Error Modelling Approach based on Artificial Neural Networks for Face Emotion Recognition

Luis-Alberto Pérez-Gaspar, Santiago-Omar Caballero-Morales,
and Felipe Trujillo-Romero

Technological University of the Mixteca, Huajuapan de Leon, Oaxaca, Mexico
`luis_335450@hotmail.com, {scaballero, ftrujillo}@mixteco.utm.mx`

Abstract. In this paper an approach based on modelling of recognition error with Artificial Neural Networks is presented to increase face emotion recognition in absence of any pre-processing or enhancement technique for feature extraction. This approach consists of two stages: in the first stage an ANN structure is defined for the recognition task by means of a Genetic Algorithm (Recognition ANN - ReANN). Then this structure is used to perform recognition on a test set Y to estimate classification error probabilities. In the second stage an additional ANN is defined to associate these error patterns with correct classification patterns using the same test set (Corrective ANN - CoANN). The composite ANN system then is tested with a different set Z. In recognition tasks performed with the ReANN it was observed that some emotions were more likely to be incorrectly classified than others. This was further corroborated with perceptual data. With the integrated ANN system (ReANN plus CoANN) it was observed that some of these emotions could be recognised more accurately. In general overall recognition was increased from 75% to 85% with this approach.

Keywords: Face Emotion Recognition, Artificial Neural Networks, Genetic Algorithms

1 Introduction

The recognition of facial emotions is a universal and innate ability in humans and is deeply involved in social communication [1]. For the development of more intuitive and intelligent artificial entities as interfaces and assistive robotic systems, face emotion recognition has become an important research subject [3].

Some works have been performed with significant results in this field. In [9] image pre-processing was performed in order to classify emotions. Methods like Discrete Wavelet Transform (DWT), PCA (Principal Component Analysis), 2D-LDA (Linear Discriminant Analysis) and Support Vector Machines (SVM) were used for feature extraction and classification tasks on the JAFFE database [4] reporting accuracies within the range of 90% to 95%.

Luis Alberto Pérez-Gaspar, Santiago Omar Caballero-Morales and Felipe Trujillo-Romero

Another work that used the JAFFE and YALE databases was presented in [6] where facial features were extracted by using the Gabor Filter in order to have representative facial deformation patterns. PCA was performed on these features for dimensionality reduction while the classification task was performed with a three-layer Artificial Neural Network (ANN) classifier trained with a backpropagation algorithm. An overall classification rate of about 96% was reported with this technique.

In [12] the combined implementation of Gabor Filter and Local Binary Patterns on images was presented for the development of a SVM classifier. Additional enhancement of this implementation was achieved with a Genetic Algorithm (GA). Classification tasks performed with the JAFFE and Cohn-Kanade databases reported a general recognition rate of about 96%.

Finally in [7] emotion recognition was performed on video data. Multi-scale morphological erosion and dilation operations were used as pre-processing techniques to extract features from eye and mouth regions. For each region an emotion recognition system was developed where each recognition system consisted of emotion-specific ANNs. Recognition rates between 73% and 87% were obtained.

In general, the works focused on facial emotion recognition require specialized pre-processing techniques for feature extraction prior to the recognition module. In this work we focus on the performance of the classifier/recogniser itself and its response, and we propose a post-processing technique which can improve the response of the recogniser independently of any pre-processing on the input data. This post-processing technique is based on the concept of error modelling which has been performed for speech recognition particularly when the patterns to be recognised have many disturbances [2,8].

With this technique an Artificial Neural Network (ANN) system is presented which is integrated by a recognition ANN (ReANN) and an error modelling ANN (corrective CoANN). Experiments performed with the JAFFE database presented a recognition rate of 85.00% which is similar to human semantic ratings [5] and is higher than the performance of the ReANN alone (77.50%).

These advances are presented in this paper as follows: in Section 2 the feature extraction method is presented while the development of the ReANN system is presented in Section 3. Then the details of the CoANN are presented in Section 4 while the performance results of the ReANN and ReANN + CoANN systems are presented in Section 5. Finally our observations and comments about future work are presented in Section 6.

