Hybrid of Fuzzy Clustering and Neural Network for Classifying the Veracity of Anger

Han Zhang

Research School of Computer Science,
Australian National University
u6541559@anu.edu.au

Abstract: The detection of authenticity of emotions can be realized depending on the human conscious subjectively or physiological signals objectively. The classification accuracy for the neural network classifier applying the physiological signal data (pupillary response) is much higher than that of the human conscious data (verbal response) [1]. In the previous experiment, the pupillary response data was directly fed into a standard feed forward neural network to predict the veracity of anger and a simple technique of adjusting the threshold of the activation function in the neural network had been applied which was originally raised in the classification of Dry Sclerophyll Forest [2,3]. The accuracy of traditional feed forward neural network can reach an accuracy of approximately 80% in a relatively slow speed. The fuzzy clustering NN architecture (FCNN) is successfully in enhancing the efficiency of the classification which performs an unsupervised pre-classification task in the early stage and classifies the pre-classified data using a standard multilayer neural network later [4]. Typically, the main motivation for FCNN is to reduce the number of features into fewer segments. However, the reduction of data can be both performed along row (sample generalisation) and column (feature generalisation) using fuzzy c-means clustering. Moreover, it is likely that the membership values generated from FCM can be applied to enrich the input data. In this paper, four FCNN models are discussed and compared for their feasibility to classify the authenticity of anger. It is noticeable that FCNN Model 1 which extracts the membership value from Fuzzy c-means process and uses membership value as additional feature apart from raw feature for classification in neural network can achieve a better result in a faster speed than baseline model though it is less stable. The FCNN Model 3 which use centroids of clustering as retained samples may have a slightly worse accuracy (75%) than using full training data (80%) though it significantly improves the efficiency for classification.

Keywords: Anger Detection, Emotion Classification, Human-centered Computing, Fuzzy Clustering, Fuzzy c-means

1 Introduction

Recently, there exists an increasing interest in the analysis of human expression, as the external facial expression reflects on internal mental status. The emotion detection can be either differentiated by the image of human expression, or the short video clips containing human expressions. In this experiment, different from distinguishing the emotion directly from the expression itself, the analysis is based on the responses of the observers who watched the video clips. There were 20 videos obtained from YouTube and were trimmed to the scenes with anger face for a few seconds [1]. Each of the video frames were pre-processed to maintain similar variations of characteristics of the protagonists (age, ethnicity, gender, etc) in order to minimise the influence on response caused by the observer's personal bias towards the protagonist [1]. The participants gave the verbal response of whether the anger in the video is posted or genuine. Meanwhile, the pupillary responses of the participants were recorded during watching the videos. In this paper, we will focus on the physiological signal (pupillary responses) of the participants due to the uncertainty and unreliability of verbal response.

The structure of the baseline network is fully connected, having 2 hidden layers, 6 input neurons, and 2 output neurons. The parameters of the networks are tuned and determined by the change of loss for network with different parameters in previous experiment [2]. The major problem in the previous experiment is that the training time is long. There are various existing solutions to solve this problem and most of them are focusing on the adjustment and improvement of back propagation algorithm [5]. There is an alternative solution that uses fuzzy c-means (FCM) clustering algorithm to group similar data into clusters so the number of training samples can be reduced in fuzzy stage before inputs are presented to neural network [4]. In this paper, four ways of hybrid of fuzzy clustering and feed forward neural network are introduced for comparison which are not limited to reduce the number of features in paper [4]. The first two ways are utilizing the membership values from FCM and the second two ways are utilizing the centroids of clustering from FCM.

The obtained membership values and centroids of clustering from FCM can be either used as additional data apart from raw data (retain raw data) or a new reduced version of data for training (discard raw data). The FCNN model using enriched version of data is expected to have a more optimal performance than the baseline model using raw data and the FCNN model using reduced version of data is expected to be more efficient than the baseline model using raw data.

