Towards a better Understanding of Human Emotions

Author: Hajer GUERDELLI
Department of Information Engineering (DINFO)
Laboratory: Media Integration and Communication Center (MICC)

PhD program in Smart Computing

Abstract

A major challenge in automatic human emotion recognition based on facial expressions is that of categorizing the very broad and complex spectrum of human emotions. In this regard, a critical bottleneck is represented by the difficulty in obtaining annotated data to build such models. Indeed, all the publicly available datasets collected to this aim are either annotated with the six prototypical emotions, or continuous valence and arousal values. On the one hand, the six basic emotions represent a very coarse approximation, and are not really useful to understand a person's emotional state. In this paper, we propose a solution to bridge the gap between the two. We propose to leverage valence and arousal to obtain a fine-grained taxonomy of emotions, interpreting emotional states as probability distributions over the valence-arousal space. In doing so, we can automatically annotate existing datasets with this new taxonomy, allowing the development of novel solutions for fine-grained emotion classification. Preliminary results obtained by training a model to classify the extended set of emotions underline the difficulty of the task, setting a new target in automatic human emotion recognition.

PROPOSED METHOD

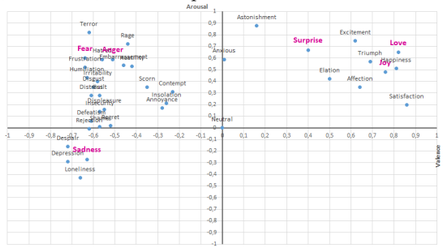
Describing an emotional state of a person with only the six prototypical emotions does describe her/his expression but does not give us the person's emotions. In the work of Russell [1], a set of 151 terms were provided with the corresponding distributions of *valence* and *arousal*. The 151 terms not only describe the emotions in a precise way but also the emotional interaction between two people. In the same context, Parrott [2] defined a taxonomy for emotion-related terms. In Parrott's theory, the classification of emotions is proposed in a tree-like structure starting with six primary, then secondary and tertiary emotions, for a total of 136 terms. The Parrott's tree does not provide the measure of valence and arousal, which we need in our work. Hence, the idea of taking the common terms between Russell's work and Parrott's classification to have the valence and arousal of each emotion with respect to a single individual.

0.1 Terms selection

To define a larger set of emotional terms, we first merged the terms of Russell and Parrott, then we selected the similar terms to get Parrott's structure and Russell's values for valence and arousal. This process, resulted into 32 terms:

- First structure emotion: Love, Joy, Surprise, Anger, Sadness, Fear
- Second structure emotion : (Love): Affection, (Joy): Happiness, Elation, Satisfaction, Excitement, Triumph,(Surprise): astonishment, (Anger): lrritability, Annoyance, Frustration, Rage, Hostility, Hatred, Scorn, Disgust, Contempt(Sadness): Depression, Despair, Displeasure, Shame, Guilt, Regret, Refeatism, Embarrassment, Humiliation, Insecurity, Isolation, Loneliness, Rejection, (Fear): Terror, Anxietty, Distress

The proposed 6 terms + 32 terms represented in the Valence and Arousal space.



EXPERIMENTS

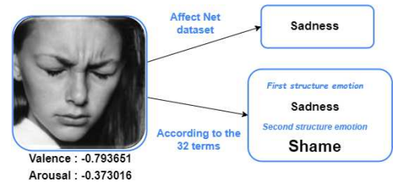
0.2 Datasets

In the experiments, we used AffectNet dataset [3]. AffectNet is a large facial expression dataset in the wild with around 0.4 million images manually annotated for the presence of seven facial expressions, along with the intensity valence and arousal.

0.3 Experimental setting

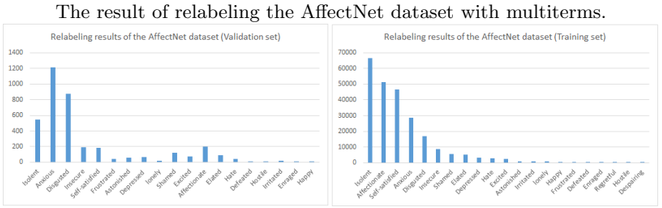
We used the Distract your Attention Network (DAN) [4]. The DAN model is proposed for the facial expression recognition, we used it as classifier using our proposed annotations as classes to perform emotion recognition.

Example of relabeling an image from the AffectNet dataset according to the proposed clustering approach.



0.4 Results

The result of relabeling the AffectNet dataset with multiterms.



The Results of our experimentation shows an accuracy of **26%** compared to the original accuracy of **64%** with 8 terms. Overall, observing the performance of DAN network gives the impression of the increased difficulty of this task, which demands for the ability of learning a very detailed and effective representation of the face so that all the slight diversities characterizing the human spectrum of emotions can be captured.

References

[1] Russell, James A. et Mehrabian, Albert., Evidence for a three-factor theory of emotions., Journal of research in Personality, vol. 11, no 3, p. 273-294, (1977)

[2] Parrott, W. Gerrod (ed.)., Emotions in social psychology: Essential readings., psychology press, (2001)

[3] A. Mollahosseini, B. Hasani, and M. H. Mahoor, Affectnet: A database for facial expression, valence, and arousal computing in the wild, IEEE Trans. on Affective Computing, vol. 10, no. 1, pp. 18–31, (2017)

[4] Z. Wen, W. Lin, T. Wang, and G. Xu, "Distract your attention: multi-head cross attention network for facial expression recognition," arXiv preprint arXiv:2109.07270, (2021)

Vote here for this poster: