Classifying Genuine and Posed Emotion: Comparing LSTM Networks, Fuzzy Logic Classification and fully-connected Neural Networks

Kangyu Chen

Research School of Computer Science, Australian National University
Canberra Australia
u6823506@anu.edu.au

Abstract. Understanding human's emotion is an important task in the research field of human-computer interaction. People could easily distinguish the genuine and posed anger while it could be difficult for computers to judge the authenticity of anger. In this paper, apart from the fully-connected Neural Networks built in the previous research, methods of LSTM Networks and Fuzzy logic classification are also built for telling the posed angers from genuine ones using limited data. Then we compared the performance of these three popular classification methods. Carefully designed modifications have been made on these models to increase the performance.

Keywords: Recurrent Neural Network, fully-connected Neural Network, Fuzzy Logic Classification, Long Short-Term Memory, Anger Detection.

1 Introduction

Emotions are psychological states which could lead to physical and physiological changes, affecting people's reactions and behaviors. Anger is one of the most common and natural emotion in our life. Understanding human's emotion is a vital task for human-computer interaction. Anger detection is one of the research fields which interests researchers a lot. It is proved that the neural network could achieve an accuracy to 99.69% to identify the veracity of human emotional expressions [1].

Sometimes, people would act anger expression when they do not carry genuine feelings, only to manipulate the perceivers. In this work, we would like to know how well the computer could understand the real human emotions, specially the posed anger and genuine anger. The problem of anger detection could be described as two-class classification problem. In the past research, some functions of anger classification are realized by using the simple three-layer fully-connected Neural Networks, although the performance is still unsatisfactory. In this experiment, the performance is improved by long-short term memory (LSTM) network, which is one of most popular recurrent neural network (RNN) architectures. Furthermore, the fuzzy c-means (FCM) method of classification is developed to classify the genuine and posed anger. This research contains a series of experiments around hyper-parameter tuning and network structures together with the comparison of performance between different neural networks models and classification methods.

Based on previous works, this paper develops the LSTM network model and fuzzy c-means method of classification method to solve the anger classification problem using the dataset from paper [2]. Then compare the performance of implemented methods with simple fully-connected neural networks method used in Chen's paper [4].

2 Methods

The genuine/posed anger recognition could be formulated as a two-class classification problem. In this paper, I have developed the LSTM network model and the fuzzy c-means method of classification to deal with this problem. Then we could figure out the performance of different network models and classification methods. The raw datasets from paper [2] is preprocessed and then normalized. In this section, we will briefly explain the dataset used in this paper (Section 2.1). Then talk about the design of LSTM network model (Section 2.2) and explain the loss function and optimizer (Section 2.3). Then introduce the fuzzy c-means method of classification (Section 2.4).

2.1 Dataset Description

The dataset is collected based on participants' reactions to different videos. Reactions to 20 different videos were collected, where 10 videos would lead to genuine anger and the rest lead to posed anger. The dataset is preprocessed and for each participant, the dataset consists of seven data which are Mean, Std, Diff1, Diff2, PCAd1, PCAd2 and Label. The dataset contains 400 patterns, the Mean is the mean value of pupillary diameter and Std is the standard deviation of pupillary diameters. The Diff is the difference between maximum and minimum pupillary diameter, and the PCAd1 and

PCAd2 are the principal component analysis results of the experiment. The label is the authenticity of anger. For this experiment, the dataset is spilt into the training and testing set. The ratio between training and testing set is 7:3 which means the training set contains 280 patterns and the testing set contains 120 patterns.

2.2 Long-short Term Memory (LSTM) network model

The Long-short Term Memory (LSTM) is a recurrent neural network (RNN) architecture used in deep learning [7]. The RNN model performs better than the normal fully-connected neural network since the RNN model provides feedbacks which could learn the short-term dependencies. However, when the input sequence is very long, the normal RNN model could have long-term dependency problems. Compared with normal RNN model, the LSTM network model could maintain information in memory for long periods of time which makes the LSTM model could have a better performance than RNN model when processing very long sequences.

An LSTM unit should contain four components: the memory cell, forget gate, input gate and output gate. The memory cell is used to save the information of hidden states. The input gates determine the extent of input goes into the current hidden state based on current unit, hidden unit and the previous content of memory cell. Similarly, the output gates determine the output coming out of the memory cell and into the current hidden unit. The forget gates determines to what extent to forget the previous data.

The figure of structure of the LSTM network is shown in Fig 1.

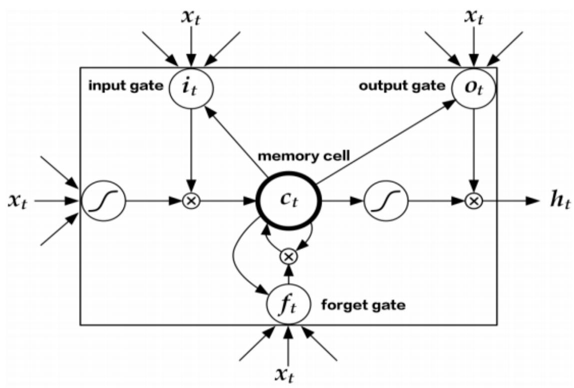

Fig. 1. The structure of LSTM network. The x_t represents input. The i_t, f_t and O_t represents input gates, forget gates and output gates respectively. The c_t represents memory cell. The h_t represents hidden output.

