A simple multi-feature classification to recognize human emotions in Images

Navin Ipe $^{1,1}$

$^{1}$M.S.Ramaiah University of Applied Sciences

November 8, 2023

Abstract

Emotion recognition by the human brain, normally incorporates context, body language, facial expressions, verbal cues, non-verbal cues, gestures and tone of voice. When considering only the face, piecing together various aspects of each facial feature is critical in identifying the emotion. Since viewing a single facial feature in isolation may result in inaccuracies, this paper attempts training neural networks to first identify specific facial features in isolation, and then use the general pattern of expressions on the face to identify the overall emotion. The reason for classification inaccuracies are also examined.
A Simple Multi-Feature Classification to Recognize Human Emotions in Images

Navin K Ipe

Abstract—Emotion recognition by the human brain, normally incorporates context, body language, facial expressions, verbal cues, non-verbal cues, gestures and tone of voice. When considering only the face, piecing together various aspects of each facial feature is critical in identifying the emotion. Since viewing a single facial feature in isolation may result in inaccuracies, this paper attempts training neural networks to first identify specific facial features in isolation, and then use the general pattern of expressions on the face to identify the overall emotion. The reason for classification inaccuracies are also examined.

Index Terms—emotion recognition, image processing, neural network.

I. INTRODUCTION

Images are fed to a computer algorithm as a series of pixels that form a pattern. Once a human face is isolated from this pattern, the recognition of emotions requires the identification of specific facial features, extraction and reduction of those features and classifying the overall understanding of the features. Facial expressions could broadly be described as motion or positions of facial muscles. Although many advanced methods of using Convolutional Neural Networks (CNN) have been developed, this paper attempts to go back to the basics to find the simplest possible technique of classifying emotions, as a process of examining the fundamentals of emotion recognition. Emotion recognition techniques generally utilize concepts of optical flow, flow vectors, principal component analysis (PCA), hidden Markov models (HMM) etcetera to recognize emotions. However, the realities of how the human brain processes emotion recognition, appears to stem from a far more fundamental process that accounts for context and associativity to one’s own capabilities to form expressions, than a mere interpretation of facial features. The interpretation of emotion from emoticons, birds or other animals are one such example. Therefore, rather than merely creating a database of facial expressions, a decision was taken to allow neural networks to learn from a set of models of each facial feature, to classify an emotion. The faces chosen for the experiment were from the JAFFE dataset containing faces pre-classified (as a ground-truth) into seven emotions as shown in Fig. 1.

Section II presents similar research on emotion recognition, Sect. III presents the design of the models, Sect. IV presents the results and the paper concludes with Sect. V.

Author acknowledges the support received from M S Ramaiah University of Applied Sciences, Bengaluru, India. N. K. Ipe is with the Department of Computer Science and Engineering, M S Ramaiah University of Applied Sciences, Bengaluru, India (e-mail: navinipe@gmail.com).

Manuscript received July 15, 2020

Fig. 1: Seven emotions: Angry, disgusted, fearful, happy, neutral, sad and surprised. Row1: Subject KA. Row2: Subject YM.

II. RELATED WORK

Certain emotion detection techniques utilize Bayesian networks, decision trees, Support Vector Machines (SVM), k-Nearest Neighbours (kNN), boosting and bagging. Other approaches have extracted geometric facial features, ducial areas and texture of skin (furrows, wrinkles etc.), supplied facial features as inputs to a single backpropagation neural network and classified the results using data about the eyes and mouth. Some researchers have used Sobel filters to emphasize prominent edges and extracted features using pixel density. Work by Konar and Chakraborty identified various methods of feature extraction, including facial, geometric model, appearance-based, voice features, gesture, posture-based features and even electroencephalogram-based methods. PCA, Independent Component Analysis (ICA) and evolutionary algorithms have been attempted for feature reduction. Neural networks, SVM, Learning Vector Quantization (LVQ), Fuzzy methods, HMM, kNN and Naive Bayes methods have been used for classification and various combinations of these approaches have been used as multimodal methods. More recent methods of emotion recognition utilize CNN's that have achieved more than 95% accuracy.

III. METHODOLOGY

The objective being to examine characteristics of facial features, the fundamental nuances of feature prominence were explored.

A. Feature prominence

In order to determine the minimum number of features required for identifying emotions, a few experiments were conducted by randomly viewing images from the dataset, keeping each feature hidden. It was noted that identifying the emotion was possible even when the eyes were hidden. Therefore, the eyebrow information, the nose furrow information, the mouth...
and chin involvement information were assumed to be sufficient. Figure 2 depicts how Otsu thresholding with multiple threshold levels, captured prominent facial features with the highest intensity of pixels, while illumination gradients of less relevant expressions were captured with lower intensities. Given the varied illumination, a feed forward neural network was considered for modeling the expressions, and since the faces were already centered, specific regions of interest were selected for examination, as shown in Fig. 3. Images were inverted to ensure the neural network received prominent features as the highest pixel values. Histogram equalization was initially attempted for pre-processing, but was not used since it reduced the prominence of some features.