Moreover, a simple technique of modifying the threshold of activation function is introduced. Commonly, threshold 0.5 is used to distinguish between two classes. However, Kogan suggests that sometimes classifying two classes using threshold 0.5 may not reach the best result [6]. Therefore, to show the performance with regard to different thresholds, we vary the threshold from 0.3 to 0.8 to train the networks respectively and finally discover that the modification of threshold can improve the performance of network to some extent. The threshold adjustment is helpful to give an overview of the comparison between performance of traditional NN and FCNN. Importantly, the overall change of performance can reflect the boundary cases that fall between two classes and is easily altered to the other class [2]. The level of change of performance can be considered as a factor reflecting the level of fuzziness, thus affecting the stability of performance.

2 Method

2.1 Data Pre-processing

The raw data collected from the experiment is composed of 400 samples and the number of samples in each class is the same. Each sample has 9 attributes as in Table 1 below.

Table 1. Header of Anger V1 dataset

1	2	3	4	5	6	7	8	9
Index	Video	Mean	Std	Diff1	Dff2	PCAd1	PCAd2	Label

1. Index for participants
2. Index for video
3. Mean of in pupillary response.
4. The standard deviation of in pupillary response.
5. The change of left pupillary size after watching a video.
6. The change of right pupillary size after watching a video.
7. An orthogonal linear transformation with first principal component
8. An orthogonal linear transformation with second principal component
9. The Genuine or Posed label.

Generally, the first two attributes are identification number for participants and videos. Attribute 3 to 8 are describing the pupillary responses. The last attribute is label of anger (target). However, the designer of the experiment labels the realness of anger based on the source of video. Genuine label is assigned to the record with video extracted from documentaries while posted label is assigned to the record with video extracted from movies. Therefore, apart from the label of anger, video identification number is another ground truth label. To predict the veracity of anger, video source data should be excluded from the training data. Meanwhile, the video index is removed as it is irrelevant to the pupillary response statistics.

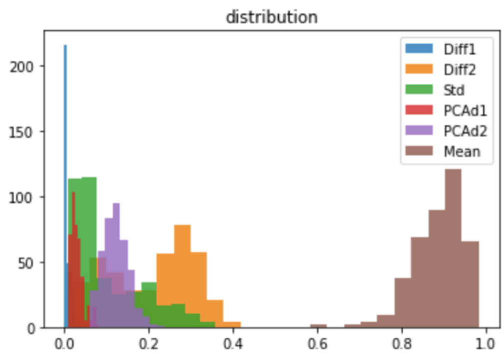

Figure 1. Distribution of each attribute

As the distribution of each attribute varies a lot from Figure 1 above, it is likely that the feature 'mean' will have greater effect on the network than other features especially 'PCAd1'. To smooth the influence brought by the distribution difference, normalization technique is applied.

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

$$x' = \frac{x - x_{mean}}{x_{std}} \qquad (2)$$

Among different normalisation techniques, the Z-score normalisation (2) can help to reach the best result with accuracy approximately 80% in testing data. While transforming the data using min-max normalisation (1) can reach the accuracy of approximately 65%. Without any transformation, the network which directly fed with the unnormalized data has an accuracy of approximately 50%. Thus, there is a significant improvement in the accuracy of predicting the two classes: 1 (Genuine) and 0 (Posted) after applying Z-score normalisation in the dataset. Then the dataset is divided into training and testing set at a proportion of 80% and 20% respectively.

2.2 Baseline Neural Network Model Construction

The baseline model is a feed-forward neural network with 6 input neurons and 2 output neurons which is trained and adjusted in previous experiment [2]. The features fed into the networks are pure pupillary statistics of participants without involving any image or video and the aim is to classify between two classes. Typically, a feed-forward network training with backpropagation algorithm to adjusts the weights is sufficient to deal with the two-class classification problem and can achieve a relative reasonable result.

To determine the number of hidden neurons, the networks with hidden neurons varying from 1 to 24 are trained separately and their loss after training are recorded for visualisation. The number of neurons is greater than 5, the loss keeps oscillating between 0.6 and 0.8. Therefore, it is feasible to choose more than five neurons. The number of neurons is set to be 20 to give the model high capacity to predict. However, there must be a trade-off between the high-capacity and over-complexity of the model, and it is necessary to introduce the L2 norm term in the optimizer (weight decay=0.001) to regularize the loss and overcome overfitting.