2   Feature Extraction

For the present work the Japanese Female Facial Expression (JAFFE) database was used for the experiments. This database contains 213 images (approximately three image samples per emotion) of 7 facial emotions or expressions (Anger,

Disgust, Fear, Happiness, Neutral, Sadness and Surprise) posed by 10 Japanese female models. Some examples of the JAFFE database are presented in Figure 1. The size of the images in the JAFFE database is 256×256 pixels. From this database the following emotions were considered: Anger, Happiness, Neutral, and Sadness [11,10].
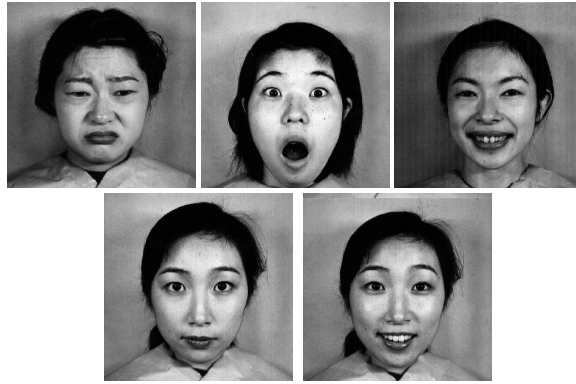
Fig. 1. Sample images from the JAFFE database.

Prior to feature extraction each image in the JAFFE database was reduced by a factor of 0.75 to reduce computational processing. This step has been performed by other works wthat involve image recognition [6]. Then each image was normalized to reduce dispersion and variability. In this way each image had a final size of 64×64 pixels.

The feature extraction consisted in the extraction of the most significant regions of the face that express emotion: when an expression is performed there are essential regions that describe the emotion and are useful to identify the associated emotional state. The regions containing the eyebrows, eyes and mouth were identified as the most important for emotion recognition. This is an extension on the regions considered in [7]. In Figure 2 the pixel intervals for each region are presented.

Each region is extracted and reshaped into a single feature vector p_{ij} which represents the i-th emotion expressed by the image of the j-th person. This process is performed for each image in the JAFFE database for the emotions considered in this work. Note that no other pre-processing technique is performed as we focus on the effect of the error modelling ANN on the classification performance of the (main) recognition ANN (ReANN).
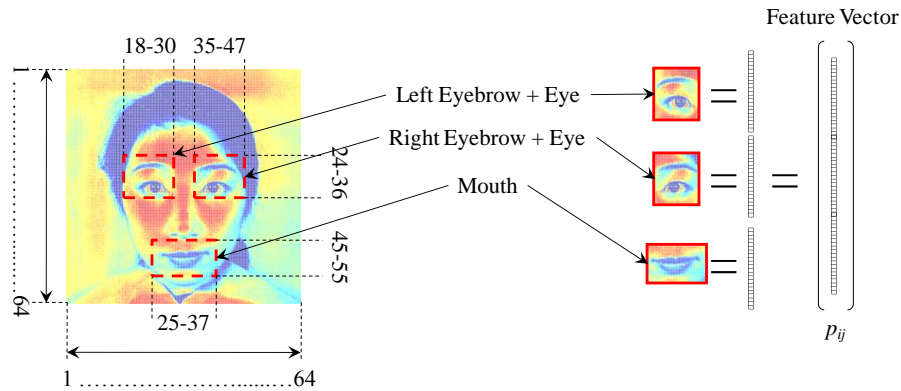
Fig. 2. Regions considered for feature extraction.

3 Estimation of the Recognition ANN

In Figure 3 the characteristics of the input and output vectors for the (main) recognition ANN (ReANN) are presented. The input vectors are the feature vectors p_{ij} while the output vectors consist of a sequence of integer numbers that represent the associated emotion for each input vector: "1" for Anger, "2" for Happiness, "3" for Neutral, and "4" for Sadness. For training of the ReANN the required output (reference) is the sequence of integer numbers describing the emotional state of each input vector from a training set X. Note that the actual output of the ReANN consists of real numbers instead of integer numbers. In the case of highly accurate recognition / classification the output would be very close to the required output and thus to the specified integer values. How different is the actual output from the required output is measured with an error metric which may consist of an euclidean distance or actual classification rate based on minimum error.

The performance of ANNs is dependent of many parameters such as the weights of the connections between neurons, the number of hidden layers and neurons, and the codification of inputs and outputs. For this work the error reduction between the actual and required output is accomplished by means of the estimation of a suitable architecture for the ReANN. Two Genetic Algorithms (GA1 and GA2) were considered for this goal and an overview of these GAs is presented in Figure 4.

