Classifying Posed and Genuine Angers from Observers’ Peripheral Physiology

Zongxin Xiao
Research School of Computer Science, Australian National University
u6709758@anu.edu.au

Abstract. Angers are important facial expressions that provides feelings and attitude to observers. For instance, a polite person will adjust his behaviors when feel his friend is angry. People can express angers from feeling or by posing or acting the anger. This report uses one of the observers’ peripheral physiology signal, Pupillary Response (PR), to classify people’ real(genuine) and posed(histrionic) angers. Twenty anger videos shown to twenty observers while asking them to identify and recording their pupillary response. This report using Bimodal Distribution Removal technique implement several artificial neural networks included recurrent neural network of LSTM. The classification accuracy was higher than human verbal response.

Keywords: bimodal distribution removal, anger recognition, pupil dilation, pupillary response, physiological signals, machine classification, RNN, LSTM

1 Introduction

Facial expression is important component of human communication and behaviors. Emotional expression can influence human cognition and judgement. The observers will be also affected at the same time [4]. When emotions run amok, negotiators lose perspective and make serious mistakes or perform poorly [1]. For some special occupations requiring calm mental state, such surgeon, astronaut and negotiator, it’s necessary to detect their real emotions to reduce risk of mission failed. Besides, detecting expressions also help monitor the emotional state of mental patient.

The research in this paper is based on the data set of twenty videos shown to twenty observers [2]. There are 20*20=400 patterns with genuine or posed label. After data normalization and pre-processing, six attributes of the dataset are shown in Figure 1. Time-series features are also used to classify pose and genuine anger. Figure 2 shows two angers have different patterns.

1. Mean of pupil dilation(mm)
2. Standard Deviation of pupil dilation(mm)
3. diff1 of pupil dilation(mm)
4. diff2 of pupil dilation(mm)
5. PCAd1 of pupil dilation(mm)
6. PCAd2 of pupil dilation(mm)

Figure 1. Six attributes of the dataset

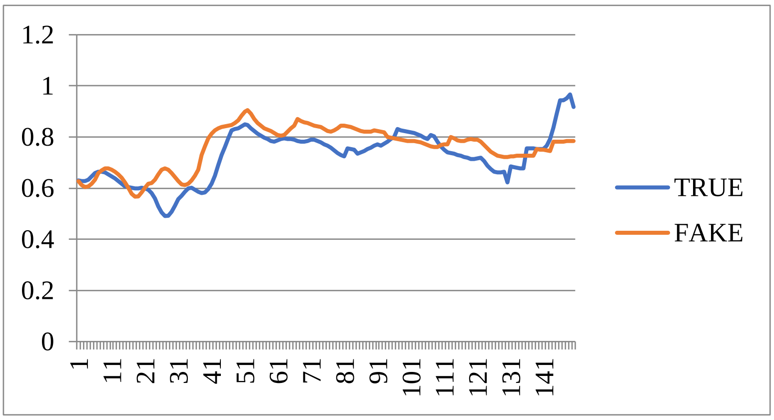

Figure 2. True and Fake time-series with normalization 0-1

I am curious about the using machine learning method to solve psychology problems, so choose the anger dataset. Chen et al. (2017) had used the same data set in research paper [2], which is the original source of the anger data set. In that paper, they didn't describe the classification method in detail. This report will discuss the method and implement the algorithm.

In order to solve the binary classification problem of this data set, firstly, I used Bimodal Distribution Removal to detect the outliers. It aims to clean up noisy and improve generalization performance. After, some artificial neural networks were constructed, adjusting the network structure and parameters converges the network training, then examine on the test sets.

2 Method

2.1 Bimodal Distribution Removal (BDR)

For improve generalization, some methods for cleaning up noisy had to be conducted. Pruning, Dynamic node Dow, Extra term in performance function, Cross validation and Outlier removal have been introduced bias. Noisy patterns seem to have two negative effect on training [6]. One is increasing the time consume of training, the other leads to overfitting problems. This report uses Bimodal Distribution Removal (BDR) [6] to detect outliers for reducing training time and improve generalization.

2.2 Implementation of Artificial Neural Network (ANN)

This paper constructed three neural networks. All neural network consists of an input layer with six nodes and an output layer with one node. Four layers network with 50 hidden nodes in each layer has the best classification performance. Then Three layers network with 50 hidden nodes in each layer has a bit weaker performance. The Two layers network with 10 hidden nodes have the weakest performance. However, it still does better job than verbal response. All networks use back-propagation model and Cross-entropy loss as loss function. Four layers network uses ReLu function for hidden neurons and batch to solve vanishing gradient problem. Sigmoid function was applied in two layers network. Although it exists efficiency problem, data set in this report is small so that the influence can be ignored. Adaptive Moment Estimation method and Stochastic Gradient Descent method are used as optimizers for comparison.