The number of epochs is set to be 1000. As the number of epochs increases, the accuracy of the training data is approaching 100% though there is no obvious improvement of accuracy in testing data. The accuracy for the testing data is oscillating around 80%. It makes no sense to continue on training and more epochs will result in the model overfitting. When the number of epochs is 1000, the curves of loss of both training data and testing data will flatten out with derivative closed to 0. Adam optimiser is determined based on the comparison between using other optimisers.

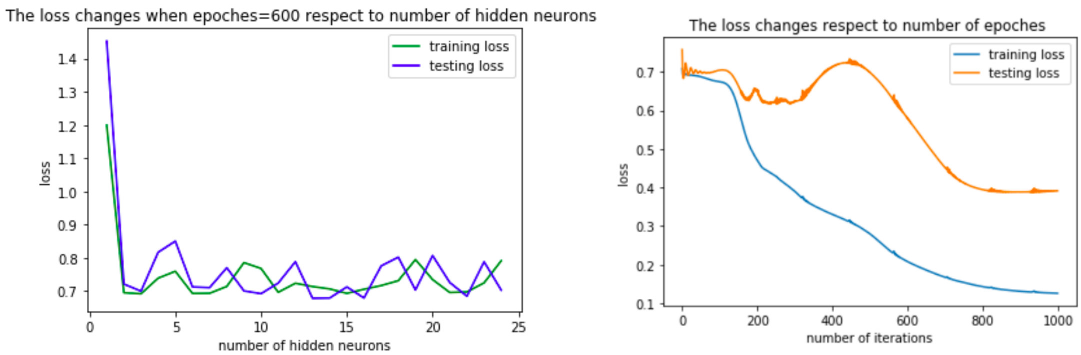

Figure 2-3. Left: The loss changes with respect to number of hidden neurons in training and testing dataset

Right: The loss changes with respect to number of iterations in training and testing dataset

To conclude, by testing different optimisation methods, learning rates, and numbers of hidden layers, the model with 2 hidden layers which trained with Adam optimiser of learning rate 0.01 works best with accuracy of approximately 80% testing data.

2.3 Fuzzy Clustering Neural Networks (FCNN) Introduction

The structure of FCNN is originally proposed in the classification of electrocardiographic beat [4]. The FCNN architecture have two major elements which is fuzzy classifier and neural network classifier. The fuzzy classifier is responsible for the pre-classification task which analyses the distribution of data and grouping similar data into clusters. Thus, the number of segments for training are reduced from original N features to N' segments (see Fig 2.) using FCM clustering in fuzzy stage before inputs are presented to the neural network. With the generalization of the data, the standard Neural Network can perform a classification task using the generalized data and obtain the final output. In summary, the topology of FCNN is the combination of unsupervised fuzzy clustering and supervised baseline neural network.

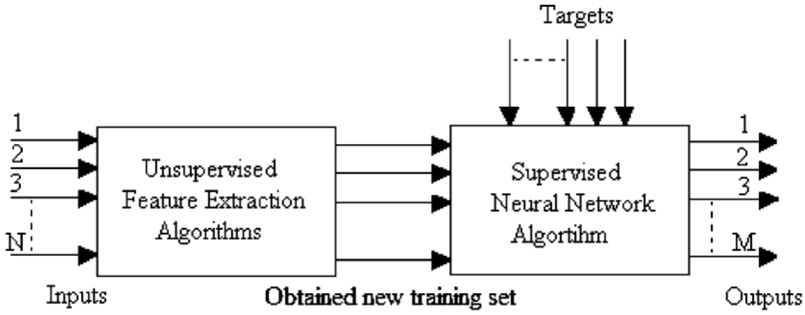

Figure 4. FCNN architecture [4]

2.4 Adaptation of Threshold

In the classification problem between two classes or multiple classes, softmax activation function can be used to output membership that the entity belonging to each class. As in the classification problem between more than 2 classes, the prediction label of the entity will be assigned to the class having higher probability. In the classification problem between only 2 classes, the class having higher probability will also be assigned. That is to say, the class having a probability more than 0.5 will be the prediction output and 0.5 is the default threshold to classify 2 classes.

However, it is possible that classifying two classes using the threshold different from 0.5 can achieve a better result [2]. Thus, a self-defined classifier which is able to take the parameter of threshold in convenience of training networks with different thresholds is introduced for classification.