B. Neural network design

A single feed forward neural network was initially trained to detect each facial feature. However, given the varied nature of eyebrow movements and other feature variations for any emotion among the JAFFE subjects, predictions from the neural network were inaccurate. This obstacle led to the idea of training a separate unifying neural network for each facial feature, which would interpret the emotion for each combination of emotions, as shown in Fig. 4. Each facial feature was examined under a grid of \( s \) number of squares as shown in Fig. 3. If greater accuracy was required for identifying any particular facial feature, the grid would be composed of smaller squares, thereby offering greater granularity. To train on the five facial features, five feed forward neural networks were used. They are abbreviated “NN” for convenience. The left eyebrow area handled by \( NN_1 \) and had \( s_1 \) number of squares. The right eyebrow area handled by \( NN_2 \), with \( s_2 \) squares and so on, where the \( n \) in \( s_n \) identifies the area of the facial feature being examined. The score value of a square is calculated as \( v = T_{\text{max}} \). Each square may contain pixels of various intensities, corresponding to various thresholds. The pixel value of the threshold in any square with the maximum number of pixels is the \( T_{\text{max}} \) value for that square. Scores are input to the neural networks as an array depicted in Eq. 1.

\[
S_n = [v_1, v_2, \ldots, v_{s_n}]
\]

Matlab’s neural network toolbox was used to create, test and validate the emotions. A sigmoid function was used as the activation function and training was performed with a scaled conjugate gradient. Data division randomly selected images during validation to find the stopping condition. Although training and validation were performed by the toolbox automatically, a decision was taken to re-evaluate the trained models on all 104 images, since the dataset was small. Using a larger dataset was problematic, since the toolbox ran into memory insufficiency errors.

Algorithm 1 Emotion detection

Step 1: Calculate scores \( S_n \) for training dataset.

Step 2: Train \( NN_n \) using \( S_n \) and ground truth \( E \).

Step 3: Train \( FN \) using outputs of \( NN_n \) as inputs and ground truth \( E \).

Step 4: Calculate \( S_n \) for test dataset and feed it as input to \( NN_n \).

Step 5: Feed outputs of \( NN_n \) to \( FN \) and obtain output \( E \).

IV. Results

Memory requirements and the fact that many images in the JAFFE dataset were sufficiently ambiguous that even a human would be unable to recognize the emotion, led to 104 images being manually chosen for trials from the entire JAFFE dataset. In trials that train with merely seven images, the set of images contain each of the seven emotions. The objective being, to observe the classification accuracy of training with just one image of each emotion.

A. Feed forward trial

In trials 1 to 3 of Table I the number of neurons used were \{\( NN_1:60, \) \( NN_2:60, \) \( NN_3:2, \) \( NN_4:2, \) \( NN_5:24 \}\). Trials 4 and 5 used \{\( NN_1:(30,10), \) \( NN_2:(30,10), \) \( NN_3:(6,3), \) \( NN_4:(6,3), \) \( NN_5:(30,10) \}\).
TABLE I: Feed forward trial with 5 thresholds

| Trial | Hidden layers | Training images | Granularity | Accuracy |
|-------|---------------|-----------------|-------------|----------|
| 1     | 1             | 104             | Low         | 25.05%   |
| 2     | 1             | 65              | Low         | 17.03%   |
| 3     | 1             | 7               | Low         | 10.57%   |
| 4     | 2             | 104             | Low         | 18.26%   |
| 5     | 2             | 104             | High        | 25%      |

Fig. 5: Regions fed to NN. The neural network actually “sees” a more feature-reduced version of these images.

\[
NN_1:(6,3), \quad NN_5: (48,20),
\]

where the numbers in parenthesis are the number of neurons used in each layer. The square sizes (in pixels) for low granularity were \{NN_1:5, \ NN_2:5, \ NN_3:15, \ NN_4:15, \ NN_5:10\} and all squares were of size 5 for high granularity trials. The final neural network \(FN\) had 20 hidden nodes in each of its 5 hidden layers in the low granularity setting. There were 104 hidden nodes in the first layer and 52 hidden nodes in the second layer for the high granularity setting. Results are presented in Table II where the pre-processing was performed with five thresholds. It was observed that classification accuracy was very poor, and increasing the number of hidden layers only lowered the accuracy. However, utilizing a higher granularity showed promise. The reason for this low performance was evaluated as shown in Fig. 5 where each of the areas being examined are shown in isolation, to view the emotion information from the perspective of the neural network. It can be observed that there is a significant similarity between the second “Happy” and the second “Surprise” images. Even the difference between the “Fear” and the first “Surprise” image are very subtle. The eyebrows and nose furrows of the “Angry” and the first “Happy” images are not significantly different. This led to the reasoning that a radial basis network may be capable of interpolating values better.

