Intelligent Systems Project

A.Y. 2020-2021

Michele Baldassini, Beatrice Lazzerini, Francesco Pistolesi

1. Overview

The aim of the first part of this project is to design and develop an intelligent system that measures a person's affective state based on various biomedical signals that are recorded by sensors.

The purpose of the second part is the classification of emotions on the basis of images depicting faces of people with different facial expressions related to a given emotion.

In general, an *affective state* (or *emotion*) can be described by two terms: *valence* and *arousal*. Valence means positive or negative affectivity, whereas arousal measures how calm or exciting the affective state is. For example, the reactions that occurred the last time you saw a scary movie or just before an important exam are the mark of high-arousal emotions. On the other hand, valence only codes emotional events as positive or negative.

An affective state *S* is thus described by a pair (v^S, a^S) in the valence-arousal space shown in Fig. 1, where v^S, $a^S \in [1, 9]$ are the valence and arousal levels, respectively.

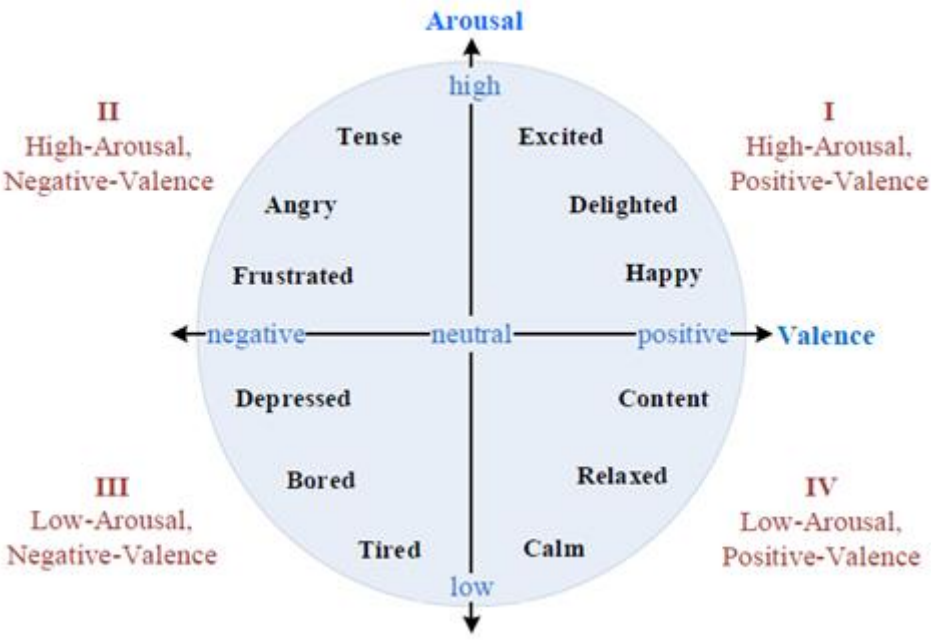

Figure 1. Two-dimensional valence-arousal space.

Emotions can also be expressed through classes characterized by appropriate values of valence and arousal, as shown in Fig. 1.

2. Datasets

2.1 Dataset with biomedical signals

A total of *58* participants took part in an experiment in which their biomedical signals were continuously recorded while they were watching a series of *36* excerpts of movie scenes. At the end of each video, each participant rated the emotion felt throughout the video, in terms of *valence* and *arousal* levels. A nine-point scale was used with a scale from 1 (very negative) to 9 (very positive) for *valence*, and from 1 (very boring) to 9 (very exciting) for *arousal*.

The types of signal recorded were:

- *Electrocardiogram (ECG*): records the electrical activity of the heart using electrodes placed on the participant's chest.
- *Electroencephalography (EEG)*: records the electrical activity of the brain using electrodes placed on the skin of the participant's head.
- *Galvanic skin response (GSR)*: records the electrical conductivity of the skin using two electrodes placed on the middle and index finger phalanges (conductivity increases in proportion to a person's sweating level).

A set F_s of features were extracted from each signal s. Let vector $\mathbf{f}_s \in \mathbb{R}^{|F_s|}$ contain the values of the features in F_s for signal s which is thus represented by $|F_s|$ features.

Dataset ***dataset.mat*** contains *1591* samples. Each sample is represented by a set of *54* features including *14* features extracted from the *ECG*, *11* features extracted from the *EEG*, and *29* features extracted from the *GSR*. Each row also contains the arousal and valence levels in the third and fourth columns, respectively.