The evaluation of the performance takes an overall consideration of various measures including the number for correctly classifying, number of FP, number of FN, rate of accuracy, rate of precision, and rate of F-score. The performance for the network with threshold varying from 0.3 to 0.8 is stored in a data frame for both training and testing data. Moreover, to better analyze the performance between the baseline model and the FCNN model, the plots are generated for visualisation.

2.5 Hybrid of FCM And Baseline Model

Typically, the fuzzy clustering is performed along column to reduce the number of features (in Fig 2) which is effective for the training data having large number of features such as large amount of pixel values for images. For the anger dataset version 1, there are 6 features fed into the neural network, so it is unnecessary to perform any data reduction along the feature. Instead, it is worthwhile to use FCM to generate new feature through the analysis of existing features. In the experiment, the membership value for each sample belonging to each class is considered as the potential new feature. Moreover, as the FCM can group similar data points into clusters, the centroid of each cluster has the ability to represent other points, it is likely that training the representations instead of the whole dataset can achieve a similar performance in a much faster manner. Additionally, it is also worth investigation that whether training on the expanded dataset that adds the centroids data onto the original data is able to have better performance than using original data.

The overview of experimental step is shown below.

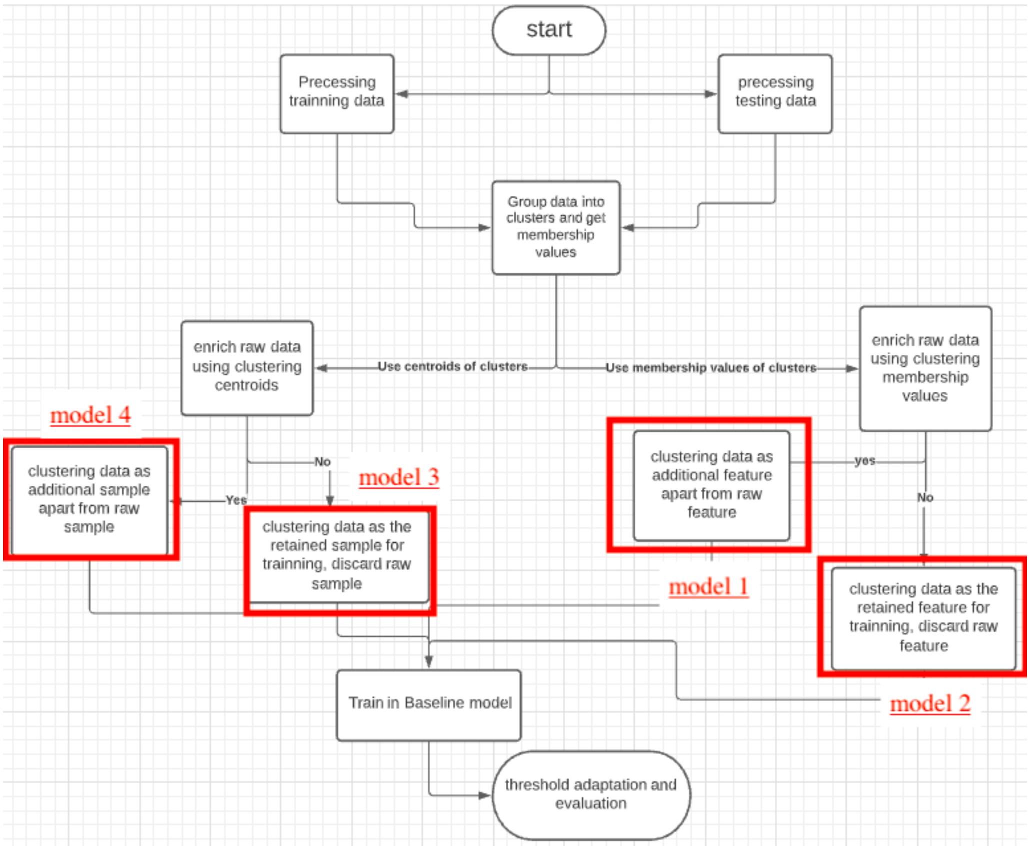

Figure 5. Experimental step

3 Result & Discussion

3.1 Evaluation Standard

The evaluation generally focuses on two aspects which is correctness and efficiency. That is, the optimal neural network is the one can have a satisfactory classification ability in an acceptable time. The aspect of classification ability is measured on accuracy, precision, recall, and F-score. The aspect of efficiency is measured by the convergent time of the neural network or the time taken for training.

