Anger Authenticity Classification Using LSTM Networks

Peiqing Yang

Research School of Computer Science,
Australian National University,
Acton ACT 2601 Australia,
u6688651@anu.edu.au

Abstract. Anger is an elemental human emotion, and an experiment shows that pupil diameters rather than human perception could indicate the authenticity of anger. In this research, a Long-Short Term Memory neural network is trained to detect whether an anger emotion is genuine or posed by physiological pupil sequences. This paper includes certain methods for adjusting hyperparameters and network structures. Results show that the proposed model could correctly classify most pupil statistics, while a better performance requires a deep hyperparameter and learning rate analysis.

Keywords: Recurrent Neural Network, Long Short-Term Memory, Classification, Anger Detection

1 Introduction

Perceiving anger by expressions is common, while it is hard to distinguish acted anger and real anger emotions since expression could be fake. Research shows that the human pupillary reflections could indicate whether anger emotions are genuine or posed effectively rather than human's perceptions.[1] Pupil diameter becomes a criterion to determine whether the anger is acted.

Artificial Neural Networks (ANN) is a traditional technique that is widely used for statistical modelling in various aspects. Even though how a neural network generates its result is hard to explain, an ANN model could outperform than part of classification machine-learning method like logistic regression.[2, 3] In early years, research shows that the ANN has already been functional to handle problems that other techniques are not able to but under some constrains such as adequate data, fundamental understanding of the problem and adequate processing power.[4] Nowadays, with the fast development of ANN, the Deep Neural Network (DNN) generated from ANNs with deep structures becomes a useful tool in various aspect.[5]

In the past research, a three-layer back-propagation network realized part of functionality to classify anger, while the performance is still not satisfying. This research aims to detect the authenticity of anger emotions with sequential pupil statistics by a specific DNN model, the Long Short-Term Memory (LSTM). In particular, this research contains a series of experiments around network structures, hyper-parameter settings as well as other optimizing tools for network performance improvements.

2 Method

2.1 Previous Research

To find correlations between pupillary diameters and anger authenticity, a three-layer back-propagation neural network was designed for anger classification by summarized pupil statics in the previous research. The network structure is simple, which consists of six input neurons, twelve hidden neurons and one output neuron. This three-layer network could properly classify around 79% of the angry patterns, so inferring anger authenticity with pupillary statistics is reasonable. Because the network performance is restricted by the network structures and limit data, a more complex network structure or different pupil patterns would help anger classification.

2.2 Data and Data Pre-processing

This research uses an experimental dataset that collects 20 participants' pupillary responses to 20 angry stimuli videos.[1] The dataset of anger in the previous research is a processed version of the dataset in this research, and the main difference between these two datasets is that this research's pupil data is sequential. Raw pupillary datasets contain both left and right eyes response of every participant to each video in every 1/60 second. To eliminate the effect of different participants' pupil size, pupillary data is normalized to 0-1 according to every person's maximum and minimum pupil diameter in this research. Besides, because of participants' eye blinks during the experiment, there are some 0 values in the dataset. Since the overall average may not appropriately interpolate in the sequence, 0 values are replaced by either the previous or post 1/60 second data near the blink or the mean of them. Nevertheless, since the left and right eyes pupil responses are not the same, left and right pupillary sequences are combined in their original forms rather than averaged into one sequence.

In this research, the pre-processed dataset consists of pupillary sequences of each participant to each video as well as the label (Genuine/Fake) for each input pattern. This dataset is supposed to be $20(participant) * 20(video)$ long, while in practice, there are some missing attendances of participants, so the final dataset size is 390.

2.3 Neural Network Structure

Recurrent Neural Network (RNN) is a type of ANN that can deal with sequential statistics. In vanilla RNN structure, every activated vector in each time step is preserved, which would lead to an increase of network depth. As a result, the training of RNN becomes difficult because of the exploding or vanishing gradients.[6] The LSTM model is a succeed version of RNN that could handle the depth or vanishing gradients problems by introducing memory cells and gate units in RNN structure.[7] Memory cells provide the capacity of sequence dependencies, which means useful information at the start of the input sequence will pass through the training process. Gate units can control the input and output memory cells flow in hidden layers.
To obtain an efficient anger-detection model, we trained a deep LSTM classification RNN. Inputs are combined left and right pupil diameter sequences and two output cells contain probabilities of genuine and fake labels for input patterns. The final anger authenticity labels for input sequences depend on the output cell that has a higher probability.
In practice, input data are separated into batches for training. When there is one pupil sequence at a time, the LSTM performance is not satisfying because the neural network would adjust for each of them rather than a general pattern. In this case, input patterns are different in time steps for various video lengths, making it hard for the model to do a general angry classification. Therefore, sequences in batches are padded by 0 to have the same time steps as the longest sequence steps, and the model could adjust its weights for a set of patterns instead of a specific sequence.