The following tables summarize the features of each type of signal (ECG, EEG and GSR).

- ***ECG Features***

Column name	Electrocardiogram (ECG)
ECG_18	% of times the feature value is above mean + std (IBI)*
ECG_19	% of times the feature value is below mean – std (IBI)*
ECG_20 - ECG_25	Statistical measurements over heart rate (HR)
ECG_26 - ECG_31	Statistical measurements over heart rate variability (HRV)

** IBI = Interbeat interval*

- ***EEG Features***

Column name	Electroencephalography (EEG)
EEG _0	Average of first derivative
EEG _1	Proportion of negative differential samples
EEG _2	Mean number of peaks
EEG _3	Mean derivative of the inverse channel signal
EEG _4	Average number of peaks in the inverse signal
EEG _5 - EEG _10	Statistical measurements over the EEG channel

- ***GSR Features***

Column name	Galvanic Skin Response (GSR)
GSR_0	Mean skin resistance
GSR_1	Mean of first derivatives of skin resistance
GSR_2	Mean of absolute values of first derivatives of skin resistance
GSR_3	Mean first derivative for negative values only
GSR_4	Percentage of time with negative first derivative
GSR_5	Standard deviation of skin resistance
GSR_8 – GSR_11	Log power density estimates; 4 sub-bands in the [0-0.4] Hz band
GSR_12	Standard deviation of skin conductance
GSR_13	Mean of first derivatives of skin conductance
GSR_14	Mean of absolute values of first derivatives of skin conductance
GSR_15	Mean of absolute values of second derivatives of skin conductance
GSR_16	Average number of local minima in the skin resistance signal
GSR_17 – GSR_26	Log power density estimates; 10 sub-bands in the [0-2.4] Hz band
GSR_27	Zero crossing rate of skin conductance low response ([0-0.2] Hz)
GSR_28	Mean skin conductance low response peak magnitude
GSR_29	Zero crossing rate of skin conductance very slow response ([0-0.08] Hz)
GSR_30	Mean skin conductance very low response peak magnitude

Some of the extracted features are statistical measurements. Each row that mentions *statistical measurements* refers to the following features, in order:

- ***Statistical Features***

Columns	Statistical Measurements
1	Mean
2	Standard deviation (std)
3	Skewness
4	Kurtosis of the raw feature over time
5	% of times the feature value is above mean + std
6	% of times the feature value is below mean - std

2.2 Dataset of images

This dataset contains photos that show faces of people expressing different emotions. As a result of a labeling process, each image is associated with one of the following emotions:
Anger, Happiness, Fear, Disgust.

Figure 2 shows four examples of images included in the dataset. All images have a resolution of *224x224* pixels, and are in *RGB* format. The images are in folders named with the labels of the emotions. In the *happiness* folder there are images depicting happy faces, in the *fear* folder there are images depicting frightened faces, and so on.

Note that it is not required to use the entire set of images, but rather it is recommended to select a representative set of images so that the number of images per emotion is balanced among the four emotions.

Since the training of a convolutional network is computationally demanding and can take a lot of time, it is suggested to start considering two sets of images labelled with different emotions (e.g., *happiness* and *disgust*) and train the network to classify the images into the two chosen emotions. The proposal of solutions for the classification in three and four classes will be highly appreciated.

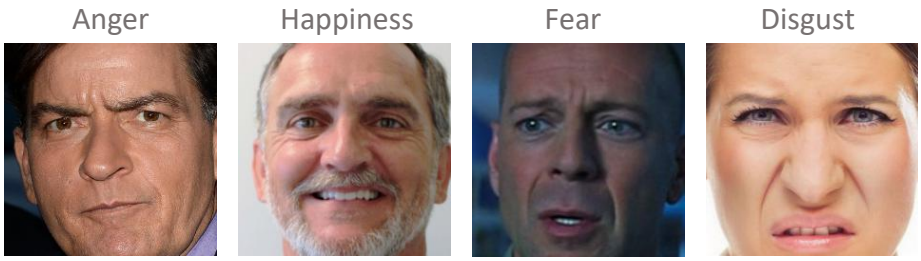

Figure 2. Examples of images contained in the dataset.

3. Requirements (Part 1)

3.1 Estimating valence and arousal with neural networks

The aim of this part of the project is to design and develop two multi-layer perceptron (MLP) artificial neural networks that accurately estimate a person's valence and arousal levels, respectively. The MLPs take as input a set of features that are selected from those described in Section 2.1 and return the corresponding valence and arousal levels, respectively.