Accuracy measures the rate of correct predictions among all predictions which is a persuasive measure to assess the quality of prediction. The formula for precision and recall is measured as $\frac{|TP|}{|TP|+|FP|}$ and $\frac{|TP|}{|TP|+|FN|}$ respectively. There is an unavoidable trade-off between precision and recall as raising the classification threshold will increase the count of TN and FN and decrease the count of TP and FP [7]. As the value of threshold is raised gradually, recall keeps decreasing and precision keeps increasing. Therefore, it is misleading to evaluate the performance of network based on the measure of precision and recall separately. F-measure combines the precision and recall as a harmonic mean, which can be considered as an effective measure for evaluation [7].

3.2 Baseline Model Evaluation

With adjustment of the threshold from 0.3 to 0.8 to classify the two classes, the overall performance of training data and testing data is shown in Appendix_A. The performance of network is also plotted for visualisation. From the statistic of baseline model, the network with threshold 0.50 or 0.55 has the optimal overall performance in training data. The network with threshold 0.45 has the optimal overall performance in testing data. There exists the inconsistency of the threshold setting for the network that can achieve the best result between training and testing set. For the efficiency of the baseline model, it consumes more than 5 seconds for training which can be reduced using FCNN.

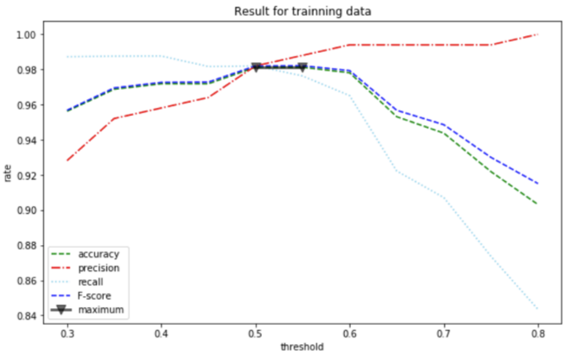

Figure 6. The performance of baseline network on training data

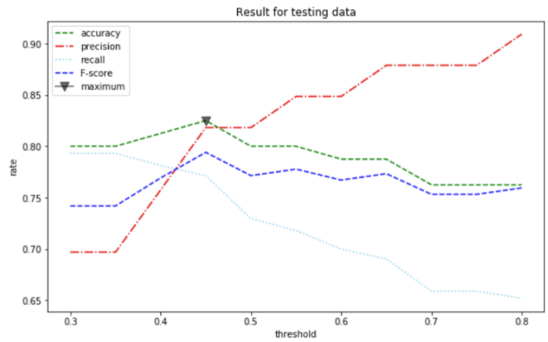

Figure 7. The performance of baseline network on testing data

3.3 FCNN Evaluation

- FCNN Model 1: Use membership value as additional feature apart from raw feature.

There are only 6 features corresponding to the pupillary response that are fed into the neural network. One of the potential ways to enrich the features is to generate a new feature from the FCM. There are two classes to classify, it is likely that performing a fuzzy clustering with two centroids and obtain the membership value for each sample belonging to each class will be a useful information for two-class classification problem. The visualization of the fuzzy c-means clustering with 2 centroids is shown in Figure 8. As the membership matrix is composed of two rows which represents the membership values for each sample belonging to the two classes, it is unnecessary to extract both of these two membership arrays as two new features. The reason is that, for two class clustering problem, the membership value belonging to one class is the supplementary of the membership value belonging to another class. The header of features for model 1 is shown below.