The chromosomes for GA1 consists of fixed-length vectors as the values of the genes representing the number of neurons and transfer function are the same for all hidden layers. Binary coding is used in order to perform uniform crossover which leads to 16-bits vectors. In contrast the chromosomes for GA2 consists of variable-length vectors as the number of neurons and transfer function are
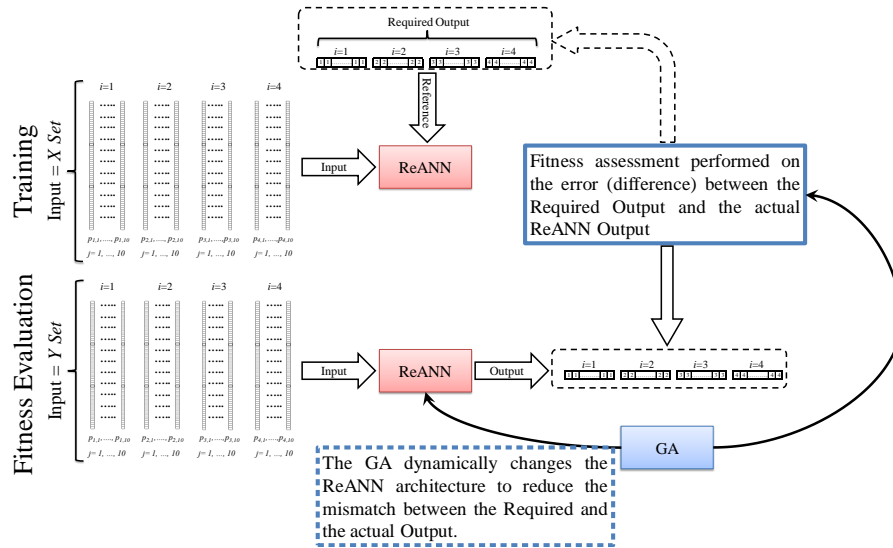
Fig. 3. Schemes for training and fitness evaluation for the ReANN and the GA.
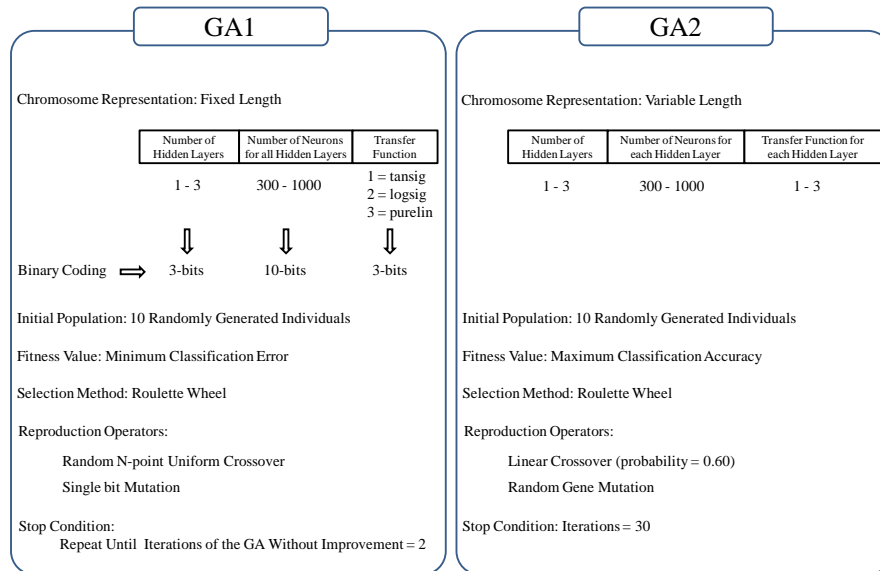

Fig. 4. Details of the Genetic Algorithms GA1 and GA2.

considered individually for each hidden layer. No binary coding is performed as linear crossover was used.

Each individual in the population for GA1 and GA2 represents an ANN configuration or structure which is trained with a set of images X (training set) and is evaluated with a set Y (GA fitness set) as presented in Figure 3. This process is performed for each individual which leads to ten output vectors where each one represents the coding of the most likely emotion i associated to each input feature vector p_{ij}. Then the euclidean distance between the actual output and the required (reference) output vector is computed and stored. The fitness value for each individual in GA1 represents the average value of the euclidean distances computed from the output vectors. In contrast for the GA2 the fitness value for each individual consists of the average correct classification percentage obtained with the associated ANN trained and evaluated on the X and Y sets respectively.