The set of extracted features should be reduced by selecting the most significant features to predict the output. One way is to use the sequential feature selection (implemented by the `sequentialfs` MATLAB function), with a neural network as a criterion function that assesses the accuracy of each subset of features in estimating the valence (or arousal) level. The suggested maximum number of features to select is 10 for each network. Once the search for the best set of features is completed, the next step is to find the best architecture for both ANNs (see Fig. 3).

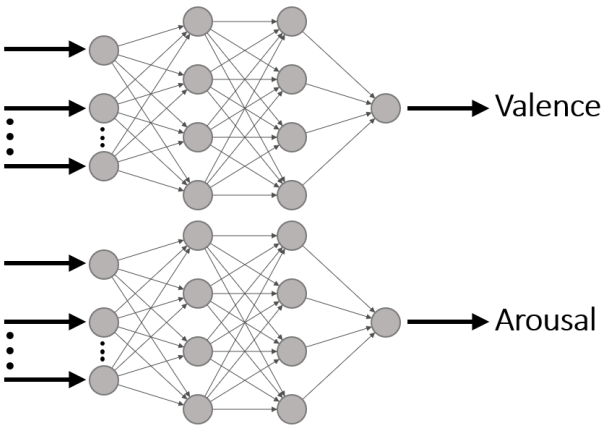

Figure 3. Overview of the two MLPs.

The last step of this part is to design and train two radial basis function (RBF) networks that do the same thing as the previously developed MLPs.

3.2 Classifying emotions with neural networks

The aim of this part of the project is to design and develop a multi-layer perceptron (MLP) artificial neural network that accurately classifies a person's emotion in one of the four quadrants of the valence-arousal plane. The MLP takes as input a set of features and returns the corresponding quadrant of the valence-arousal plane.

The set of features to use consists of the union of those selected and used to train the networks developed in Section 3.1. If these two sets have features in common, the set to use will contain a lower number of features (if a feature is in both sets, it must be considered as a single input). The goal is to find the best architecture for the MLP.

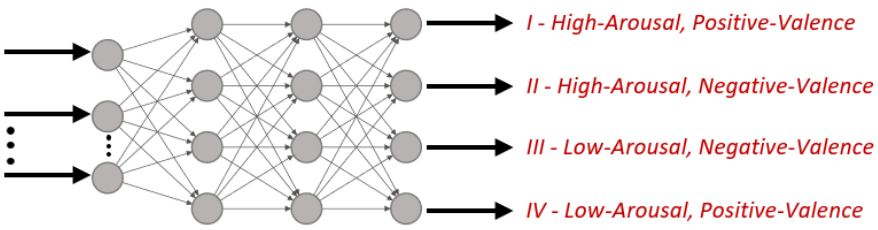

Figure 4. Overview of the classifier.

3.3 Mamdani-type fuzzy inference system

One of the major problems of the model described in Section 1 is that users are not familiar with the concepts of valence and arousal. So it is very common for people to feel a certain emotion but they are unable to correctly evaluate it using valence and arousal levels. For example, feeling fear and attributing valence and arousal levels belonging to the first quadrant of the plane means poor understanding of the model.

The aim of this part is to design and develop a fuzzy inference system to fix the deficiencies in the arousal dimension.

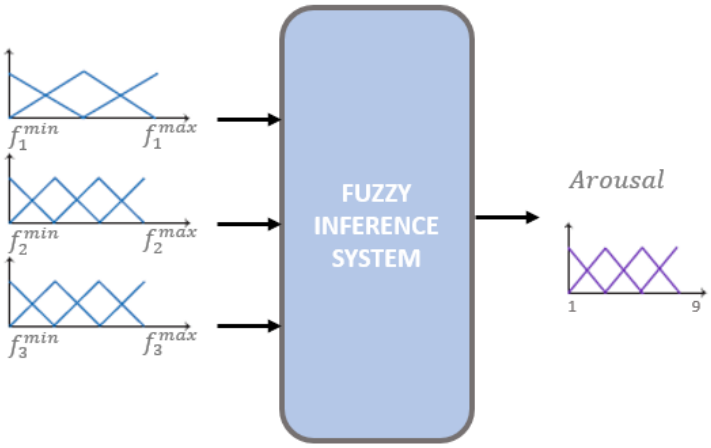

Figure 5. The fuzzy inference system.