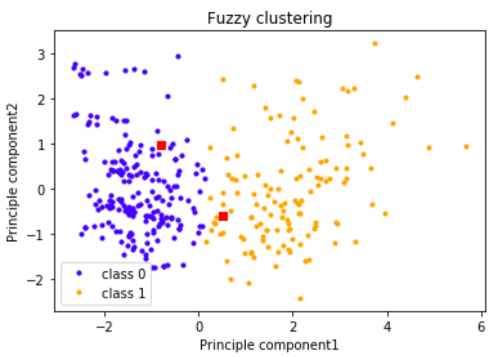

Figure 8. Fuzzy clustering with 2 centroids

Table 2. Header features for model 1

1	2	3	4	5	6	7
Mean	Std	Diff1	Dff2	PCAd1	PCAd2	Membership value for class 0

Compared to baseline model, FCNN model 1 is much efficient than the baseline and is able to achieve a slight better classification performance. That is, the testing result (along the four measures) for the baseline model after 1000 epochs of training is slightly worse than that for the FCNN model after training for 600 epochs (see Fig 9). The training time for baseline model is 5.14 seconds and it can be reduced to 1.23 seconds for the FCNN model 1.

According to the previous paper, the group of entities whose labels are easily changed by adapting the threshold can be considered as the boundary cases that lie between the two classes and is commonly to be wrongly classified [2]. As shown in Figure 10 and Table 3, the four measures all experience a much more dynamic change in the process of adjusting the threshold from 0.3 to 0.8. That is, there are more points on the boundary between two classes using FCNN model for classification, thus making the performance unstable and sensitive to the threshold. This phenomenon is mainly caused by involving the fuzzy feature membership. Different from K means clustering, all the entity is classified to one class with the crisply with membership value either 0 or 1. Fuzzy c-means assign soft membership value to each entity. For the sample getting membership value 0.49 for 'Posted Class' and 0.51 for 'Genuine Class', the entity is easily become an ambiguous boundary case after classification using FCNN that involving the membership value. As a result, the performance will be less stable when choosing different threshold values due to the fuzziness of FCNN model which is a major drawback for the FCNN model.

Therefore, FCNN Model 1 which extracts the membership value from Fuzzy c-means process and uses membership value as additional feature apart from raw feature for classification in neural network can achieve a better result in a faster speed than baseline model though becomes more sensitive to the threshold setting.

Table 3. Difference between best and worst performance

	Accuracy	Precision	Recall	F-score
Baseline	6.2%	21.3%	14%	3.6%
FCNN 1	9.3%	28.1%	21%	5.3%

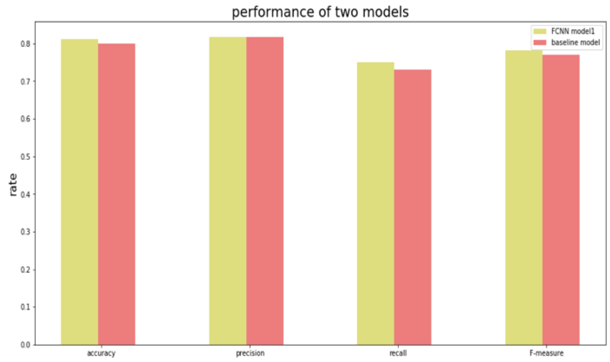

Figure. 9. Performance of FCNN and baseline NN when threshold equals 0.55

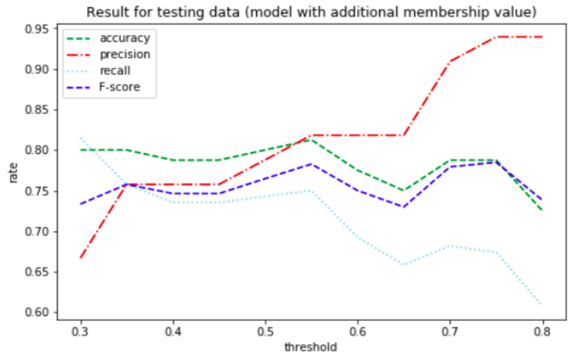

Figure. 10. The performance of FCNN on testing data with different threshold

- FCNN Model 2: Use membership value as retained feature, discard raw data.

As discussed above, it is unreasonable to perform any data reduction along the feature as there are limited number of features. The membership value after performing FCM with only 2 clusters cannot have the generalization ability to replace the raw feature. While performing FCM with more clusters makes it meaningless as the initial motivation is to reduce the number of features, so discarding the raw data and feeding the membership value into the neural network is not appropriate in this scenario.

- FCNN Model 3: Use centroids of clustering as retained data, discard raw data.