TABLE II: Radial basis trial

| Trial | Thresholds | Training images | Granularity | Accuracy |
|-------|------------|-----------------|-------------|----------|
| 1     | 3          | 65              | High        | 62%      |
| 2     | 3          | 104             | High        | 100%     |
| 3     | 3          | 7 (YM)          | Low         | 25%      |
| 4     | 3          | 7 (KA)          | Low         | 8.65%    |
| 5     | 3          | 65              | Low         | 66.34%   |
| 6     | 3          | 104             | Low         | 100%     |
| 7     | 5          | 104             | High        | 100%     |

V. Conclusion

Many emotion detection techniques utilize certain prominent features of the face, like the eyes, nose furrows, eyebrows and lips to classify emotions. Although the results from Sect. IV show promise in the capability of neural networks to recognize emotions by assimilating predictions of multiple features to estimate the overall facial emotion, this paper also seeks to highlight that techniques which rely on illumination or prominent facial features, require more information to accurately judge emotion. This work was limited by the number of images used for testing and training and the fixed positions of the area of the face being examined. A few techniques for future work to improve upon are:

- The ability to locate relevant pixels of a facial feature based on the distance from various parts of the face and interpolating the complete feature extents even with insufficient illumination or feature obscurity, based on past training of what a complete feature should be like. This is because the facial muscles subtly affect various portions of the face. These include the edges of the eyes, the forehead, the cheeks and the eyelids.
- Illumination of the face can be an obstacle in identifying features. Rather than use a neural network that generalizes features, neural networks would be capable of learning and classifying better, if categorized into networks based on the overall illumination (light source direction and intensity) of the image. This would assist in improved pre-processing for each category.
- Even context plays a large role in emotion detection. The incorporation of context from the image background can improve results (for the JAFFE dataset, the context would be the unremarkable background which implies a staged photo session, and an overall view of all photos of a...
candidate which would give the algorithm an estimate of which candidates expressed emotions prominently).

REFERENCES

[1] M. Z. Uddin and T.-S. Kim, “Facial expression recognition using independent component features and hidden markov model,” Emotion Recognition: A Pattern Analysis Approach, p. 113, 2014.

[2] P. MS and K. N. GM, “Modified emotion recognition system to study the emotion cues through thermal facial analysis,” Biomedical Research (0970-929X), vol. 28, no. 20, 2017.

[3] M. J. Lyons, S. Akamatsu, M. Kamachi, J. Gyoba, and J. Budynek, “The japanese female facial expression (jaffe) database,” in Proceedings of third international conference on automatic face and gesture recognition, 1998, pp. 14–16.

[4] N. Sebe, Y. Sun, E. Bakker, M. S. Lew, J. Cohen, and T. S. Huang, “Towards authentic emotion recognition,” in 2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583), vol. 1. IEEE, 2004, pp. 623–628.

[5] J. G. Rázuri, D. Sundgren, R. Rahmani, and A. M. Cardenas, “Automatic emotion recognition through facial expression analysis in merged images based on an artificial neural network,” in 2013 12th Mexican International Conference on Artificial Intelligence. IEEE, 2013, pp. 85–96.

[6] I. Perikos, E. Ziakopoulos, and I. Hatzilygeroudis, “Recognizing emotions from facial expressions using neural network,” in IFIP International Conference on Artificial Intelligence Applications and Innovations. Springer, 2014, pp. 236–245.

[7] A. Konar and A. Chakraborty, Emotion recognition: A pattern analysis approach. John Wiley & Sons, 2015.

[8] L. Nwosu, H. Wang, J. Lu, I. Unwala, X. Yang, and T. Zhang, “Deep convolutional neural network for facial expression recognition using facial parts,” in 2017 IEEE 15th Intl Conf on Dependable, Autonomic and Secure Computing, 15th Intl Conf on Pervasive Intelligence and Computing, 3rd Intl Conf on Big Data Intelligence and Computing and Cyber Science and Technology Congress (DASC/ICDCS/DataCom/CyberSciTech). IEEE, 2017, pp. 1318–1321.

[9] H.-D. Nguyen, S. Yeom, G.-S. Lee, H.-J. Yang, I.-S. Na, and S.-H. Kim, “Facial emotion recognition using an ensemble of multi-level convolutional neural networks,” International Journal of Pattern Recognition and Artificial Intelligence, vol. 33, no. 11, p. 1940015, 2019.