Distinguishing genuine and fake anger emotion with fuzzy clustering for data augmentation

Yaozhong Liu[1]

Research School of Computer Science, Australian National University, Acton ACT 2601, AUSTRALIA
u6686404@anu.edu.au

Abstract. Emotion recognition is one of the meaningful tasks that attract significant attention in the computer vision community in the past decade. Whether can classify the emotion correctly has become a metric to measure the machine possesses the humanized characteristic. Emotion recognition lies in the intersection of applications like human-computer interaction and crowd analytics [9]. To distinguish genuine and fake anger images, some previous research using complex machine classification has reached 95 percent in prediction accuracy. Rather than design a complex model, in this research, we plan to use fuzzy clustering to create more features to train a neural network model to deal with facial emotion recognition problems. Here, we adopt the BDNN approach to train a neural network while using fuzzy clustering to create new features to augment our input data. We believe augmented input data by fuzzy clustering will improve the classification accuracy when differentiating genuine and acted anger images. After the augmentation of data's feature by fuzzy clustering, we find our best model exhibits 75.4 percent accuracy in the test set, which is 3% higher than the model trained by the same data without feature augmentation. This suggests that using fuzzy clustering for data augmentation benefits anger classification task and it is promising to be applied to other classification tasks.

Keywords: Bidirectional Neural Networks · fuzzy logic · fuzzy clustering · anger recognition

1 Introduction

1.1 Report Statement

Analyzing the human's facial expression is an essential part of the artificial intelligence domain since it can be regarded as the symbol that machines possess humanized characteristics. Face emotions significantly expose the expresser's internal mental state [2], and due to its potential application to human-machine interaction [3], it is pretty meaningful to understand the facial expression correctly and precisely. However, aside from recognizing the emotion type, the difficulty we may encounter is that some of the facial expressions are not veracious, which increases the error when recognizing the facial expression. Understanding the real emotion on people's faces and avoiding manipulation by acted emotions becomes critical. According to the findings of previous researchers, the pupillary response of a person in viewing two types of anger stimuli can achieve remarkable accuracy compare to his verbal response. This report proposes a deep learning method using Bidirectional Neural Networks (BDNN) [2] trained by participants' pupillary response data to classify genuine and acted anger images. The dataset we used in this work comes from an experiment that researchers at Australian National University conducted consisting of pupillary responses of the participants when watching images of two types of anger [1]. Besides simply using BDNN to do classification, we also implement fuzzy clustering and add the membership value of each cluster as the new features to every training data.

1.2 Report Outline

The report consists of 4 chapters and is organized as follows:

- **Chapter 2 Method** presents the methods we used in this research, including the data processing method and the fuzzy clustering method to augment input features as well as the strategy of training and evaluating different models(the BDNN model trained by augmented data and the BDNN model trained without augmented data).
- **Chapter 3 Results and Discussion** will display the classification performance of different models with straightforward visualizations using tables and plots and the effect of our processing techniques is analyzed to draw final conclusion.
- **Chapter 4 Conclusion and Future Work** firstly gives an account of the findings in our research, then we demonstrates the limitations of our design and introduce the implication of our research. At last, we conduct future work about possible direction to improve our research and the next step of this work will be presented .

2 Method

In this chapter, we will illustrate the techniques we used in the stage of processing data, the introduction of fuzzy logic and the BDNN model we used in the research.

2.1 Data processing and Feature augmentation

Normalization and transformation: The dataset we use is the same as the one in this paper [1]. The total number of the data is 400, with half of the data represent posed anger and the other half represent genuine anger. For each training instance, it has six features and the type of these features are continuous value, including `Mean` (The mean of in pupillary response), `Std` (The standard deviation of in pupillary response), `Diff1` (The change of left pupillary size after watching a video), `Diff2` (The change of right pupillary size after watching a video), `PCAd1` (An orthogonal linear transformation with first principal component), `PCAd2` (An orthogonal linear transformation with second principal component), and associated with the `label` indicating the emotion expression is genuine or posed. Since these features are measured in different measurement units, we notice that the value range of different features is quite different. In order to avoid the situation that the model is dominated by a certain particular feature and increase the stability of the model, we normalize every feature in the input data by L2 Normalization.