In practical applications, there are many problems related to sampling from large amount of real-life data [8]. In the past, people always randomly selected a few of them as samples for learning, but it is likely that the randomly selected sample set does not contain the characteristics of all the samples, and there will be large errors occurred in the predicted results due to the randomness. The determination of the number of samples to choose and which samples to choose from the real-life data has become an urgent problem [8].

In the danger dataset V1, there are a total number of 400 samples, which is not extremely large as real-life data, but it can still be generalized and transformed into smaller size. Fuzzy c-means clustering is applied along all samples to group similar samples and the corresponding centroids may conclude the characteristics of samples in that particular clustering. Therefore, centroid of each clustering is collected as training data and the size of training data is reduced from 320 (400×80%) to the number of centroids. Three FCNN models with different number of clusters are constructed for comparison. The training time for models are shown in the table below. It is noticeable that the more of reduction of data, the less time is required for training.

Table 4. Training Time for Model

Model	Baseline	FCNN 50 Clusters	FCNN 100 Clusters	FCNN 200 Clusters
Time	5.14 s	1.78 s	2.77 s	3.94 s

Figure 11 shows that the accuracy changes with respect to iterations obtained from three different training sets. From the graph, for all three models with different training patterns (different clusters in FCM), the accuracy rate increases as the number of iterations increases. It is noted that the baseline model can generally reach a higher accuracy than most of the FCNN model. The FCNN model with 100 clusters and applies these centroids as training data can reach the nearly same accuracy as the baseline model.

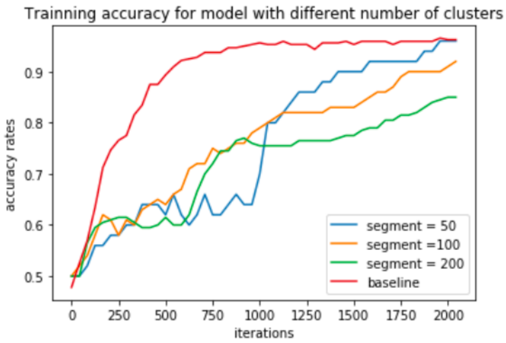

Figure 11. Classification results for training data

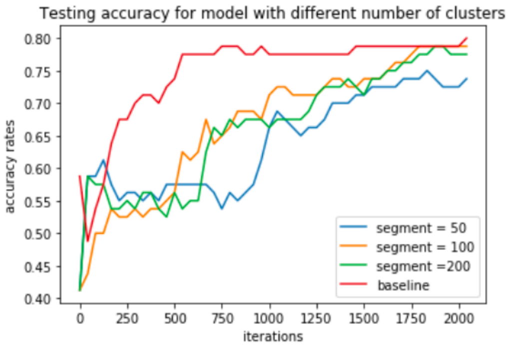

Figure 12. Classification results for training data

- FCNN Model 4: Use centroids of clustering as additional data apart from raw data.

This model is appropriate for the scenario that there is limited number of samples for training. By testing, there is no improvement either in classification ability or training efficiency compared to baseline model. The optimal accuracy for the FCNN model is 75% which is less than 80% in baseline model. The training time for FCNN is 5.2 seconds which is longer than 5.14 seconds for baseline model. Therefore, the model is not appropriate for this dataset.

4 Conclusion & Future Work

4.1 Conclusion

In summary, the study shows how a simple feed-forward neural network can be used to classify the veracity of anger by feeding the pupillary response data of participants. Moreover, a threshold adaptation technique is applied to the neural network to adjust the performance of neural network. From the evaluation statistics, it is proved that the baseline network with threshold other than 0.5 is possible to reach a better result than that with a standard threshold 0.5 which is consistent with the statement of Kogan [6]. The modification of threshold provides a higher perspective to filter the data of ambiguity that is easily altered to the other class in the process of threshold adaptation.

Additionally, the FCNN architecture has been proposed, developed and presented to classify the veracity of anger by using different training sets generated by Fuzzy c-means clustering. A comparative evaluation between different FCNN models is conducted. The result show that with the addition of membership value obtained by FCM, the neural network can have a better result more efficiently. However, in the process of threshold adaptation, the performance becomes less stable and sensitive to the threshold due to the involvement of fuzzy membership value. While reducing the number of training samples using FCM can also achieve an acceptable performance closed to baseline neural network in a much faster speed.