2.4 Hidden Layer

As shown in Fig. 1, a hidden unit in the LSTM model consists of three gates (input, output and forget) and a memory cell. In this research, an input gate combines the current time step input with last layer' outputs by the Sigmoid function and transfer the result into the memory cell. The forget gate adds weight on the memory cell to decide when the current information will expire, and the output gate controls the output of its hidden neuron. A memory cell contains previous time-series memory and current inputs, and the output of a hidden unit is $OutputGate \otimes tanh(MemoryCell)$.

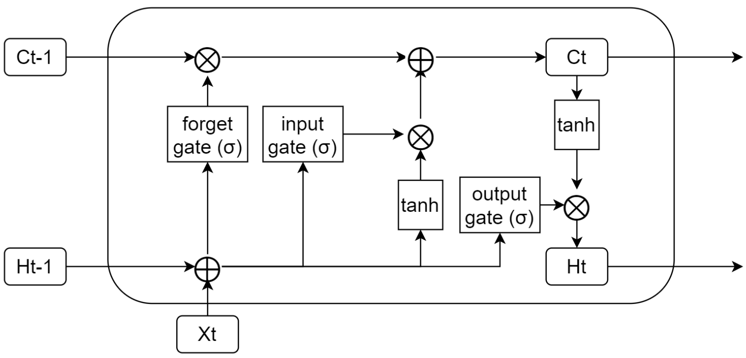


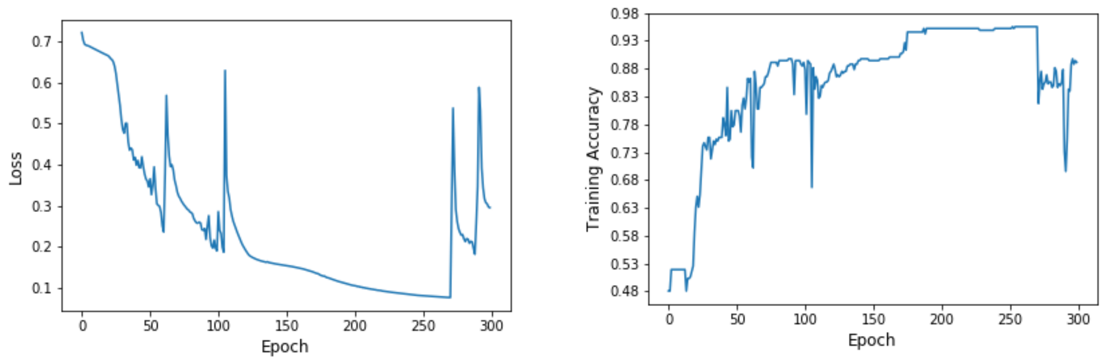
Fig. 1: A LSTM unit for every time step.

For the number of hidden layers, when there is one hidden layer, the training is faster than a deep LSTM, and the highest accuracy is 91.03%. When adding hidden layers, training process slows down and the test accuracy of 2 and 3 hidden layers is the same, 93.6%. As a consequence, there are two hidden layers because of the higher accuracy and shorter training time compared with 1 and 3 hidden layers.
In each hidden layer, there are 12 neurons. When the number of neurons is too large, the network is easy to collapse and the final model would classify all patterns into one class. However, when the hidden neurons' number is too small, the network will lake of training, for example, the accuracy is around 83% when there are 3 hidden neurons.

2.5 Hyper-parameters

Epoch The epoch is set to 265. From Fig. 2, theoretically, the best epoch for this model is 270 because the loss reaches its lowest at this point. However, there is a sharp loss increase right after running 270 epochs, which means drastic changes in loss may happen before 270 epochs when the train test patterns change. As a consequence, the epoch is set to 265 to prevent fluctuations in training.

