artificial Intelligence AI'95, Canberra, (1995).
5. Wong, PM, Gedeon, TD and Taggart, IJ: An Improved Technique in Porosity Prediction: A Neural Network Approach. In: IEEE Transactions on Geoscience and Remote Sensing, vol. 33, n. 4, pp. 971-980, (1995).
6. Tamás D. Gedeon: "Data Mining Of Inputs: Analysing Magnitude and Functional Measures."
7. Sepp Hochreiter; Jürgen Schmidhuber (1997). "Long short-term memory". Neural Computation. 9 (8): 1735–1780. doi:10.1162/neco.1997.9.8.1735. PMID 9377276. S2CID 1915014.