Predicting the authenticity of anger by applying three-layer neural network, LSTM model and decision tree model

Nannan Ruan

Research School of Computer Science,
Australian National University,
Canberra ACT 2601
U6472765@anu.edu.au

Abstract. Human observable unconscious responses can be used to detect the authenticity of emotions, such as pupil responses. Deep learning models, such as the three-layer neural network, LSTM model, and the decision tree model were trained to predict the authenticity of anger in the video, and the three results were compared with human verbal responses. Finally, we found that the three-layer neural network reached a prediction accuracy of 49.28%, and LSTM model reached a prediction accuracy of 87.5%. The decision tree model enhanced the interpretation capability while having a lower prediction accuracy than LSTM model, which was 67.74%.

Keywords: Three-layer neural network, LSTM, decision tree, analysis of anger authenticity

1 Introduction

1.1 Background

Interacting with emotional faces can focus people's attention strongly, and the inner mental state of people as observers can be affected. Chen, Gedeon, Hossain and Caldwell found that humans have the ability to unconsciously judge the authenticity of emotions, like anger, through the human pupil response, that is, pupil changes. [1]

This ability of humans to unconsciously recognize the authenticity of anger could be applied to the genetic risk research of autism spectrum disorder (ASD) [2] and could be used in chronic condition management and aged care [1].

1.2 Problem Statement and the Method Performed

The problem studied in this paper is to use deep learning models to detect whether humans have the ability to unconsciously judge the authenticity of anger through the human pupil response, that is, pupil changes. When the participants watched the angry video, their pupil response data was collected, preprocessed and used as the input of the three-layer traditional neural network, LSTM model and the decision tree model, and the authenticity of the angry emotion in the video was used as the output. The research in this paper would follow the steps below.

First, the paper would compare the prediction accuracy of the basic three-layer traditional neural network with the result of human verbal responses provided in the research of Chen, Gedeon, Hossain, and Caldwell.

Second, the LSTM model would be applied, and its prediction accuracy would be compared with the results of the first two models. The reason LSTM model was chosen to do further research is that the provided raw data was arranged in time series. In this case, we knew that the human pupil response data collected in each video was not isolated but related to the front and back frames. Therefore, it could be reasonably inferred that the use of a recurrent neural network (RNN) model that could better process sequence information would improve the prediction accuracy.[3] Considering that the basic RNN model was prone to gradient dispersion when facing long sequence data, I decided to use the Long Short Term Memory model (LSTM) with long memory capabilities to process the data to try obtaining higher prediction accuracy results.

Finally, since deep neural networks are habitually regarded as black box models, I decided to use interpretable decision trees to explore the rules of prediction to better analyze the data.

1.3 Hypotheses

Since the decision tree model was generally believed to improve interpretability while reducing the prediction accuracy, it is hypothesized that the prediction accuracy of the decision tree model will be lower than the results of the other two deep learning neural network models, but higher than that of human verbal responses. And among the two deep learning neural network models, the results of the LSTM model should be better than the traditional three-layer neural network model.

1.4 Dataset

The raw dataset was created to test the ability of humans to consciously detect the authenticity of anger, and to further test the ability of humans to unconsciously detect the authenticity of anger with pupil response. [1] This dataset collects the pupil size changes of 22 experimental participants watching 20 processed videos, which are 10 performing anger and 10 real anger videos from YouTube and are processed to eliminate the external influence on the participants. Since pupil changes are human unconscious behaviors, the collected data can effectively reflect the ability of humans to unconsciously detect the authenticity of anger.

The raw data is stored in two worksheets, which are the pupil differences of the left and right eyes of each participant in the 20 videos collected in the experiment.

Since the LSTM model requires the input data to be time series information, it uses time series information of 20 videos. The dataset for LSTM contains 20 columns, each representing the information of a video. The length of each column varies according to the length of the video. There are three worksheets, which can be treated as three features, namely the left eye pupil difference (LeftPD), the right eye pupil difference (RightPD), and the mean pupil difference (MeanPD). The target of the data is the video tag in the top column in each worksheet, where the real anger is 'T' and the performing anger is 'F'.