In Table 1 the constant parameters for the ReANN are presented while the architectures estimated by each GA are presented in Table 2.

Table 1. Constant settings of the ReANN

Setting	Value
Type	Feedforward
Training Algorithm	RPROP backpropagation
Epochs	1000
Error	0.0001

Table 2. Architectures estimated with the Genetic Algorithms GA1 and GA2

Genetic Algorithm	Architecture
GA1	Two hidden layers with 961 neurons for each one. Logsig transfer function for all hidden layers. Purelin transfer function for the output layer.
GA2	Three hidden layers: 1st layer with 859 neurons and purelin transfer function, 2nd layer with 456 neurons and logsig transfer function, 3rd layer with 144 neurons and tansig transfer function. Purelin transfer function for the output layer.

4 Corrective ANN based on Error Modelling

In order to improve classification an additional ANN termed as corrective ANN (CoANN) was designed. This ANN was established to refine the output of the ReANN considering the knowledge of the classification errors estimated during fitness evaluation. This process is similar to the modelling of a confusion matrix [2]. In Figure 5 the use of the CoANN is presented.
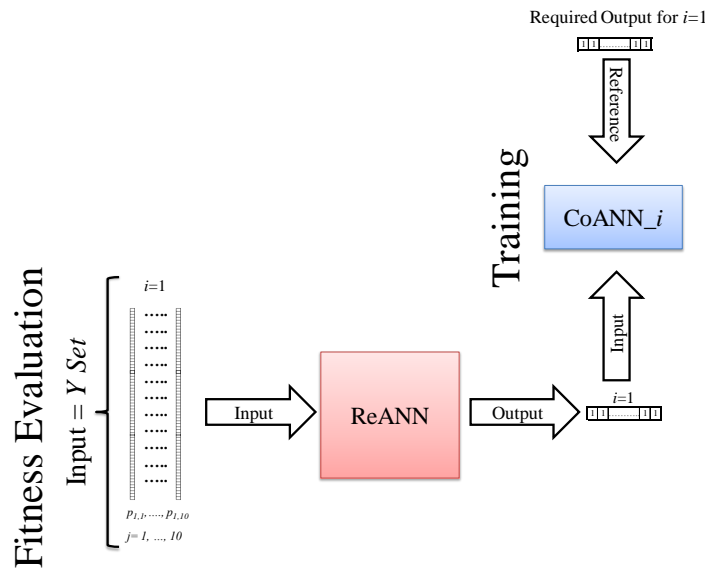

Fig. 5. Training of a CoANN for emotion $i=1$.

A CoANN was created for each i-th emotion. Due to the input and output vectors for the CoANNs being shorter than those for the ReANN the structure of the CoANNs was much simpler as presented in Table 3. No optimization was performed on these ANNs.

Under the integrated system ReANN + CoANN the recognition task is performed in two stages:

1. Initial classification is performed by the single ReANN where the input consists of image feature vectors and the output consists of real numbers associated to the most likely emotion.
2. The output of the ReANN is decoded by each CoANN$\_i$ in order to provide a refined output where the CoANN with the minimum error or euclidean distance is associated to the most likely i-th emotion.

Table 3. Constant settings of each *i*-th CoANN

Setting	Value
Type	Feedforward
Training Algorithm	Scaled Conjugate Gradient (SCG) backpropagation
Epochs	1000
Error	0.0001
Hidden Layers	2
Neurons	400 for 1st hidden layer, 200 for 2nd hidden layer
Transfer Function	Purelin

5   Performance of the ANN Systems

Each user in the JAFFE database has three image samples of each emotion. This made the separation of the database into three sets of images suitable as follows: the first sample images were used for the set X, the second sample images were used for the set Y, and the third sample images were used for the set Z. As presented in Figure 3 the sets X and Y were used for training and fitness evaluation purposes respectively. Thus final testing was performed with the set Z.

In Table 4 the recognition accuracy obtained with the GA-optimized ReANN on the set Z across all users in the JAFFE database is presented.

Table 4. Recognition Performance on the Test Set Z: GA-Optimized ReANN

ANN	Anger	Happiness	Neutral	Sadness	Total Recognition
with GA1	90.00%	90.00%	90.00%	30.00%	75.00%
with GA2	100.00%	80.00%	80.00%	50.00%	77.50%

While performance of the ReANN optimized with GA2 was higher than the performance of the ReANN optimized with GA1 (77.50% > 75.00%) both performances are under baseline reference performance available on the literature (up to 96% [6]). However this is expected as no complex feature extraction is performed (e.g., PCA or SVM extraction). In Table 5 the recognition accuracy obtained with the GA-optimized ReANN and the CoANNs on the set Z across all users in the JAFFE database is presented.

