ISSN 1870-4069

Design of a Soft Emotion Sensor for Food Recommendation Using Deep Learning

Alberto Espinosa-Juárez, Marco A. Moreno-Armendáriz

Instituto Politécnico Nacional,
Centro de Investigación en Computación,
Laboratorio de Ciencias Cognitivas Computacionales,
Mexico

{aespinosaj2021, mam_armendariz}@cic.ipn.mx

Abstract. The analysis of emotions and moods in people allows us to know, among many things, tastes, which allows us to offer personalized products or services. This article shows a way of detecting and classifying people's moods using selfie photographs for subsequent use in restaurant dish recommenders. For this solution we propose to make use of two neural networks; the first one will be used for the detection of Action Units (AU) present in the works of Paul Ekman, and a second neural network to classify the correct mood once the AUs have been previously detected and classified.

Keywords: Emotions, facial, classification, neural networks.

1 Introduction

The emotions are present in us as humans at all times and is often the reason for choices in different areas of our lives, from whether or not to go out with a friend to whether or not to see a movie.

Paul Ekman, the pioneering psychologist in the study of human emotions and expressions, mentions in his numerous investigations the importance of studying them to understand part of human behavior and, in his study Universal and cultural differences in facial expression of emotion [6], he classifies for the first time the seven facial expressions associated with universal emotions: anger, disgust, fright, surprise, happiness, sadness, and contempt.

Although recent research, such as that published in Nature [5] shows that there can be up to 16 facial expressions, this is because they take into account the culture and the confusion or association that can be made of an expression to a particular emotion.

For a better reading of facial expressions,published The Facial Action Coding System (FACS)[8] and an update in 2002. FACS is a system that aims to measure all visually distinguishable facial movements and while this system has dozens of applications, mood analysis and states of mind is one of them. FACS is based on single action units (AU) distinguishable by a number (code). Each AU corresponds to a visually distinguishable facial activity. It also has a collection of head and eye movements and positions.

Fig. 1. Images present in the training set of FER2013.

FACS describes all visually distinguishable facial activity based on single action units and several categories of head and eye positions and movements. Each AU has a numerical code (the designation of which is quite arbitrary). Ekman and Friesen [7] first coined the term microexpressions, defining them as those that show a hidden emotion and that can last half a second or less, so, for this work, macro expressions will be used, these being longer lasting (between half a second and four seconds).

Paul Ekman [3] classifies facial expressions associated with emotions; however, in this article, we work with moods for the following reasons: One, conceptual clarity is a basis of science, and two, emotion biases behavior, whereas mood biases cognition.

The relationship between emotion and food is complex. Being happy, stressed or sad can make us choose a certain type of food, additionally, nowadays, compared to previous decades, we have available in restaurants a wider variety of dishes that we can choose from and this can also influence how we feel; there is bidirectionality between what we eat and part of our mood and also in the food we choose depending on our mood, for example, eating foods high in glycemic index (parameter present in carbohydrate foods that classify them taking into account the speed at which they are digested, absorbed and metabolized affecting blood glucose and insulin levels) is possibly a causal effect of depression [13].

Computationally, intelligent recommenders have become an important part of the industry, because based on information provided by the user, such as tastes, searches, weather and even location, it is possible to recommend a product or service. We are looking to create a food recommender that takes into account as a main parameter, the emotion of people through artificial vision. We will make use of convolutional neural networks, which, according to the state of the art, have proven to be highly efficient in image classification problems. However, the scope of this paper is the classification of Action Units (AUs) to human emotions.

2 State of the Art

In 2020, the book *What the face reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)* [12] was published, in which different applications of FACS are discussed.

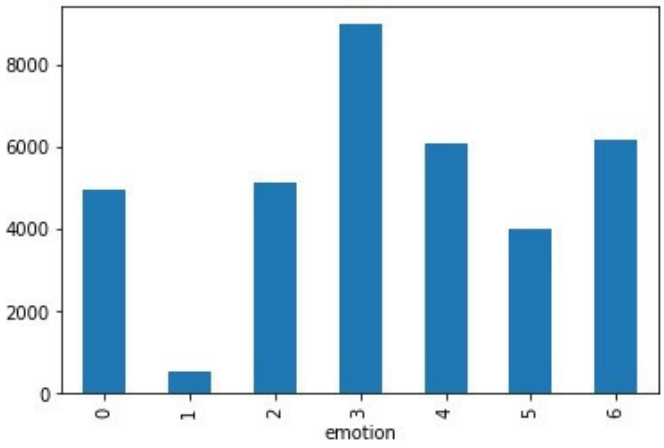

Fig. 2. Number of images per class in the FER2013 training set.

