A

AI-Based Emotion Recognition

Mousami Prashant Turuk
https://orcid.org/0000-0002-0103-9789
Pune Institute of Computer Technology, India

Sreemathy R.
Pune Institute of Computer Technology, India

Shardul Sandeep Khandekar
Pune Institute of Computer Technology, India

Soumya Sanjay Khurana
https://orcid.org/0000-0001-7336-1382
Pune Institute of Computer Technology, India

INTRODUCTION

Due to the recent advancements in technology, humans can interact with computers in ways that were previously unimaginable. Human-computer interaction is a multi-disciplinary field that focuses on designing computer technology to ease the interaction between computers and humans. New modalities such as voice and gestures are used to interact with computers that extend the traditional methods confined to keyboard and mouse. For human-to-human communication, voice and vision play a significant role. Thus, it is desirable for computers to comprehend the environment from visual as well as audio cues. This desire is supported by the growth in computer vision, natural language processing and the era of machine learning and deep learning which has helped to model the real world. Machine learning has provided a means for machines to extract useful information from images as well as speech. Various machine learning applications like image classification, image segmentation, object detection, text understanding and pattern recognition are being used on a day-to-day basis. Even with such advancements in the field, machines still fail to understand the 'emotion' of the person and this might lead to a failure in understanding the context provided entirely. In the current era of Industry 4.0, due to the availability of huge amounts of data, industries in every field are using artificial intelligence to tackle the problem of pattern recognition. Emotion is a mental or psychological state which is mainly associated with feelings, thought process and behavior of humans. Emotional state of a person conveys not only his mood but also his personality. Humans are able to exchange information through multiple domains like speech, text and visual images. In verbal communication, the same word expressed in different emotions can convey different meanings. Identification of emotional states using only audio cues is hence inadequate and needs to be in fusion with visual cues. This chapter aims to analyze and present a unified approach for audio-visual emotion recognition based on back propagation algorithm.

Emotion is a concept involving three components:

DOI: 10.4018/978-1-7998-9220-5.ch049

- Subjective experience.
- Expressions (audio-visual: face, gesture, posture, voice intonation, breathing noise).
- Biological arousal (heart rate, respiration frequency/intensity, perspiration, temperature, muscle tension, brain wave).

After recognizing universality within emotions despite the cultural differences, (Ekman et al., 1978). classified six emotional expressions to be universal: happiness, sadness, anger, disgust, surprise and fear.

Computer vision techniques have enabled the computer to understand the environment. Interacting with computers in voice and gesture modalities is much more natural for people, and the progression is towards the kind of interaction between humans. Despite these advances, one necessary ingredient for natural interaction is still missing, that is emotions. Emotions play an important role in human-to-human communication and interaction, allowing people to express them beyond the verbal domain. The ability to understand human emotions is desirable for the computer in some applications such as improving driver safety, medical conditions and lie detection. This chapter recognizes human emotions based on audio-visual cues.

Prosodic features in the audio and facial emotions exhibited on the face can help the computer make some inferences about the user's emotional state. The emotional frontier is the next obstacle to be surmounted in understanding humans. Facial expressions can be considered not only as the most natural form of displaying human emotions but also as a key non-verbal communication technique. If efficient methods can be brought about to automatically recognize these facial expressions, striking improvements can be achieved in the area of human computer interaction. Research in facial emotion recognition has been carried out in hope of attaining these enhancements. Moreover, there are other applications which can benefit from automatic facial emotion recognition. Artificial Intelligence has long relied on the area of facial emotion recognition to gain intelligence on how to model human emotions convincingly in robots. Recent improvements in this area have encouraged the researchers to extend the applicability of facial emotion recognition to areas like chat room avatars, video conferencing avatars, lie detection etc. The ability to recognize emotions can be valuable in face recognition applications. Suspect detection systems and intelligence improvement systems meant for children with brain development disorders are some other beneficiaries.

The area of human-computer interaction (HCI) will be much more effective if a computer is able to recognize the emotional state of human beings. Emotional states have a greater effect on the face which can predict the mood of the person. Faces are accessible windows into the mechanisms which governs emotional and social lives. About 70% of human communication is based on non-verbal communication such as facial expressions and body movements.

Despite the many theories, it is evident that people display these expressions to various degrees. One frequently studied task is the judgment of emotions—how well can human observers tell the emotional expressions of others, in the voice, on the face, etc.? Related questions are: Do these represent their true emotions? Can they be convincingly portrayed? How well can people conceal their emotions? In such tasks, researchers often use two different methods to describe the emotions.

One approach is to label the emotions in discrete categories, that is, human judges must choose from a prescribed list of word labels, such as joy, fear, love, surprise, sadness, etc. One problem with this approach is that the stimuli may contain blended emotions.

Also, the choice of words may be too restrictive, or culturally dependent. Another way is to have multiple dimensions or scales to describe emotions. Instead of choosing discrete labels, observers can indicate their impression of each stimulus on several continuous scales, like pleasant–unpleasant, atten-