$$p_i' = \frac{p_i}{\sqrt{\sum_{i=1}^{N} {p_i}^2}} \tag{1}$$

p_i represents the value of a certain feature of the i-th training instance. p_i' is the value after normalization.
The L2 Normalization has a wide range of applications in image classification tasks, and the improvement of accuracy has been witnessed compared to the training data that is not normalized [10]. This research finds that under the premise of fixed hyperparameters, the model trained by normalized training data achieves 10 percent accuracy higher than the model trained by non-normalized training data. Recall that our research aims to do binary classification on genuine and acted anger images, and the labels we have in the data are two categorical values, genuine and posed. To fit the goal of the task, we change them into 0 and 1 since the value type of the rest of the features is all numerical. After using L2-Normalization on the features of each training instance and doing the transformation of the labels, we randomly split the data representing two kinds of images (genuine or acted) at the ratio of 7 :3 respectively and combine the larger parts of these two categories into training data while making the residual data be the test set. Therefore, whether in the training set and test set, the number of data of genuine anger is equivalent to the number of the data of acted anger.

Augment feature through fuzzy clustering: Fuzzy c-means clustering is a form of clustering in which each data point can belong to more than one cluster. The data points assigned to the same cluster are as similar as possible, while the data points belonging to different clusters are as dissimilar as possible. The membership values are associated with every data point, indicating the degree to which data points belong to each cluster. At the end of the clustering, we will get the membership values of each data point for each cluster, and the highest membership value determines the cluster it belongs to. In some previous research, the result of the fuzzy clustering is added to training data to provide extra information for model training. By applying fuzzy clustering, the performance of the image classification model increase about ten percent.[11]. Additionally, in some features such as `Diff1`, the values of genuine images and acted images are overlapped. Due to the peculiarity of fuzzy clustering, it will assign membership for every cluster to each data, which is suitable to data overlapped problem[12].

Inspired by this idea, we do not care which cluster the data belongs to but use the membership values of different clusters in the result of fuzzy clustering as the new features to augment the training data. The number of extra features for every data point is corresponding to the number of clusters that we set before clustering, i.e., if we set the number of clusters equals two, then we will have two extra features for every data points(the first feature value represents the degree of this point belongs to `cluster 1`, the second feature value represents the degree of this point belongs to `cluster 2`). Because we want our model to do binary classification, then we start by setting the number of clusters equals two and increase to four to find the effect it causes to the classification accuracy.

2.2 Training method

After the data pre-processing and feature augmentation, we feed the training data to the BDNN model we built. BDNN consists of two different neural networks. One of them is a feedforward learning neural network using backpropagation to optimize, while the other is used for backward learning while utilizing forward propagation. BDNN allows outputs to inputs mapping, which means the output of the feedforward neural network will be used as input to the backward learning neural network. BDNN learns from both directions by continuously alternating forward learning and backward learning, and the weights are shared in both directions. We calculate the loss of feedforward learning and backward learning and update the weights when the model finishes one back and forth training.