This book is divided into chapters and although there are many works dedicated in the book to emotion recognition and classification, in the chapter *The Next Generation of Automatic Facial Expression Measurement* a project on face detection and classification of facial expressions using support vector machines is described, achieving an accuracy of up to 93% accuracy in all the categories to be classified.

Li and Deng conducted a study on FER (*Facial Expression Recognition*) by analyzing the most popular datasets in this field: JAFFE, FER2013, SFEW 2.0, TFD, MMI and CK+. Each dataset was analyzed using different methods, such as the use of convolutional neural networks, recurrent neural networks, and MSV (support vector machines). In conclusion, they mention that the use of Deep Learning techniques (especially the use of convolutional neural networks) gives better results than the use of other types of neural networks.

However, despite finding congruence between what is learned by CNNs (Convolutional Neural Networks) and FACS Action Units, it is shown that they are unable to capture powerful convolutional features.

In *Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network* [11] the use of convolutional networks in attention for facial expression classification is proposed, achieving accuracies of 70% for FER2013, 99.3% for FERG, 92.8% for JAFFE and 98% for CK+.

Agrawal and Mittal [1] perform a study about how kernel size and several filters affect a convolutional neural network architecture proposed by the authors for facial expression classification, showing as conclusions a significant affectation between kernel size and several filters with the accuracy of the neural network using the FER2013 dataset as a basis, adding that they are of great help because they are not only simple in their architecture but also unique in terms of hyperparameter selection in the layers of the network.

In *Facial expression recognition with convolutional neural networks via a new face cropping and rotation strategy* [10], they achieve accuracies of up to 97% on CK+ and JAFFE datasets using a new face cropping and rotation method.

	emotion	pixels	Usage
0	0	70 80 82 72 58 58 60 63 54 58 60 48 89 115 121...	Training
1	0	151 150 147 155 148 133 111 140 170 174 182 15...	Training
2	2	231 212 156 164 174 138 161 173 182 200 106 38...	Training
3	4	24 32 36 30 32 23 19 20 30 41 21 22 32 34 21 1...	Training
4	6	4 0 0 0 0 0 0 0 0 0 0 0 0 3 15 23 28 48 50 58 84...	Training

Fig. 3. FER2013 dataset architecture.

Their method consists of generating new data to compensate for the imbalance in the datasets by randomly rotating and flipping the images with important conditions, such as rotating an image such that both eyes of the face are aligned in a straight line, cropping the faces, and spanning only from the edge of the eyebrows to the bottom edge of the chin.

Gogic and Ahlberg [9] propose a new facial expression classification method based on a trainable feature extraction process that uses sets of decision trees producing sparse binary feature vectors (LBF) and a shallow neural network. Such a method, according to its authors, is ideal for facial expression classification in resource-constrained environments such as mobile and embedded platforms.

2.1 Difference between this Proposal Concerning the State of the Art Presented Above

Currently, the use of FACS for facial expression analysis is rarely used and the preprocessing of the data often lacks measures for low and very high contrasts, as well as for varying light levels during the day.

We propose to use FACS to make a more complete dataset and in the case of machine vision, to set key points of interest in muscles using action units; we plan a modification of the datasets by including photographs taken in different locations (not only in the studio), as well as at different times of the day.

To balance the categories to be classified, we suggest the generation of random data using techniques seen in the state of the art, such as flips and flips of photographs looking for eye leveling. In the case of the neural network, we will continue to use convolutional neural networks because of their high efficiency in images and video, as demonstrated in the state of the art.

3 Solution Development

3.1 Description of the Solution to the Problem

The first step is the choice of the data sets to be used. The most used so far is called FER2013. This dataset contains 28,709 48x48 pixel grayscale images for use in training, 3,589 images for validation and 3,589 images for testing.