Table 1. Partial dataset for LSTM model (LeftPD sheet)

T1	T2	…	F10
0.845300262	0.911463586	…	0.778781149
0.845929662	0.913615828	…	0.780947302
0.843716191	0.915253082	…	0.78344886
…	…	…	…

The three-layer neural network and decision tree model use the dataset after data preprocessing and extended features. The size of the dataset is 400×9, and there are a total of 20 videos with statistics of 20 observers for each video. Its first two columns represent observers and videos. Middle six columns are processed statistical data collected in the experiment. They are respectively the mean value of pupil response (Mean), standard deviation of pupil response (Std), observers' left pupil size change after watching videos (Diff1), observers' right pupil size change after watching videos (Diff2), orthogonal linear transformation with first principal component (PCAd1) and orthogonal linear transformation with second principal component (PCAd2). The last column is the video label, where the real anger is labeled as Genuine and the performing anger is Posed.

Table 2. Partial sataset for three-layer neural network and decision tree model

	Video	Mean	Std	Diff1	Diff2	PCAd1	PCAd2	Label
O1	**T1**	0.8431838	0.18523455	0.01220263	0.11444589	0.01739607	0.10219691	Genuine
O2	**T1**	0.8592473	0.10634656	0.003125	0.21364198	0.01705165	0.10749746	Genuine
O3	**T1**	0.84861888	0.12974385	0.00668682	0.05941109	0.01784606	0.10847376	Genuine
…	…	…	…	…	…	…	…	…

1.5 Outline for the following content

Section2. Methodology: State the methods of data preprocessing and the application and evaluation of the three models, namely the three-layer traditional neural network, LSTMs, and the decision tree model.
Section3. Result and Discussion: Display the prediction accuracy of the three models, and the results will be analyzed and discussed.
Section4. Conclusion and Future Work: Summarize the content of this paper and provide the directions for further research.

2 Methodology

2.1 Data Preprocess

2.1.1 Data preprocessing for LSTM model

The **LSTM** model uses the sequence information of the three sheets in the dataset as features, and the label row information in the first row of the sheet as the target.

For the features data, the information of the three sheets is integrated into a feature data frame. Each row of each column has three dimensions, can be represented as [LeftPD, RightPD, MeanPD]. At the same time, for columns with different sequence lengths for the LeftPD and RightPD, delete the redundant data values of the longer columns to make the data lengths of the three dimensions consistent to avoid runtime errors. Moreover, since the data has been standardized and scaled to the [0, 1] interval, no further processing is required.

For the target data, generate target columns based on the video tag pair in each column. The target of the column whose tag contains 'T' is expanded to a column with the same length as the video sequence and padded with 1. As for the column whose tag contains 'F', its corresponding target is expanded to a column with the same length as the video sequence and padded with 0.

Finally, the data columns are randomly arranged, and the dataset is divided into a training set (60%) and a test set (40%) according to the new arranged columns. The 6:4 ratio is used because the number of videos is quite small, which is only 20. I try to expand the test set size while randomly repeating training on the training set videos to make up for the insufficient number of training set videos.

2.1.2 Data preprocessing for baseline neural network and decision tree model

The generation of the **three-layer neural network** and the **decision tree model** uses the data of six statistical columns in the dataset as input and the label column as output. Therefore, we first drop the first two non-statistical data columns, namely the observer and video columns.

For the input columns, it is observed that part of the input data may affect the weight. Therefore, all statistical data is normalized to make the indicators comparable. At the same time, all data is scaled to the [0,1] interval to make data processing convenient and fast. The normalization method used here is Min-Max scaling.

For the output data, since the problem to be dealt with in this paper is a binary classification problem, I relabeled the output label column for the convenience of data processing, where Genuine anger is set to be 1 and Posed anger to be 0.

Finally, randomly arrange the dataset after the above processing. The dataset is divided into test set (80%) and test set (20%).

2.2 Neural Network

This paper uses three layers of traditional neural networks, including input layer, hidden layer and output layer. The number of input neurons in the input layer is the number of input features in the data. Since the study is a binary classification problem, the number of output neurons is 2, namely Genuine (0) and Posed (1).

In the process of adjusting the parameters of the neural network model, the main parameters that need to be adjusted are the number of hidden neurons, the learning rate, and training epochs. According to the influence on the model training results, the first thing I adjust is the learning rate, which is the most important hyperparameter. The learning rate is set in the range of 1-0.0001 for ten tests. I found that the training effect is most obvious when the learning rate is 0.1, and the accuracy of the test is relatively high.