2.3 The Loss function and Optimizer

The loss function is used to compute the difference between the actual output and the predicted output. The cross entropy and mean square error are common loss functions in the deep learning model. If a problem is a regression problem, we usually use mean square error as the loss function. If it is a classification problem, we usually use cross entropy. In this experiment, we aimed to classify the posed anger from genuine ones, so that I choose to use cross entropy as loss function. The cross-entropy loss could be described as:

$$CEloss(x, class) = -x[class] + \log\left(\sum_j \exp\left(x[j]\right)\right)$$

The $x[j]$ is the predicted label, the $x[class]$ is the original label.

The optimizers are algorithms used to change the weights or learning rate of neural network to reduce losses. In this experiment, I choose to use the first-order Adam optimizer algorithm as optimizer. That is because the Adam optimizer is a replacement optimization algorithm for stochastic gradient descent for training the deep learning models that costs little memory. Compare with other optimization algorithms like Stochastic gradient descent optimizer, we could find that the Adam optimizer is more efficient and stable which produces better results.

2.4 Fuzzy c-means (FCM) method of classification

The fuzzy c-means (FCM) is a method of clustering which allows one piece of data to belong to two or more clusters [5]. The fuzzy c-means clustering could be very powerful compared with traditional clustering method where each point should be assigned to the exact label. This algorithm gets the membership degree of sample data points to centers and then determines the class of the sample points, so as to achieve the purpose of classification of the sample data points. Given dataset $X = \{x_1, x_2,\ x_3,\ \dots, x_n\}$, k is the class number, $m_j (j = 1, 2, .., k)$ is the centers of each cluster. $\mu_j(x_i)$ is the membership function of i th sample data to j th class. Then the clustering loss function based on membership function can be written as [5]：

$$J_f = \sum_{j=1}^{c} \sum_{i=1}^{n} [\mu_j(x_i)]^b \| x_i - m_j \|$$

Where b is a constant, usually we take b as 2.

When the fuzzy c-means algorithm converges, the clustering centers and membership values of each sample are obtained, thus the fuzzy clustering classification is completed.

3 Results and Discussion

In this part, I will evaluate the performance of simple fully-connected neural network (section 3.1) and the LSTM network (section 3.2) as well as the fuzzy c-means method of classification (section 3.3). Then compare and discuss the effectiveness of different methods and find the factors (hyper-parameters) that might affect the performance of models.

3.1 The prediction result of simple fully-connected neural network

The hyper-parameters play important role in neural networks. For this experiment, I would like to change the size of hidden layer and different loss function (Mean square error/ Cross-entropy loss) for the fully-connected neural network. The Adam optimizer is used to update the parameters of network. To prevent overfitting, I choose to stop the training when the training loss does not change. Five experiments have been done and the average train accuracy and test accuracy are computed. The results could be found in table 1.

Table 1. Result of simple fully-connected neural network

Loss function	Train accuracy	Test accuracy
Hidden layer = 16		
Mean Square Error	60%	37.5%
Cross-Entropy Loss	61.43%	47.33%
Hidden layer = 32		
Mean Square Error	69.38%	62.5%
Cross-Entropy Loss	75.71%	73.33%
Hidden layer = 48		
Mean Square Error	55.36%	47.57%
Cross-Entropy Loss	67.14%	52.86%

From the results, it could be found that the number of hidden layer equals 32 gives the best testing accuracy. Also, we find that the cross-entropy loss performs slightly better than mean square loss which means the cross-entropy loss is more suitable for solving classification questions.

3.2 The prediction result of LSTM network

For the experiment of testing performance of LSTM network model, the Adam optimizer is used to update the parameters of network. Since the cross-entropy loss function is more suitable for classification problem, so that I choose to use the cross-entropy loss as the loss function. In this experiment, I would like to tune the batch size for LSTM model with one hidden layer. The experiments have been done and the averaged test accuracies are computed. The results could be found in table 2. Figure 2. shows how training loss changes as epoch number changes.

Table 2. Results of LSTM network with different batch sizes

Batch size	Test accuracy
4	50%
8	75%
10	80%

From the table 2, we could find that when the batch size is 10, the model produces best performance. For this experiment, I choose the learning rate of this LSTM model as 0.01 and the total epochs as 250. Figure 2. shows how training loss changes as epoch number changes. Figure 3. shows how training accuracy changes as epoch number changes.