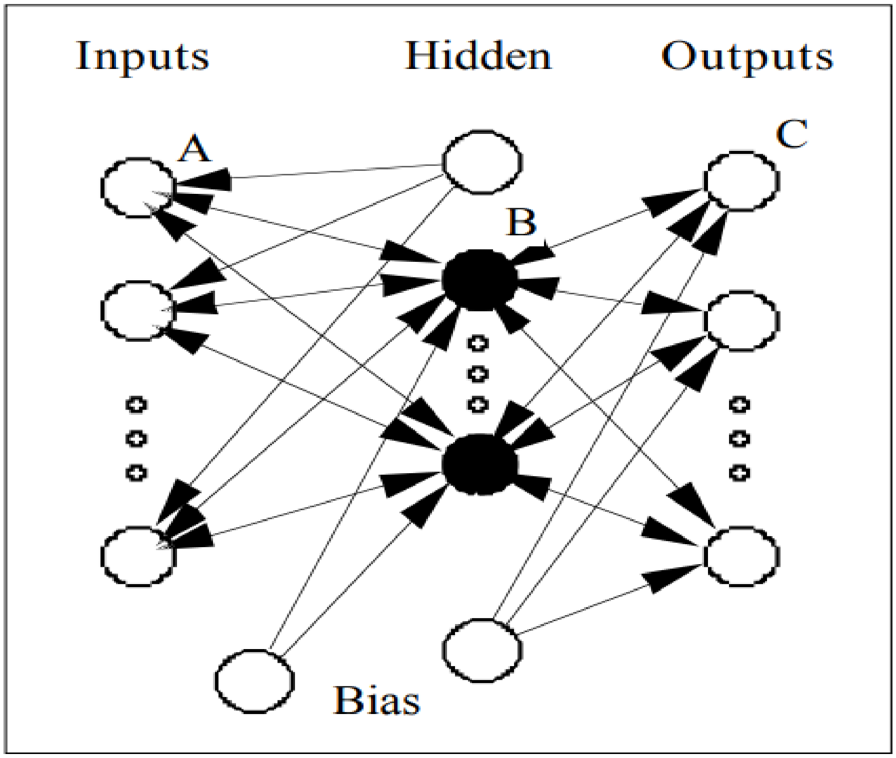

Fig. 1. BDNN structure diagram

The image below, from layer A to layer C, is the feedforward learning neural network; from layer C to layer A is the backward learning neural network. One thing that needs to be noticed is that the input neurons and output neurons should be one to one relationship. Our dataset does not meet this requirement since we only have two outputs (0 or 1) to indicate genuine or acted anger while having more than six input features. To solve this problem, we add extra nodes with the value of -1 to the output layer.

2.3 Evaluation method

We use accuracy to measure the performance of the models on the whole test set without distinguishing the data's label. Accuracy is the measure of all the correctly identified cases which is most used when all the classes are equally important. To calculate accuracy, we use confusion matrix. Note that the entries inside of a confusion matrix (TP, TN, FP, FN) are counts [14]:

- **True Positives (TP)**: The number of instances that the model predict as positive and the prediction is correct(True).
- **True Negatives (TN)**: The number of instances that the model predict as negative and the prediction is correct(True).
- **False Positives (FP)**: The number of negative instances that the model incorrectly classified as positive (i.e. the negative examples that were falsely classified as "positive")
- **False Negatives (FN)**: The number of positive instances that the model incorrectly classified as negative (i.e. the positive examples that were falsely classified as "negative")

This is the equation of **Accuracy**.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \tag{2}$$

3 Results and Discussion

In this chapter, the research result will be recapitulated and marshalled into the following aspects:

- The performance of the BDNN model trained by original training data with six features and the performance of the models trained by augmented training data using fuzzy clustering with 2 cluster centers.
- The comparison between the performance of the different BDNN models trained by augmented training data using fuzzy clustering with 2, 3 and 4 cluster centers.
- Compare the stability of classification accuracy between BDNN model trained by 2 features augmented training set and 3 features augmented training set.

3.1 Original training data vs. augmented training data

At first, we set the 15 hidden neurons, 1000 epochs, and 0.001 learning rate and uses the original training data to train the model. After that, we tune these hyperparameters and change the training set into the augmented one which includes two extra columns of membership values. The BDNN uses the ReLu activation function in the hidden layer and the Softmax activation function in the output layer for classification.

Using original training set				
Hidden neurons	Epochs	Learning rate	Training accuracy	Test accuracy
15	1000	0.001	63.1%	53.4%
15	1500	0.001	67.1%	57.2%
15	1500	0.005	67.4%	55.9%
15	2000	0.001	68.5%	56.5%
20	1000	0.001	70.3%	60.8%
20	1500	0.001	72.7%	63.5%
20	1500	0.005	72.8%	63.8%
20	2000	0.005	75.6%	60.6%
25	1000	0.001	74.9%	65.3%
25	1500	0.005	78.7%	69.4%
25	1500	0.01	83.5%	72.4%
25	2000	0.01	71.7%	66.4%
Using 2 features augmented training set				
15	1000	0.001	65.4%	56.3%
15	1500	0.001	68.8%	59.7%
15	1500	0.005	69.1%	56.1%
15	2000	0.001	70.5%	56.5%
20	1000	0.001	73.3%	61.9%
20	1500	0.001	74.1%	65.1%
20	1500	0.005	74.7%	66.3%
20	2000	0.005	73.6%	62.8%
25	1000	0.001	75.4%	66.3%
25	1500	0.005	79.9%	70.4%
25	1500	0.01	84.7%	75.4%
25	2000	0.01	68.8%	64.2%