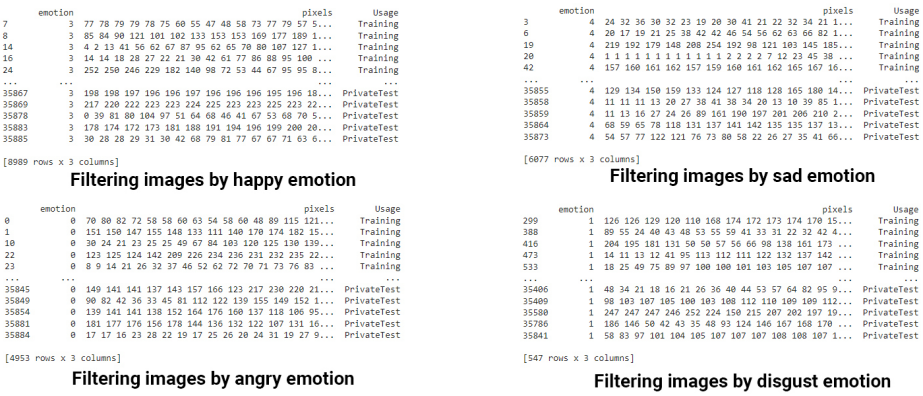

Fig. 4. Filtering of emotions in dataset FER2013.

The dataset has labels for universal emotions, represented as 0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral. However, when exploring the amount of data per class within FER2013, as seen in Figure 2, class 1 is found to be unbalanced. The reason for the unbalance is the small amount that the class has, which although a neural network model can be trained in that way, it would cause the training in that class to be null.

The architecture of FER2013 is based on 3 columns:

1. *Emotion*, corresponds to the type of emotion.
2. *Pixels*, containing a list of the number of values that make up the image (values from 0 to 255).
3. *Usage*, Usage, which identifies the recommended data type to use: *Training*, *Private* (Validation) and *Public* (Testing).

To make use of the pixels that form the images it was necessary to manipulate the data set. The original data type found in the column containing the pixels is of string type, so it was decided to transform a 1-dimensional tensor to achieve a more accurate and simple manipulation of the data. First, a conversion to list data type was made and consequently, the list was converted to a NumPy array (which allows us to represent collections of data of the same type in several dimensions and belongs to the NumPy numeric calculation library of Python), which allowed us to perform a reshape; for this reshape a size of 48x48 was chosen, which corresponds to the size of the image in the dataset indications (48x48 pixels) and only one channel (because the images are in black and white).

In the final part of the treatment, we chose to convert this NumPy array into a tensor (collection of n-dimensional vectors) of dimension 1, thus preparing the data for training and also allowing us to easily visualize it using the matplotlib library (Python library that allows us to make data graphics).

As the objective of this classifier of facial expressions corresponding to moods is to recommend a food dish in a restaurant, not all facial expressions are helpful. For this reason, we have decided to eliminate some facial expressions, modifying the dataset to focus on the ones we are interested in.

Fig. 5. Sample of the *Happy* category in FER2013.

Fig. 6. Sample of the *Sad* category in FER2013.

To achieve this, it was necessary to manipulate the original dataset using the pandas and NumPy Python libraries. With pandas, we filtered and eliminated those rows that correspond to facial expressions that are not of interest to us, and with NumPy we manipulated the list of pixels that correspond to each image to convert it into a NumPy array, to form a tensor that will be used to train the neural network and at the same time, show examples such as those presented below for each kind of facial expression that corresponds to a emotion. The final dataset was organized by the presence of only the following facial expressions:

– *Happy.*
– *Sad.*
– *Angry.*
– *Disgust.*

Fig. 7. Sample of the *Disgust* category in FER2013.

Fig. 8. Sample of the *Angry* category in FER2013.

Figures 5, 6, 7 and 8 represent examples found within the FER2013 dataset. We have recreated, thanks to the pixels provided by the same dataset a representative image.

It has been decided to leave only the categories of facial expressions mentioned above because as shown in some research, such as Salari-Moghaddam et al. [13] and AlAmmar et al. [2] and the correspondence between mood and food is between happy, sad and depressive moods.

To better encompass such mood states, the facial expressions of *Angry* and *Disgust* will become part of the sad or depressive mood state according to our next selection progress. Subsequently, manipulation of the dataset will be made, making modifications to the data it contains, that is to say, making a preprocessing.

During this preprocessing, we will try to use images with different shades of contrast and brightness, as well as with different backgrounds and in different locations, to have adequate variation and not fall with examples in the "ideal state" where all photographs used to comply with perfect shades of illumination.