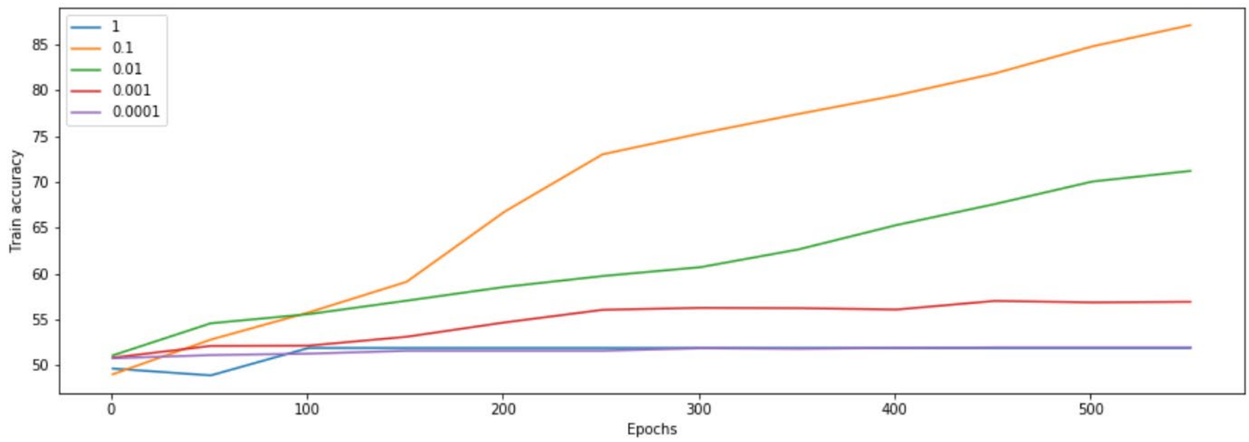

Fig. 1. Average train accuracy of different learning rate of ten tests

The second is the number of hidden neurons. Hidden neurons affect the accuracy and complexity of the model. Increasing the number of hidden neurons to a certain extent can improve the accuracy of model training. However, too many hidden neurons will increase the complexity of the model and may lead to overfitting. After a series of debugging, I chose to set up 20 hidden neurons. Finally, adjusting training epochs. Since the dataset used is not very large, I first tried 1000 training epochs to check the overfitting of the trained model. It is found that the training model is relatively mature when the training epochs is about 400.

The model will use the sigmoid function as the activation function and the cross-entropy function as the loss function. The Sigmoid function normalizes the data through a non-linear method, and the cross-entropy function can be widely used in classification problems. [4] The two are in line with the requirements of our dataset.

After several attempts, I chose to use the Adam optimizer as the optimizer of the baseline neural network. Compared to SGD (Stochastic gradient descent), Adam has a better performance in this dataset. This is reflected in the Adam optimizer that makes the accuracy of the dataset after each round of training have a more obvious improvement compared to the SGD optimizer.

Finally, the parameters of the baseline neural network model are 6 input neurons, 20 hidden neurons, 2 output neurons, 0.1 learning rate, 600 training epochs with printing loss values and accuracies every 50 times, sigmoid activation function, cross entropy loss function, and Adam optimizer.

2.3 LSTMs

Recurrent Neural Network (RNN) is considered to be the most effective method for time series forecasting. [5] The biggest difference between RNN and traditional neural network is that it will bring the previous output result into the next hidden layer for training every time. However, due to the short-term memory and gradient disappearance of RNN, in this article I choose to use the long-term short-term memory (LSTM) model to predict the authenticity of anger. LSTM is based on ordinary RNN and solves the problem of memory and forgetting by adding some multi-threshold gates. [6]

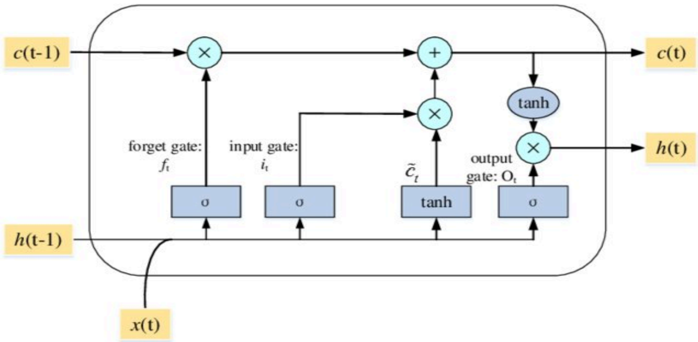

Fig. 2. The structure of the LSTM unit [7]

Since the angry dataset is a time series set, it is reasonable to infer that the LSTM model should be a pretty good way to predict this data set. The input dimension of the LSTM model is 3, which is the three dimensions for each feature, that is LeftPD, RightPD and MeanPD. The output dimension is 2, which are prediction result and target respectively. In each epoch, the columns of the training set are randomly repeated to achieve the expansion of the training set. Then apply the LSTM model to each video sequence to get the prediction result.