The three inputs of the fuzzy system to design stem from an appropriate linguistic modeling of three features. The three features to model are the most relevant for estimating the arousal (i.e., the three most relevant features selected in Section 3.1). Linguistic modeling of inputs must be performed using appropriate sets of membership functions. Pay particular attention when designing linguistic modeling of features by analyzing the samples that are misclassified by the MLP, trying to fix the problem. The output of the system is a linguistic variable that expresses the arousal based on the values of the three input features. The linguistic modeling of the output must be performed by means of appropriate membership functions.

3.4 Improving MLP's performance using the fuzzy inference system

Use the output of the fuzzy inference system to train the MLP described in Section 3.1 to estimate the arousal level. This should be done with the aim of improving the accuracy, thereby reducing the bias. In order to achieve this result, the output of the fuzzy inference system must first be defuzzified and then used to replace the desired arousal level, where needed.

4. Requirements (Part 2)

4.1 Classifying facial expressions with convolutional neural networks

The aim of this part of the project is to design and develop a convolutional neural network (CNN) that accurately classifies a person's emotion, based on facial expression. The CNN takes an image as input and returns a class that represents an emotion. The goal is to find the best architecture for the CNN with the optimal configuration of hyperparameters.

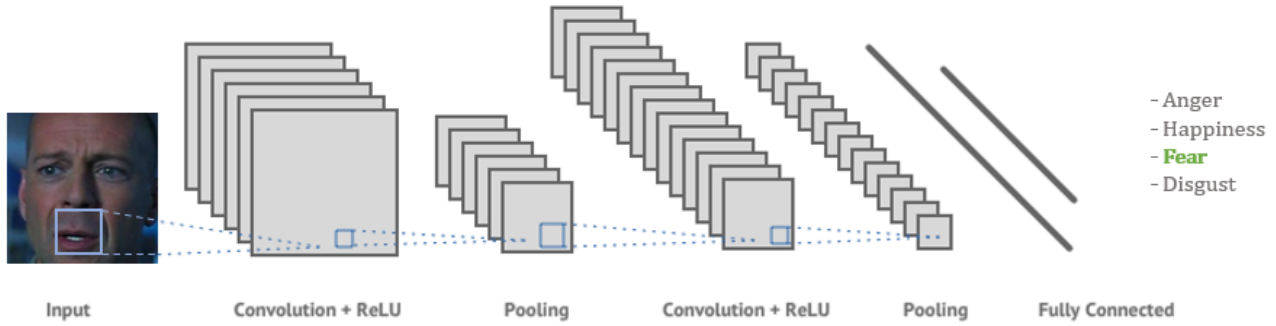

Figure 6. Overview of the CNN.

4.2 Classifying facial expressions with a pretrained CNN

The aim of this part of the project is to fine-tune a pretrained CNN to perform the task described in the previous section. The suggested pretrained network is A*lexNet* (*https://it.mathworks.com/help/deeplearning/ref/alexnet.html*). *AlexNet* is a CNN for object recognition in images, trained with a dataset of images that contain objects belonging to *1000* classes.

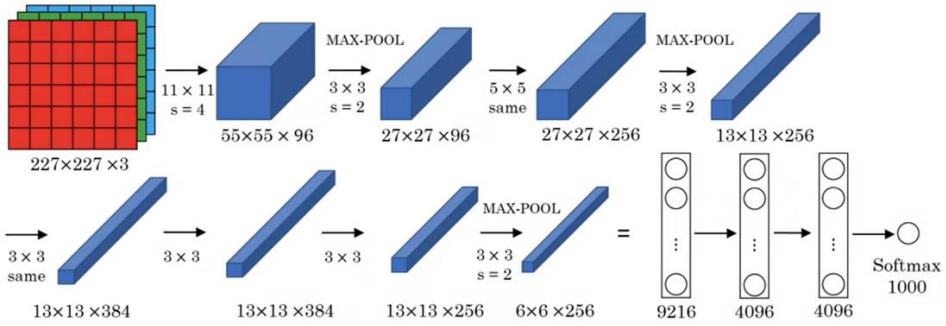

Figure 7. Overview of AlexNet.

Tasks to be carried out

This project can be carried out by groups of up to three students.

The optional tasks for each type of group are as follows:

- one-member groups: **3.2**, **3.4**, **4.1** (or **4.2**).
- two-member groups: **3.2**, **3.4**.
- three-member groups: **3.4**.