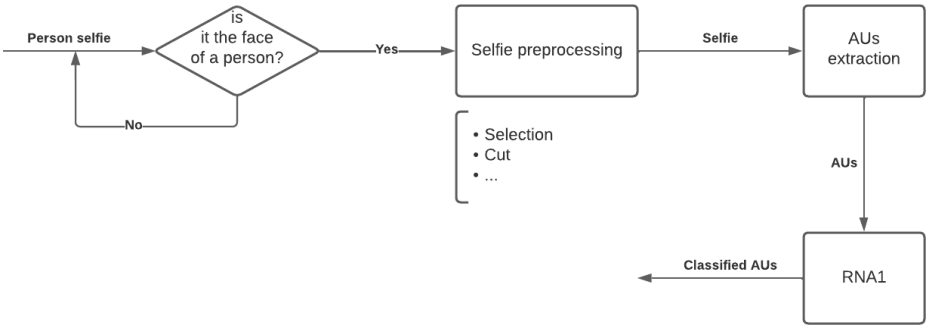

Fig. 9. General operation of the AU classifier.

In addition, we will try to generate new images to compensate for the unbalance of the dataset classes. To make sure that we correctly label the images we will use, we will make use of FACS, a system that will allow us to know, according to the AUs, to which facial expression each photograph corresponds. For this purpose, a scheme has been created using AUs that allows us to correctly classify the facial expressions.

Of all the main AUs [4], those that are not used for this work have been purged. The result is as follows:

0. Neutral Face,
6. Cheek Raiser: The orbicularis oculi muscle is lifted (below where dark circles usually appear),
4. Brow Lowerer: Wrinkling the glabella area (area between the two eyebrows) and home of the procerus muscle,
9. Nose Wrinkler: Actuating the levator muscle of the upper lip and the wing of the nose,
12. Nasolabial Deepener: Lifting the zygomaticus major muscle (located in the cheek); causes the smile,
15. Lip Corner Depressor: Actuating the depressor muscle of the angle of the mouth (located below the final corner of the lips),
16. Lower Lip Depressor: Movement in charge of the depressor labii inferioris muscle (located in the lower left middle part of the mouth).

For emotions, the AUs necessary to achieve them are already defined. In the case of the present work, the AUs will be used:

– Happyness: 6+12 (UA 6 plus UA 12),
– Sadness: 1+4+15 (UA 1 plus UA 4 plus UA 15),
– Disgust: 9+15+16 (UA 9 plus UA 15 plus UA 16),
– Neutral: 0.

Once all the images are labeled in the new dataset, the necessary features of the face are extracted based on FACS, and the image is cropped taking into account only the area of interest. With these data, a first neural network will learn to classify these AUs.

Once we have these individual AU classifications, we enter a second neural network, although this time, it will only give us a facial expression classification according to the previously detected AUs. In this way we make sure to find a facial expression more efficiently, guided by the FACS and implementing our solution.

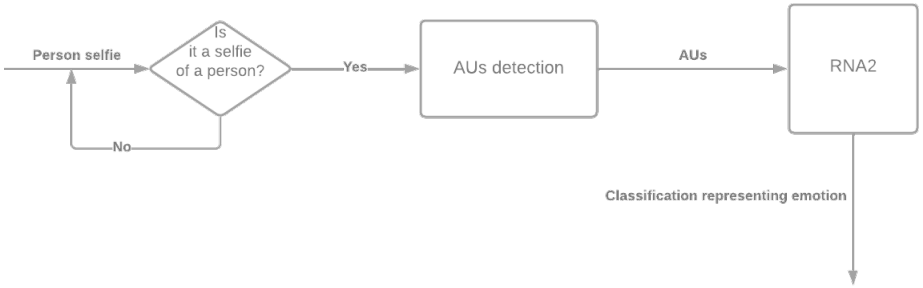

Fig. 10. General operation of the emotion classifier.

3.2 Scientific Novelty

Add to the state of the art a new way to analyze facial expressions that correspond to moods in images using FACS and deep neural networks.

3.3 Evaluation

The way to evaluate the correct performance of this facial expression classifier is based on the outputs of our neural network against different parameters, for example, making use of a k-fold cross-check, which allows us to give an estimate of the performance of the neural network based on unseen data, dividing the training set into subsets and then training all but one of them independently. This process is repeated until all subsets have been separated from the training iteration and the result is averaged across all models created.

4 Conclusions

The state of the art and theoretical research on saucer recommenders allows us to prove the feasibility of the work. The analysis of moods for the recommendation of products and services is an area of great interest that will allow us to make more appropriate recommendations to people, and the use of deep neural networks for the analysis of facial expressions that correspond to moods can be of help in the investigation of these because as Paul Ekman mentioned in his research and various psychologists today, moods lead us to choose, do and think differently than we would do depending on that state.