The hyperparameters that this LSTM model needs to adjust mainly include the number of hidden neurons, layers of LSTM and dropout. The process of hyperparameter adjustment is similar to the baseline neural network. Different from baseline model, I introduced the dropout parameter for LSTM. Since the training data of this dataset is quite small, over-fitting is prone to occur during the training process. In order to prevent over-fitting, the dropout parameter should be introduced to LSTM, as it is the most effective regularization method to avoid over-fitting in neural networks. Moreover, the use of dropout can also speed up training.

Take the average of train accuracy after ten tests. The parameters of the LSTM model are 3-dimension input, 2-dimension output, 5 hidden neurons, 2 layers of LSTM, 0.4 dropout, 100 training epochs, cross-entropy loss function and Adam optimizer.

2.4 Decision Tree and Decision Tree Rules

Decision tree is a supervised learning algorithm based on if- else rules. Considering that this thesis is studying the two-class problem, the decision tree method is used to predict the input of the test set to get the prediction accuracy and get the corresponding rules. When adjusting the parameters of the decision tree model, the feature selection criterion uses entropy. Although theoretically the gini coefficient will make the decision tree model run faster because it does not require logarithmic operations, our dataset size is not large, and after debugging the criterion, the entropy enables the decision tree model to get a higher accuracy rate. In addition, due to the over-fitting problem easily caused by the small size of the dataset, I have limited the maximum depth of the decision tree to 4 to simplify the model and to extract the main classification rules of the decision tree.

Finally, either the threshold function or the plot function built in the tree structure can be used to extract the classification rules of the decision tree according to the composition of the tree.[8] This research uses the plot function to observe the generated rules more vividly.

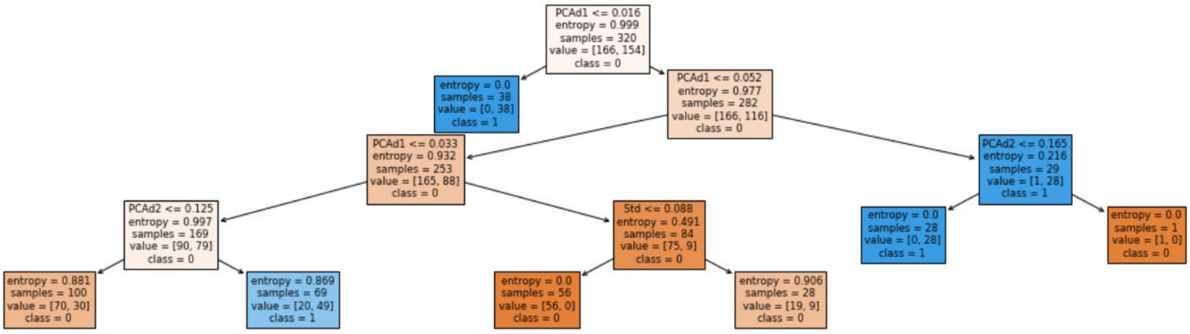

Fig. 3. Generated rules from the decision tree model with max-depth limited to 4

2.5 Evaluation

This paper uses f1 score to evaluate three models. The formula for f1 score is

$$f1 = \frac{2 \times Recall \times Precision}{Recall + Precision}, where\ Recall = \frac{TP}{TP + FN}, Precision = \frac{TP}{TP + FP}$$

TP, FP, FN are the elements of confusion matrix.

Table 3. Confusion matrix

Actual class \ Predicted class	Positive	Negative
Positive	TP	FN
Negative	FP	TN

Recall is the proportion of actual positive observations that are correctly classified.
Precision is the proportion of positively predicted observations that are correctly classified.
F1 score also considers recall and accuracy to weighted average them to get a more accurate evaluation metric.
This study uses the built-in funtion f1_score in the sklearn.metrics package to directly evaluate the three models.

Table 4. F1 score of three models

	F1 Score of Train Set	F1 Score of Test Set
Baseline Neural Network	0.890323	0.709677
LSTM Model	0.824644	0.738279
Decision Tree Model	0.641350	0.838710

We can observe that the baseline neural network and LSTM model have similar f1 scores for train set and test set, and the test set scores are lower than those of train sets. But the scores of the decision tree model are just the opposite. Its score of the test set is much higher than that of the train set.