Table 1. The results of hyperparameter tuning and using different training sets

Table 1 shows the best BDNN model trained by the original training set with 25 hidden neurons, 1500 epochs training, and 0.01 learning rate. This model achieves 83.5 % accuracy at the end of training and 72.4% accuracy in the test set. From Tabel 1. we also notice the effect of different hyperparameter tuning strategies. Increasing the epoch number can make the model perform better in the test set; however, the test accuracy decreases if the epoch number is more than 2000. The number of hidden neurons affects the test accuracy significantly compared to training epochs. It is evident that 15 hidden neurons are not enough to learn the implicit pattern of the training data, and the test accuracy is improved over 5 percent every time we increase five hidden neurons and fix other hyperparameters. Only when enough hidden neurons are offered can the neural network sufficiently learn from training data. Another finding is that the influence of learning rate to improve test accuracy is relatively small or even damage test accuracy when there are not too many hidden neurons. However, when hidden neurons become 25, their influence is remarkable. Similar to what we just described, enough hidden neuron helps to learn the pattern from training data, which is a foundation of the training and enhances the neural network's capability to process data. Only based on enough hidden neurons, tuning learning rate improves accuracy largely otherwise it contributes little or exacerbates the accuracy.

For the BDNN models trained by the augmented training sets, the overall test accuracy is higher than the BDNN models trained by the original training set. Even though the improvement in test accuracy is 2% approximately, when we use the same hyperparameters while change the training set to train the BDNN model, it suggests that our data augmentation strategy using fuzzy clustering indeed produces a positive effect. Two extra membership value columns give more information to allow the BDNN model to learn from other aspects. Despite the fact that we can not ensure for every data in the training set, both of these two membership values make a positive contribution to the result that the data is classified correctly. The overall test accuracy improvement indicates that these two membership values play essential parts in their correct classification for most of the data.

3.2 Classification result using different augmented training sets

After we compare the classification results of BDNN models trained by original training set and augmented training set with 2 extra features, we move one step further to figure out what will happen if we still enlarge the features of training set.

We establish another two augmented training sets using fuzzy clustering to add extra three columns of membership value and four columns of membership value, respectively. These new training sets train more BDNN models, and their classification results are listed below. We use the 2-FA model to represent the BDNN model trained by two features augmented training set; similarly, we use 3-FA and 4-FA model to represent the other two kinds of BDNN model. According to the table, 3-FA models' classification performances are pretty close to 2-FA models' performances, and the difference of classification accuracy between these two types of model is within 2%. Nevertheless, if the number of extra features becomes four, the classification is dramatically affected, and the accuracy reduces by about 7%. The difference becomes more conspicuous when the BDNN has 25 hidden neurons, 4-FA models' performances are more than 10% lower than the other two types of models. Intuitively, we may expect apparent difference test accuracy occurs between 3-FA models and 2-FA models, just like 4-FA models perform worse than 3-FA models. The doubt such as the third extra feature contains misleading information and would be detrimental to the training process compared to only two extra features. We contrast the training set with two extra features and the training set with three extra features detailedly. We find that for most of the data points in the 3-FA training set, their first two features resemble their corresponding values in the 2-FA training set, and their values of the third features are pretty small. This finding gives solid evidence to explain why 2-FA models' test accuracy is close to 3-FA models' test accuracy. There are six identical feature values, two very similar feature values for every data point in these two training sets; the minor difference is that the 3-FA training set has one extra feature value close to zero. Thus, the BDNN model trained by these two training sets learns similar information from data and obtains close test accuracy. Now we may ask, is the huge difference between 4-FA models and 3-FA/2-FA models caused by the huge difference between feature values? We list the data in the 4-FA training set and find our suspect is correct. The values of the first three extra features in the 4-FA training set are dissimilar to the corresponding values in the 3-FA training set for nearly one-fourth of the data, making it is more different from the corresponding data point in the 2-FA training set. It goes without saying that we will have more distinct test accuracy if we use the training data with more different feature values.