Learning Rate When the learning rate is 0.0001, the loss decreases slowly during the training process, but when the learning rate is 0.01, the loss tends to fluctuate and make it difficult to reach a desirable result. Because the fluctuation is acceptable without influencing the final model performance when learning rate equals to 0.001 and the learning time is much shorter than learning at 0.0001, the network will first work at the learning rate 0.001. Fig. 2.(a) shows that the loss could fast decrease from 0.7 to 0.2 in 120 epochs at this learning rate. However, an accurate classification requires learning inputs stably, so the learning rate turns to 0.0001 after the loss becomes less than 0.15.



(a) The relation between loss and epoch. The lowest loss is 0.0826 when epoch equals 270.

(b) The relation between training accuracy and epoch. The maximum learning accuracy is 96% when epoch is greater than 225.

Fig. 2: The effect of loss on the anger detection model.

2.6 Optimizer and Loss Function

This experiment compares two basic optimizers: the Stochastic Gradient Descent (SGD) optimizer and the Adam optimizer[8], and the optimizer for this LSTM model uses the Adam algorithm.
An SGD optimizer without momentum updates the model parameters by subtracting $LearningRate*Gradients$, where $Gradients$ are the parameters' gradients. When the learning rate is below 0.001, the SGD optimizer updated the network slowly, while the learning rate turns to 0.01, the SGD optimizer sometimes stayed in a local minimum and no longer updated in real practice.
The Adam algorithm is a powerful optimizer that costs little memory and could be applied to wide ranges of machine-learning problems.[8] In this research, compared with SGD, a simple Adam optimizer is more efficient and stable with different learning rates.
For calculating the loss after every training process and updates the model parameters, this LSTM model uses cross-entropy as the loss function. Because anger detection is a binary classification problem, the loss is $-IsFaked * log(ouput_0) - IsGenuine * log(output_1)$, where $IsFaked$ and $IsGenuine$ are binary and $output_i$ denotes the network's two outputs.

3 Result

Table 1 compares the three-layer back-propagation model in previous research with a summarized pupil dataset and the LSTM model with pupil diameter sequential data in this research. The training and test accuracy of the back-propagation model is different, while the train test accuracy of the LSTM model is similar. Besides, the overall accuracy of the LSTM model is better than that of the three-layer NN.
Table 2 and Table 3 shows the confusion matrix of training and testing set for one of the train test patterns. The training accuracy is 96%, and the test accuracy is 93.6%. This result based on specific train test split dataset that the training data occupies 80% of anger statistics and the rest is the data for testing.

Table 1: Average train and test accuracy (± standard deviation) over 10 repeated runs for BP and LSTM

Model	Training Accuracy (%)	Test Accuracy (%)
Three-Layer NN	94.72 (±4.95)	78.85 (±11.25)
LSTM	95.87 (±0.63)	92.95 (±1.02)

Table 2: Confusion Matrix of Training

Predicted / Actual	Genuine	Posed
Genuine	158	4
Posed	7	138

Table 3: Confusion Matrix of Testing

Predicted / Actual	Genuine	Posed
Genuine	31	1
Posed	4	42

Also, this research carries out a 10-fold cross validation for the anger detection model to validate the stability as well as reliability of the model. The maximum accuracy is 96.77%, the minimum accuracy is 87.10%, the mean accuracy is 92.95%, the standard deviation is 0.0279 and the 95% confidence interval is [0.9085,0.9506].

4 Discussion

4.1 Three-layer Network Performance

In previous research, Table 1 shows that the three-layer back-propagation model for anger classification is overfitting for the training accuracy is greater than the test accuracy by around 10%. But even though the model is overfitted, this network might not learn the training set well, since the test accuracy is not stable and only 90%. The overall underfitting could attribute to the number of hidden neurons or training times that they are not enough for the network to completely learn the dataset. Also, a three-layer network structure might result in overfitting if the number of hidden neurons is too large when there is no other hidden layer to smooth the learning process. However, if there are few hidden neurons, then both of the accuracies of training and testing will decrease. Besides, the output of the three-layer network is not stable for its fluctuating accuracy because of a 0.01 learning rate.