Considering the size ratio of the train sets to test sets, it might be that the f1 scores of the deep neural network models are positively correlated with the size of the dataset, while the score of decision tree model is negatively correlated.

3 Result and Discussion

Table 5. Prediction accuracy of the three models

Method	Prediction accuracy
Human Verbal Response	60%
Neural Network	55%
LSTM	87.5%
Decision Tree	81.25%

As can be seen from Table 5, the prediction accuracy rate of the dataset using the three-layer baseline neural network is 55%, which is slightly lower than the result of human verbal response of 60% and is much lower than that of the LSTM model (87.5%) and the decision tree model based on the If-else rule (81.25%).

Compared with Zhu's Predicted the authenticity of anger through LSTMs and three-layer neural network and explain result by causal index and characteristic input pattern, the prediction accuracy of the three-layer traditional neuron in the paper, even if I set the same parameters, I still cannot achieve similar results of up to 77%.. This may be because my neural network structure is too simple and needs further optimization.

But the prediction accuracy of LSTM model obtained in this paper can reach 87.5%, which is far higher than the prediction accuracy of the baseline neural network. This can strongly prove the efficiency and superiority of the LSTM model for time series forecasting.

It is observed that the prediction accuracy rate of using the decision tree method has reached 81.25%, which is 21.25% higher than the accuracy rate of the oral response, 26.25% higher than the accuracy rate of the baseline neural network, and only 6.25% lower than the result of LSTM model. This proves that the performance of the decision tree model is good. The if-else rule extracted from the model can determine the threshold of each feature to the prediction result, which increases the explanatory power of the model.

There is a certain discrepancy between the research results and the hypothesis. Among the three models, the performance of the three-layer baseline neural network is much worse than I expected, and the performance of the decision tree model is better. This might be because the structure of the baseline model needs to be optimized, or it might be a result of the small size of the dataset. But one point is certain, that is, the LSTM model has obvious superiority for time series analysis, as its result is better than other models.

4 Conclusion and Future Work

The prediction accuracy of using the LSTM model to process time series is significantly better than other models, and the prediction accuracy obtained by the decision tree model based on if-else rules is much higher than that of human verbal response and baseline neural networks. We can conclude that deep learning models with memory capabilities such as LSTM have high accuracy in predicting the authenticity of anger using human detectable unconscious behavior sequences. And the decision tree model based on if-else rules is also good for predicting the authenticity of anger, the interpretability of the model is greatly increasing at the same time.

In the future, we can use other recurrent neural networks, such as GRU, or use a random forest composed of multiple decision trees to further improve the accuracy of predicting the authenticity of emotions through detectable human unconscious behavior.

Reference

[1] Chen, L., Gedeon, T., Hossain, M. Z., & 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). ACM.

[2] Wendt, F. R., Carvalho, C. M., Pathak, G. A., Gelernter, J., Polimanti, R. (2020, September). Polygenic risk for autism spectrum disorder associates with anger recognition in a neurodevelopment-focused phenome-wide scan of unaffected youths from a population-based cohort. PLoS Genet 16(9): e1009036. https://doi.org/10.1371/journal.pgen.1009036

[3] Li, Y., Zhu, Z., Kong, D., Han, H., Zhao, Y. (2019). EA-LSTM: Evolutionary attention-based LSTM for time series prediction. Knowledge-Based Systems, Volume 181, 2019, 104785, ISSN 0950-7051, https://doi.org/10.1016/j.knosys.2019.05.028

[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] Rumelhart, D., Hinton, G. & Williams, R. Learning representations by back-propagating errors. Nature 323, 533–536 (1986). https://doi.org/10.1038/323533a0

[6] Hochreiter, S., Schmidhuber, J. (1997). Long Short-Term Memory. Neural Comput, 1997, 9 (8): 1735–1780. doi: https://doi.org/10.1162/neco.1997.9.8.1735

[7] Yuan, X., Li, L., Wang, Y. (2019). Nonlinear Dynamic Soft Sensor Modeling With Supervised Long Short-Term Memory Network. IEEE Transactions on Industrial Informatics. PP. 1-1. 10.1109/TII.2019.2902129.

[8] Gedeon, T. D., Turner, H. S. (1993) Explaining student grades predicted by a neural network. Proceedings of 1993 International Conference on Neural Networks (pp. 609-612), Kensinton.