3.3 Stability comparison between 2-FA model and 3-FA model

In the result above, 3-FA models show similar performance to the 2-FA model in test accuracy. In spite of this fact, we also care about whether 3-FA BDNN models can perform as stable as the 2-FA BDNN models. Therefore, we train six models for each type and use the box plot to present their test accuracy. Based on the box plot, we observe that indeed the 3-FA models achieve similar test accuracy to 2-FA models for some time. However, the fluctuation

Add 2 features to training set				
Hidden neurons	Epochs	Learning rate	Training accuracy	Test accuracy
15	1000	0.001	65.4%	56.3%
15	1500	0.005	69.1%	56.1%
15	2000	0.001	70.5%	56.5%
20	1000	0.001	73.3%	61.9%
20	1500	0.005	74.7%	66.3%
20	2000	0.005	73.6%	62.8%
25	1000	0.001	75.4%	66.3%
25	1500	0.01	84.7%	75.4%
25	2000	0.01	68.8%	64.2%
Add 3 features to training set				
15	1000	0.001	64.8%	55.1%
15	1500	0.005	68.1%	56.0%
15	2000	0.001	70.5%	55.5%
20	1000	0.001	73.1%	59.9%
20	1500	0.005	73.7%	62.3%
20	2000	0.005	73.6%	62.7%
25	1000	0.001	74.9%	66.1%
25	1500	0.01	82.1%	73.6%
25	2000	0.01	69.6%	63.2%
Add 4 features to training set				
15	1000	0.001	55.7%	47.2%
15	1500	0.005	54.4%	49.1%
15	2000	0.001	60.3%	50.7%
20	1000	0.001	63.5%	53.0%
20	1500	0.005	63.7%	53.3%
20	2000	0.005	62.6%	52.8%
25	1000	0.001	67.4%	56.3%
25	1500	0.01	70.7%	57.4%
25	2000	0.01	71.8%	52.2%

Table 2. The results of hyperparameter tuning and using different augmented training sets