The future work is to find the ideal characteristics for the dataset, work on the formation of new data for this dataset that contemplate the modifications previously mentioned, work on an artificial vision to select the areas of interest in the faces and design the architecture of the convolutional neural network that will allow us to make the correct classification of the data.

Acknowledgments. This work has been possible thanks to the support of the Mexican government through the FORDECYT-PRONACES program of Consejo Nacional de Ciencia y Tecnología (CONACYT) under grant APN2017 – 5241; the SIP-IPN research grants SIP 2083, SIP 20220533; and IPN-COFAA and IPN-EDI.

References

1. Agrawal, A., Mittal, N.: Using CNN for Facial Expression Recognition: A Study of the Effects of Kernel Size and Number of Filters on Accuracy. The Visual Computer, vol. 36, no. 2, pp. 405–412 (2020). DOI: 10.1007/s00371-019-01630-9.
2. AlAmmar, W.A., Albeesh, F.H., Khattab, R.Y.: Food and Mood: The Corresponsive Effect. Current Nutrition Reports, vol. 9, no. 3, pp. 296–308 (2020). DOI: 10.1007/s13668-020-00331-3.
3. Beedie, C., Terry, P., Lane, A.: Distinctions Between Emotion and Mood. Cognition & Emotion, vol. 19, no. 6, pp. 847–878 (2005). DOI: 10.1080/02699930541000057.
4. Cohn, J.F., Ambadar, Z., Ekman, P.: Observer-Based Measurement of Facial Expression with the Facial Action Coding System. The Handbook of Emotion Elicitation and Assessment, vol. 1, no. 3, pp. 203–221 (2007). DOI: 10.1093/oso/9780195169157.003.0014.
5. Cowen, A.S., Keltner, D., Schroff, F., Jou, B., Adam, H., Prasad, G.: Sixteen Facial Expressions Occur in Similar Contexts Worldwide. Nature, vol. 589, no. 7841, pp. 251–257 (2021). DOI: 10.1038/s41586-020-3037-7.
6. Eckman, P.: Universal and Cultural Differences in Facial Expression of Emotion. In: Nebraska Symposium on Motivation, vol. 19, pp. 207–284, https://www.paulekman.com/wp-content/uploads/2013/07/Universals-And-Cultural-Differences-In-Facial-Expressions-Of.pdf (1972)
7. Ekman, P., Friesen, W.V.: Nonverbal Behavior and Psychopathology. In: Friedman, R.J., Katz, M.M. (eds), The Psychology of Depression: Contemporary Theory and Research, pp. 203–232, https://www.paulekman.com/wp-content/uploads/2013/07/Nonverbal-Behavior-And-Psychopathology.pdf (1974)
8. Ekman, P., Friesen, W.V.: Facial Action Coding System. Environmental Psychology & Nonverbal Behavior, (1978). DOI: 10.1037/t27734-000.
9. Gogic, M., Manhart, M., Pandzic, I.S., Ahlberg, J.: Fast Facial Expression Recognition Using Local Binary Features and Shallow Neural Networks. The Visual Computer, vol. 36, no. 1, pp. 97–112 (2020). DOI: 10.1007/s00371-018-1585-8.
10. Li, K., Jin, Y., Akram, M.W., Han, R., Chen, J.: Facial Expression Recognition with Convolutional Neural Networks Via a New Face Cropping and Rotation Strategy. The Visual Computer, vol. 36, no. 2, pp. 391–404 (2020). DOI: 10.1007/s00371-019-01627-4.
11. Minaee, S., Minaei, M., Abdolrashidi, A.: Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network. Sensors, vol. 21, no. 9, pp. 3046 (2021). DOI: 10.3390/s21093046.
12. Rosenberg, E.L., Ekman, P.: What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS). Oxford Academic, (2020). DOI: 10.1093/acprof:oso/9780195179644.001.0001.
13. Salari-Moghaddam, A., Saneei, P., Larijani, B., Esmaillzadeh, A.: Glycemic Index, Glycemic Load, and Depression: A Systematic Review and Meta-Analysis. European Journal of Clinical Nutrition, vol. 73, no. 3, pp. 356–365 (2019). DOI: 10.1038/s41430-018-0258-z.