2.3 Implementation of Long short-term memory (LSTM)

Traditional neural network architecture has been unable to solve sequence problems. The recurrent neural network (RNN) can "inherited" input information to the output through the loop. RNN is not good for the anger dataset timeseries. It has the problem of gradient disappearing. This report conducts Long short-term memory (LSTM) to implement classification. In addition to using preprocessed data, normalize raw data to expend dataset. Batch size is 1.0. Activation function adopts softsign function. RMSProp, AdaGrad and momentum（Nesterovs）is tried in the model. Gradient clipping is applied in the experiments. Regularization l1 and l2 are used to prevent overfitting.

2.2.1 Adam Algorithm [3]

Adam Algorithm is a popular gradient descent method currently. It can be considered as a combine of RMSprop and Momentum. Similar to RMSprop using exponential moving average for the second-order momentum, Adam uses the exponential moving average for the first-order momentum. When Adam is selected as the optimizer, the iterations have a smooth process and a shorter converge time than SGD algorithm. The three layers network cost 3000 times to converge reaching 100% accuracy and the loss decreased to 0.0035.

2.2.2 Stochastic Gradient Descent (SGD) Algorithm [3]

Stochastic gradient descent (SGD) is classic gradient descent method. When the convergence is not enough, the learning rate can be slightly reduced to improve the convergence. The model sometimes cannot convergence well using Adam Algorithm. SGD can get rid of saddle point by using momentum. The convergence speed of SGD will be slower, but the final convergence result is generally better.

2.3 Neural Network Reduction

Neural network reduction technique simplifies the network structure and reduce computing cost through remove insignificant neurons [7]. If the vector angle of pair units is less than 10 degree, one of them can be deleted for similarity feature. When the vector angle of pair units is larger than 170 degree, both of them can be deleted for complementary feature.

2.4 Evaluation method

The pre-processed dataset was divided into two parts. 80% data were used as training set and 20% data were used as validation set. Then randomly select a part of the sample in the raw data and normalize it as the test set. Two performance evaluation are applied in this report. Accuracy of different neural networks is compared. Besides, calculate the sum time of training and testing process to compare the performance.

3 Results and Discussion

3.1 Detecting Outliers

The first picture in Figure 3-1 is error distribution in epoch 0. It’ s typical bimodal distribution. After 12 times iterations, the normalized variance of the errors in whole data set was reduced below 0.1. The lower peak vanished and the other peak move to left and divided into two peaks.

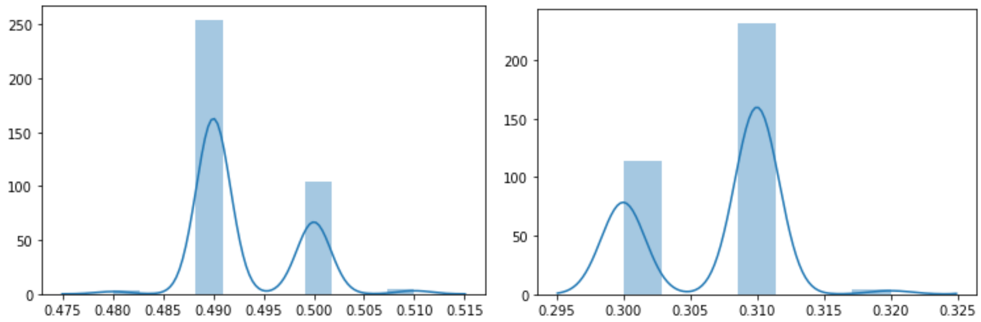

Figure 3-1. Error distribution at epoch 0 (left) & normalized variance v_{ts} errors below 0.1(right)

Remove all patterns with error $\geq \bar{\delta}_{ss} + \alpha\sigma_{ss}$ in the training set, where α was chose 0.5. Figure 3-2 shows the trend that error distribution converges to well-distribution through BDR process. It's good at fixing bias-variance dilemma and speed up convergence.

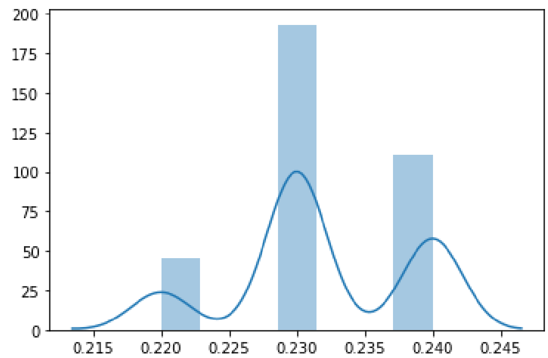

Figure 3-2. Error distribution at epoch 50

3.2 Balanced dataset

The dataset contains 200 genuine samples and 200 posed samples. Every video has ten samples. Random sampling is divided into training set and test set to ensure datasets balanced. Figure 4 show the details.