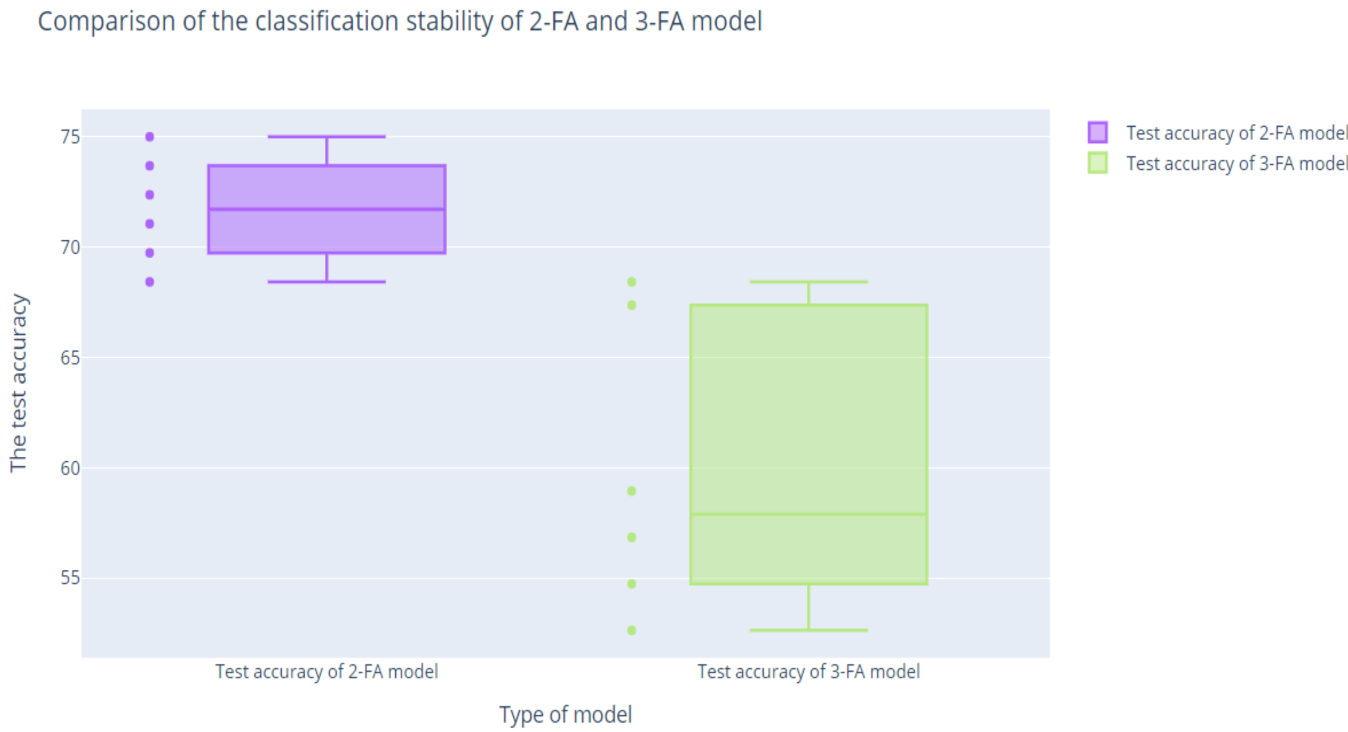

Fig. 2. Test accuracy of the 2-FA model and the 3-FA model

of the 3-FA models' test accuracy is much bigger than the fluctuation of the 2-FA models' test accuracy. In other words, the 3-FA model is more unstable than the 2-FA model, even though they occasionally have comparable performance. The main reason for this situation is also the values of the extra feature. We have discussed in the last section. First, there is a high similarity between the two extra features in the 2-FA training set and the value of the first two extra features in the 3-FA training set. Second, the minor influence of the value of the third extra feature in the 3-FA training set. As a result, the 3-FA model achieves almost the same performance as the 2-FA model. However, our extra features are produced by the fuzzy clustering method, which has a certain degree of randomness when generating the final membership value for each cluster. Due to randomness, we can not expect either the values of the first two features in the 3-FA training set are close to the values of the extra features in the 2-FA training set or the values of the third feature do not impact the training process seriously. Therefore, when huge differences occur between the value of the extra features in the 2-FA training set and the 3-FA training set, the 3-FA models' performance of test accuracy tends to oscillate dramatically.

3.4 Discussion

In this research, we propose a data augmentation method using fuzzy clustering to generate extra features to improve the classification accuracy of the BDNN model. We successfully implement the method in genuine or faked anger image classification tasks and find around 3% test accuracy improvement when fix training hyperparameters using an augmented training set. Our best model has 75.4% test accuracy trained for 1500 epochs with a 0.01 learning rate and 25 hidden neurons using two features augmented training set. This accuracy is 3% higher than the test accuracy of the model using the same hyperparameters but using the original training set without feature augmentation. Extra features provide extra information allowing the BDNN model to learn more implicit data patterns to predict better. We augment our training set further by setting the number of clusters equal to three and four to investigate their influence on test accuracy. The result shows the 3-FA models have the ability to do classification almost as well as 2-FA models sometimes because of the similar value of the extra features between training sets. However, their performances are not stable compared with the 2-FA models. The 4-FA models' test accuracy is worse than the 3-FA models and 2-FA models. Compared with the 2-FA training set, too many different extra feature values are added in the 4-FA training set for the corresponding data, which become noise impeding the model to do classification correctly. We try to shrink the size of our neural network model by applying weight pruning while damaging the test accuracy immensely; thus, we do not record this test accuracy in the result part. It is common knowledge that the more data we have access to, the deep learning model we train, the more effective it can be[13]. However, it is expensive to collect a large amount of data and transform them into the training set. The idea that exploits the current data by some augmentation methods draws much more attention from researchers. The fuzzy clustering is able to afford this mission to enhance learning ability and improve the classification accuracy for models. This data augmentation method is inspired by the application of C-means clustering in image segmentation [12] since our task is also related to the problem about image.