4.2 Future Work

This paper generally focuses on the 'Anger V1' dataset which is structured and regular with limited number of features. Chances are that the using 'Anger V2' can reveal the advantages of FCNN over traditional NN as the length of features for each sample is not fixed and there exist some missing values. Fuzzy c-means is able to generalize the raw feature into a fixed length thus provide a set of structured features for neural network. Moreover, the size of 'Anger V2' dataset is much bigger than 'Anger V1' dataset so it is highly possible to experience long time of training which can be relieved by FCNN. Therefore, 'Anger V2' dataset will be analyzed to compare the FCNN and traditional NN in the future.

Reference

[1] L. Chen, T. Gedeon, M. Hossain and S. Caldwell, "Are you really angry?", Proceedings of the 29th Australian. Conference on Computer-Human Interaction, 2017. Available: 10.1145/3152771.3156147

[2]. H. Zhang, 2021 A Standard Feed-Forward Backpropagation Neural Network for Classifying the Veracity of Anger: Construction, Parameter Tuning, and Evaluation, 4th ANU Bioinspired Computing Conference

[3] L.K. Milne, T.D Gedeon, and A.K. Skidmore, 1995. Classifying Dry Sclerophyll Forest from Augmented Satellite. Data: Comparing Neural Network, Decision Tree & Maximum Likelihood. training, 109(81), p.0.

[4]. Y. Özbay, R. Ceylan and B. Karlik, "A fuzzy clustering neural network architecture for classification of ECG arrhythmias", Computers in Biology and Medicine, vol. 36, no. 4, pp. 376-388, 2006. Available: 10.1016/j.compbiomed.2005.01.006 [Accessed 28 May 2021].

[5] Y. He, L. Xiang, the analysis and application of the BP Neural Network based on the fuzzy clustering.

[6] Kogan, "Neural networks trained for classification cannot be used for scoring," IEEE Trans. on Neural Networks, 1991

[7] P. Christen and K. Goiser, "Quality and Complexity Measures for Data Linkage and Deduplication", Quality Measures. in Data Mining, pp. 127-151, 2007. Available: 10.1007/978-3-540-44918-8_6

[8] A. Parlos, An accelerate learning algorithm for multilayer perceptron networks. 1992, 1(3) : 62-72

Appendix

Appendix_A: Performance of baseline model in training data and testing data

Threshold	correct	FP	FN	Accuracy	Precision	Recall	F-score
0.30	306	12	2	95.6%	92.8%	98.7%	95.7%
0.35	310	8	2	96.9%	95.2%	98.8%	97.0%
0.40	311	7	2	97.2%	95.8%	98.8%	97.2%
0.45	311	6	3	97.2%	96.4%	98.1%	97.3%
0.50	314	3	3	98.1%	98.2%	98.2%	98.2%
0.55	314	2	4	98.1%	98.8%	97.6%	98.2%
0.60	313	1	6	97.8%	99.4%	96.5%	98.0%
0.65	305	1	14	95.3%	99.4%	92.2%	95.7%
0.70	302	1	17	94.4%	99.4%	90.7%	95.0%
0.75	295	1	24	92.2%	99.4%	87.3%	93.0%
0.80	289	0	31	90.3%	100%	84.3%	91.5%

Threshold	correct	FP	FN	Accuracy	Precision	Recall	F-score
0.30	64	10	6	80%	69.7%	79.3%	74.2%
0.35	64	10	6	80%	69.7%	79.3%	74.2%
0.40	65	8	7	81.2%	75.8%	78.1%	76.9%
0.45	66	6	8	82.5%	81.8%	77.1%	79.4%
0.50	64	6	10	80%	81.8%	73.0%	77.1%
0.55	64	5	11	80%	84.8%	71.8%	77.8%
0.60	63	5	12	78.6%	84.8%	70%	76.7%
0.65	63	4	13	78.6%	87.9%	69%	77.3%
0.70	61	4	15	76.3%	87.9%	65.9%	75.3%
0.75	61	4	15	76.3%	87.9%	65.9%	75.3%
0.80	61	3	16	76.3%	90.9%	65.2%	75.9%