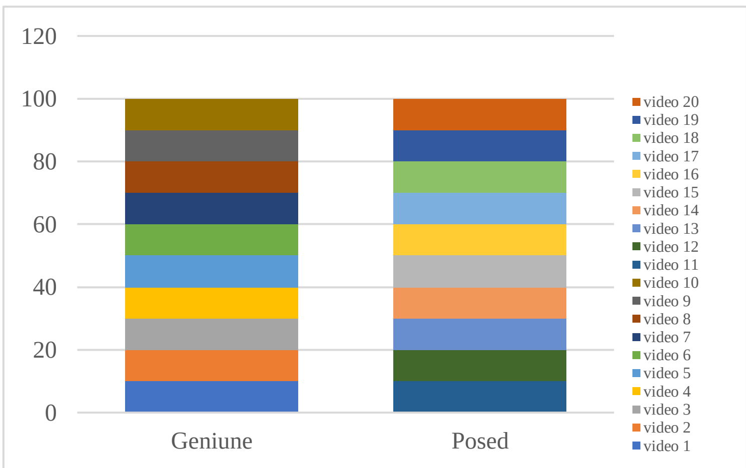

Figure 4. Balanced dataset

3.3 Parameter adjustment

It's important to initialize parameter. Random methods may slow down the convergence speed, affect the convergence results, and even cause a series of problems such as Nan. Uniform distribution and Gaussian distribution initialization have similar result. SVD initialization is better on RNN (Saxe, 2013).

Jozefowicz (2015) argues that the bias of forget gat initializing with 1 or larger can make better performance. I set it 1.0 in the experiments which improve convergence speed.

Adam and Adadelta experiments are not as effective as SGD. The convergence speed of SGD will be slower, but the final convergence results are generally better. Use Ada to run first, when it almost converges, change to SGD to continue training. Adadelta works better on classification problems.

Gradient clipping is important. When the model is without gradient clipping, sometimes the loss becomes nan after training for a while.

3.4 Comparison of results with different neural network model

Three neural networks were designed in this paper. Three layers and four layers network have better accuracy. The hardware is Intel Core M (5Y31). The results can be seen in Table 5:

NN	Validation Accuracy	Test Accuracy	Test Time(second)
Two layers ANN	68.7%	51.4%	7
Three layers ANN	85.5%	72.5%	14
Four layers ANN	92.5%	65.3%	25
LSTM	93.3%	82.0%	58

Table 5. Accuracy and consume time between several NN models

3.4 Comparison with others' work

Chen et al. (2019) had used the same data set in their research paper [2]. In that paper, the machine classification through pupillary response has 95% accuracy on anger detecting. This result is similar to my work, but my classification's performance is not very stable.

4 Conclusion and Future Work

The neural networks conducted in this report work well. Machine classification through pupillary have 72.5% accuracy while verbal response has only 60% accuracy [2]. Bimodal Distribution Removal has good effect on solving bias-variance dilemma. It improves generalization ability cause the models achieve good accuracy in the test set. In the future, it's a good point to collect heartbeat information during the experiments. Combining heartbeat information and pupillary implements a multi-timeseries classification. It should have better generalization ability. 1D CNN also can be applied into timeseries classification and prediction. We can compare the performance between two methods. Besides, more psychology experiment can be combined with deep learning and other machine learning methods.

References

1. Adler, R. S., Rosen, B., & Silverstein, E. M. (1998). Emotions in negotiation: How to manage fear and anger. Negotiation journal, 14(2), 161-179.
2. 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).
3. Kingma, D. P., Ba, J.: Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. (2014)
4. Lerner, J. S., & Tiedens, L. Z. (2006). Portrait of the angry decision maker: How appraisal tendencies shape anger's influence on cognition. Journal of behavioral decision making, 19(2), 115-137.
5. Li, M., Zhang, T., Chen, Y., Smola, A. J.: Efficient mini-batch training for stochastic optimization. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining (pp. 661-670). ACM. (2014)
6. Slade, P., & Gedeon, T. D. (1993, June). Bimodal distribution removal. In International Workshop on Artificial Neural Networks (pp. 249-254). Springer, Berlin, Heidelberg.
7. Wilson, D. R., & Martinez, T. R.: Reduction techniques for instance-based learning algorithms. Machine learning, 38(3), 257-286. (2000)
8. Glorot, X., & Bengio, Y. (2010, March). Understanding the difficulty of training deep feedforward neural networks. In Proceedings of the thirteenth international conference on artificial intelligence and statistics (pp. 249-256).
9. He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 770-778).
10. Saxe, A. M., McClelland, J. L., & Ganguli, S. (2013). Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. arXiv preprint arXiv:1312.6120.
11. Jozefowicz, R., Zaremba, W., & Sutskever, I. (2015, June). An empirical exploration of recurrent network architectures. In International conference on machine learning (pp. 2342-2350).