4.2 LSTM Performance

This research generates a reliable anger detection model that could correctly classify most of the sequential physiological pupil data into genuine or fake anger classes. The LSTM structure is stable and capable to handle sequences because most of the experimental accuracy of this model is above 90% according to the 10-fold cross-validation. Since the test and training accuracy is similar and the difference is around 3% as shown in Table 3, the model is slightly overfitting but in an acceptable range.

Compared with the previous back-propagation network with 2000 epochs and 0.01 learning rate, this LSTM has two hidden layers with a different hidden structure, running for 265 epochs with 0.001 learning rate. The LSTM model is stable because a low learning rate is more likely to help the network reach a convergence state while a large rate may easily cause a fluctuation and the final state will have randomness.[9] On the other hand, the 2000 training epochs may lead to the overfitting of the back-propagation network.

The input pattern may also be a reason for the outperformance of the LSTM model. Because in the three-layer network, input patterns are summarized pupil data with 6 features that descript sequences' distribution. But the input pattern for the model in this research preserves every detailed information in a sequence, the neural network could adopt the information it considers helpful for its learning.

4.3 Limitations and Further-works

Even though the LSTM anger classification model performs better than the three-layer network, the classification accuracy could reach at 95% with another model[1], which means it is possible to improve current LSTM model for a higher accuracy than 93.6%.

In current LSTM model, there are some limitations. Firstly, due to the running environment and time limitations, adjusting hyper-parameters like hidden layer neurons is manually randomly and abundant combinations of different parameters have not tested in the current LSTM model. Besides, there is a sharp increase in loss after training the network after 270 epochs, which may cause by the learning rate. When loss is below 0.1, the learning rate 0.0001 might have negative impacts on the Adam optimizer to update its stepsizes Δ_t. In Adam optimizer, $\Delta_t = LearningRate * \hat{m}_t/\sqrt{\hat{v}_t}$, where $\hat{m}_t$ and $\hat{v}_t$ are bias-updated first and second moment estimates.[8] When the epoch increase, the loss decreases below 0.1, and then $\hat{v}_t$ may be close to zero while the $LearningRate$ keeps at 0.0001, leading to the Δ_t increase sharply while the learning rate can not eliminate the sharp increase.

Therefore, different adjusting hyper-parameters method like random search[10] or a specific learning rate decrease by loss may help improve the performance of the LSTM classification model.

5 Conclusion

We have proposed an LSTM model to classify time sequences about human pupil as genuine or fake anger. In this research, different hyper-parameters such as learning rate, epoch and hidden layer numbers are examed. The around 93% test accuracy indicates that this model could effectively detect the authenticity of anger for most pupil reaction statistics. Besides, compared with a three-layer back-propagation network, this model is more stable and functional in anger classification. However, this model has a slightly overfitting problem and the performance is not superb in currently existing models with the same data source. Further improvements require a reasonable hyper-parameter setting method and a changing learning rate for different epoch ranges.

References

1. Chen, L., Gedeon, TD., Hossain, MZ., & Caldwell, S. (2017, November). Are you really angry? Detecting emotion veracity as a proposed tool for interaction. In Proceedings of the 29th Australian Conference on Computer-Human Interaction (pp. 412-416).
2. Andrews, R., Diederich, J., & Tickle, A. B. (1995). Survey and critique of techniques for extracting rules from trained artificial neural networks. Knowledge-based systems, 8(6), 373-389.
3. Dreiseitl, S., & Ohno-Machado, L. (2002). Logistic regression and artificial neural network classification models: a methodology review. Journal of biomedical informatics, 35(5-6), 352-359.
4. Anderson, D., & McNeill, G. (1992). Artificial neural networks technology. Kaman Sciences Corporation, 258(6), 1-83.
5. Liu, W., Wang, Z., Liu, X., Zeng, N., Liu, Y., & Alsaadi, F. E. (2017). A survey of deep neural network architectures and their applications. Neurocomputing, 234, 11-26.
6. Jozefowicz, R., Zaremba, W., & Sutskever, I. (2015, June). An empirical exploration of recurrent network architectures. In International conference on machine learning (pp. 2342-2350).
7. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.
8. Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.
9. Yu, C., Qi, X., Ma, H., He, X., Wang, C., & Zhao, Y. (2020). LLR: Learning learning rates by LSTM for training neural networks. Neurocomputing.
10. Bergstra, J., & Bengio, Y. (2012). Random search for hyper-parameter optimization. Journal of machine learning research, 13(Feb), 281-305.