4 Conclusion and Future Work

Adding extra feature values for the training set using fuzzy clustering increases the test accuracy of the BDNN model when the number of extra features is relatively small and the feature values do not contain too many noises that interfere with the training. Although the performance of the BDNN models trained by augmented data exceeds other BDNN models that are not, a noticeable difference is still witnessed compared to the result in this paper[1] which is 95% test accuracy. There are some limitations to our research. First, because of the randomness of the c-means clustering, we can not guarantee that the values of extra features are identical if we apply c-means in the same data twice. Second, the data we can access is limited. We only have about 280 data in the training set; thus we do not know if applying data augmentation in an adequate training set can improve the classification accuracy. Third, our fuzzy clustering uses six features for each data in training set to generate cluster centers, but we lack an effective evaluation method to assess whether we need all of these six features. Nevertheless, we would expect that such data augmentation using fuzzy clustering might be applied to benefit classification tasks lacking sufficient data and contribute to the other algorithms for the classification task.

For future work, our goal is still to give a solution to classify genuine and acted anger images correctly; therefore, we should not get restricted to one specific data processing technique and one specific neural network model. More complex models like SVM, convolutional neural network(CNN) will be tested on this task. More data processing methods like outliers detecting and Generative Adversarial Nets(GAN) for data augmentation will also be used in order to promote model performance. We will continue to explore the applicability of our method to similar task like genuine or acted smile classification which is an equivalently significant emotion recognition problem need to tackle. If we can obtain a large enough training set, we would figure out the influence on test accuracy after using fuzzy clustering. We may also extend this method to other classification problems like audio and video classification.

References

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. Nejad, A. F., Gedeon, T. D. (1995, November). Bidirectional neural networks and class prototypes. In Neural Networks, 1995. Proceedings., IEEE International Conference on (Vol. 3, pp. 1322-1327). IEEE.
3. Zhang, T., Zheng, W., Member, IEEE, Cui, Z., Zong, Y. and Li, Y. (2018, January). Spatial-Temporal Recurrent Neural Network for Emotion Recognition. In IEEE Transactions on Cybernetics (pp. 839-847). IEEE.
4. Liu, Y., Sun, C., Lin, L., Wang, X. (2016, May). Learning Natural Language Inference using Bidirectional LSTM model and Inner-Attention. arXiv preprint arXiv:1605.09090.
5. Hansen, L., Salamon, P. (1990, Oct)"Neural network ensembles," in IEEE Transactions on Pattern Analysis and Machine Intelligence, (vol. 12, no. 10, pp. 993-1001), IEEE.
6. Shanker, M., Hu, M. Y., Hung, M. S. (1996, Aug). Effect of data standardization on neural network training. Omega, 24(4), 385-397.
7. Wöllmer, M., Metallinou, A., Eyben, F., Schuller, B., Narayanan, S. (2010). Context-sensitive multimodal emotion recognition from speech and facial expression using bidirectional lstm modeling. In Proc. INTERSPEECH 2010, Makuhari, Japan (pp. 2362-2365).
8. Lin, C. T., Lee, C. S. G. (1991, Dec). Neural-network-based fuzzy logic control and decision system. IEEE Transactions on computers, 40(12), (pp. 1320-1336).
9. Ng, H. W., Nguyen, V. D., Vonikakis, V., Winkler, S. (2015, November). Deep learning for emotion recognition on small datasets using transfer learning. In Proceedings of the 2015 ACM on international conference on multimodal interaction (pp. 443-449).
10. Dai, Z., Chen, W., Huang, X., Li, B., Zhu, L., He, L., ... Zhang, H. (2018, December). CNN descriptor improvement based on L2-normalization and feature pooling for patch classification. In 2018 IEEE International Conference on Robotics and Biomimetics (ROBIO) (pp. 144-149). IEEE.

11. Maulik, U., Saha, I. (2010). Automatic fuzzy clustering using modified differential evolution for image classification. IEEE Transactions on Geoscience and Remote Sensing, 48(9), 3503-3510.
12. Cai, W., Chen, S., Zhang, D. (2007). Fast and robust fuzzy c-means clustering algorithms incorporating local information for image segmentation. Pattern recognition, 40(3), 825-838.
13. Perez, L., Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. arXiv preprint arXiv:1712.04621.
14. Confusion matrix, *https://en.wikipedia.org/wiki/Confusion_matrix*. Last accessed 14 Apr 2021