While the overall performance is still under baseline reported by other works an increase in classification was achieved over previous performance. It is important to mention that while recognition improved for Neutral and Sadness a decrease was obtained for Anger. However human perceptual data (semantic ratings) available for the JAFFE database [5] showed significant confusion of Sadness and Happiness with Neutral. By normalizing the semantic ratings pre-

Table 5. Recognition Performance on the Test Set Z: GA-Optimized ReANN with CoANN

ANN	Anger	Happiness	Neutral	Sadness	Total Recognition
with GA2 + CoANNs	80.00%	80.00%	100.00%	80.00%	85.00%

sented in [5] to recognition percentage it was obtained that the average level of correct classification was within the range of 80% to 90%. The performance presented in Table 5 is within this range.

6 Future Work

Considering the results obtained with the CoANN future work will consist in developing an emotion recognition system with the integration of pre-processing techniques as PCA, DWT or 2D-LDA. Because the corrective ANN (CoANN) produced a significant gain on the performance of the baseline ReANN, it may produce a significant gain on a recognition system with pre-processing. However this needs to be further studied with additional experiments. Also recognition experiments with more emotions and other databases must be performed in order to evaluate the effectiveness of the proposed method.

References

1. Anderson, I., Shippen, C., Juhasz, G., Chase, D., Thomas, E., Downey, D., Toth, Z., Lloyd-Williams, K., Elliot, R., Deakin, W.: State-dependent alteration in face emotion recognition in depression. The British Journal of Psychiatry 198, 302–308 (2011)
2. Caballero, O., Cox, S.J.: Modelling confusion matrices to improve speech recognition accuracy, with an application to dysarthric speech. In: Proc. of the International Conference on Spoken Language Processing (Interspeech 2007). pp. 1565–1568 (2007)
3. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.: Emotion recognition in human computer interaction. IEEE Signal Process Ma. 18(1), 32–80 (2001)
4. Lyons, M., Budynek , J., Akamatsu, S.: Automatic classification of single facial images. IEEE Trans. Patt. Anal. Mach. Intell. 21, 1357–1362 (1999)
5. Lyons, M. J.: Semantic Ratings Data: Japanese Female Facial Expressions (JAFFE), Database of digital images (1997). In: http://www.kasrl.org/jaffe_info.htmlm (Accessed on 12/05/2014)
6. Owusu, E., Zhan, Y., Mao, Q. R.: A neural-AdaBoost based facial expression recognition system. Expert Systems with Applications 41, 3383–3390 (2014)
7. Rao, K.S., Saroj, V.K., Maity, S., Koolagudi, S.G.: Recognition of emotions from video using neural networks models. Expert Systems with Applications 38(10), 13181–13185 (2011)

8. Seong, W.K., Park, J.H., Kim, H.K.: Dysarthric speech recognition error correction using weighted finite state transducers based on contextdependent pronunciation variation. In: Computers Helping People with Special Needs, Lecture Notes in Computer Science, Volume 7383. pp. 475–482 (2012)
9. Shih, F.Y., Chuang, C.-F., Wang, P. S. P.: Performance comparisons of facial expression recognition in JAFFE database. International Journal of Pattern Recognition and Artificial Intelligence 22(3), 445–459 (2008)
10. Yildirim, S., Bulut, M., Lee, C.M., Kazemzadeh, A., Busso, C., Deng, Z., Lee, S., Narayanan, S.: An acoustic study of emotions expressed in speech. In: Proc. Int. Conf. Spoken Language Processing (ICSLP '04). vol. 1, pp. 2193–2196 (2004)
11. Yu, F., Chang, E., Xu, Y. Q., Shum, H.Y.: Emotion detection from speech to enrich multimedia content. In: Proc. IEEE Pacific-Rim Conf. Multimedia 2001. vol. 1, pp. 550–557 (2001)
12. Zavaschi, T.H.H., Britto, A.S., Oliveira, L.E.S., Koerich, A.L.: Fusion of feature sets and classifiers for facial expression recognition. Expert Systems with Applications 40, 646–655 (2013)