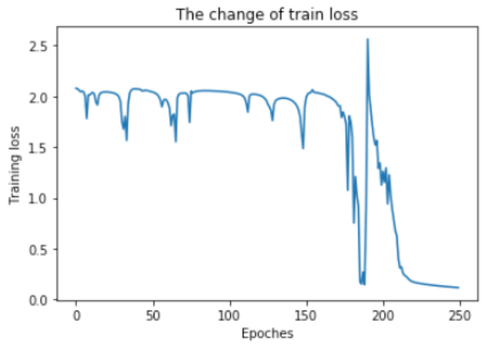

Fig. 2. The relation between training loss and epochs. The lowest loss is 0.116 when epoch equals 250.

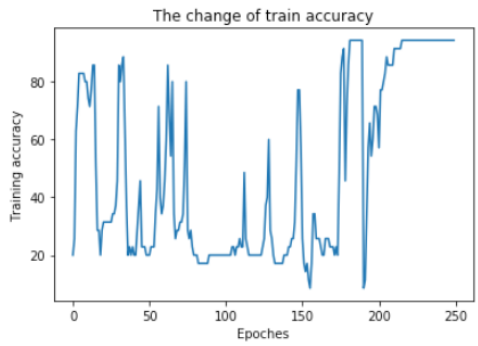

Fig. 3. The relation between training accuracy and epochs.

The maximum training accuracy is 94.29% when epoch is greater than 200.

3.3 The prediction result of fuzzy c-means method of classification

For the experiment of testing performance of fuzzy c-means method of classification, I would like to tune the number of clusters for the c-means algorithm. Then compute the train accuracy and test accuracy on training set and testing set. The results could be found in table 3.

Table 3. Results of fuzzy c-means method of classification with different number of clusters

Number of clusters	Train accuracy	Test accuracy
2	51.78%	50.0%
4	55%	51.66%
8	58.57%	46.66%

From the table 2, we could find that when the number of clusters is 4, the algorithm produces best test accuracy. We could also find that when we use 8 as number of clusters, the accuracy on training set increases while the accuracy on test set decreases. My understanding is that too many clusters might cause the overfitting problem.

3.4 Discussions

From the experiment results, we could find that the LSTM network performs better than the simple fully-connected neural network and fuzzy c-means method of classification. This research shows that the LSTM network model is a stable and reliable anger recognition model which could classify majority of the posed anger from the genuine ones.

Furthermore, from the experiment results of the simple fully-connected neural network model, we could find that the model used for classifying genuine/posed anger is overfitting and the training accuracy is normally higher than the test accuracy by approximately 15%, which means the model is not stable and the model depends too much on training dataset. Compared with fully-connected neural network model, the LSTM network model is more stable and suitable for solving the classification problem since the experimental test accuracies are much higher than the test accuracies for other models and algorithm. With proper values of learning rate and epochs, the LSTM network model could achieve the best performance.

Although the fuzzy c-means method of classification does not perform better than the LSTM network model and simple fully-connected neural network model in classification problem, but from the experiment results, we could find that the training accuracy and testing accuracy are close. This indicates that the fuzzy c-means method of classification has good generalization ability and has minor possibility to overfit.

From paper [2] we could find that the accuracy for human to predict the emotion correctly is around 60%. From the experiment result, we could find that the accuracy for LSTM network model on testing dataset is much higher than the human's prediction. The simple fully-connected neural network model could normally perform better than human if the dataset is perfectly distributed. The fuzzy c-means method of classification does not perform as well as human's prediction.

4 Conclusion and Future work

In this work, we compare the performance of three classification methods on the anger recognition dataset. In conclusion, the LSTM network model performs better than the simple fully-connected neural network and fuzzy c-means method of classification. The accuracy of a fully trained LSTM network model with one hidden layer could reach 80%. However, compared with the previous research in paper [2] whose techniques could achieve 95% accuracy on anger recognition dataset, more methods could be used to improve model and boost the accuracy. In the future, we could improve the method of hyper-parameter tuning and find the suitable hyper-parameters which could achieve the optimal performance of these models. What's more, we could do research with some other efficient classifiers like random decision forests algorithm [6] which could be a good choice to improve the accuracy of classification.

References

1. Qin, Z., Gedeon, T., Caldwell, S.: Neural networks assist crowd predictions in discerning the veracity of emotional expressions. In: International Conference on Neural Information Processing. pp. 205–216. Springer (2018)
2. Chen, L., Gedeon, T., Hossain, M. Z., & Caldwell, S. 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. (2017)
3. Milne, Linda, Tom Gedeon, and Andrew Skidmore. "Classifying Dry Sclerophyll Forest from Augmented Satellite Data : Comparing Neural Network, Decision Tree and Maximum Likelihood." (1995)
4. Chen, K.: "Classifying Genuine and Posed Emotion: Comparing Neural Network, Decision Tree Classification, 4th ANU Bioinspired Computing Conference". (2021)
5. Dunn, J.. "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters." (1973).
6. Tin Kam Ho. 1995. Random decision forests. In Proceedings of the Third International Conference on Document Analysis and Recognition (Volume 1) - Volume 1 (ICDAR '95). IEEE Computer Society, USA, 278.
7. Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long Short-Term Memory. Neural Comput. 9, 8 (November 15, 1997), 1735–1780. DOI:https://doi.org/10.1162/neco.1997.9.8.1735