

Affective Computing: Research Advances and Future Challenges

Shiyi Shen^{1,a,*}

¹*School of Computer Engineering and Science, Shanghai University, Shanghai, China*
a. shensy2004@shu.edu.cn

**corresponding author*

Abstract: Emotion, which is an unavoidable part of human beings, is getting hot in cognitive science these years. Affective computing, which is based on the research progress of human intelligence and emotion in psychological science, as a research field composed of many disciplines, has attracted much attention in recent years. There are many research directions and challenges to be solved. At the same time, it also has broad application prospects. As a result, it is particularly difficult for those eager to understand affective computing to get a simple but general understanding quickly. By collecting the research results mainly distributed in the past four years, this paper summarizes the overall research status in this field, including the development status of technology, practical application, and possible ethical issues. The goal of this paper is to provide a basic but comprehensive picture for researchers who are eager to get a preliminary understanding of this field.

Keywords: Effective computing, Emotion recognition, Emotion analysis, Artificial intelligence.

1. Introduction

In psychology, emotions are identified as a state of belief that causes mental and physical changes, which is an unavoidable part of all types of communication. There are three main theories of emotion. Physiological theories suggest that there is something in the human body that causes emotions. Neurological theory suggests that emotions are generated by activity in the brain. Cognitive theory holds that thoughts and other mental activities are important in shaping emotions [1]. Changes in expression and physiological indicators always accompany changes in emotion. For a long time, emotion did not receive academic attention. It was not generally accepted until the end of the 20th century. In modern times, cognitive scientists have regarded the interaction of emotion with other cognitive processes as a research hotspot. Affective computing is also getting hot.

Affective computing aims to create computing systems that can recognize emotions and act on them in a friendly way. Affective computing can help people understand the emotional world of others and themselves, and it can also give people a better emotional experience when they use apps. Therefore, affective computing has a wide range of applications. Nowadays affective computing involves many research fields and is a highly integrated research direction. It not only depends on the research progress of human intelligence and emotion in psychological science but also needs the support of technological updates in the field of computer science. This multidisciplinary situation has brought many challenges, including how to obtain emotional information, how to identify emotion,

and how to express emotion. In practical applications, there are many ethical issues to be discussed. In summary, due to the many research directions and future challenges, it is very difficult to have a preliminary general understanding of this field. The purpose of this paper is to provide an overview of the current state of research and future challenges in different directions and to give those interested in affective computing a general picture of the field.

The second part of this paper describes the development of technology, including emotion feature extraction, emotion analysis, and emotion generation. In the third part, the application status and prospects of affective computing in four fields of education, games, medical care, and security are summarized. In the fourth part, the ethical dilemmas of affective computing in affective definition, data acquisition, and practical application are introduced. Finally, the paper makes a conclusion in the fifth part.

2. Development of Techniques

2.1. Affective Feature Acquisition

2.1.1. Emotion Model

There are two kinds of emotion models. One is the Categorical Emotion Model, which divides emotions into discrete categories. The most basic one is the binary method, which divides emotions into positive, negative, and neutral. Some more detailed categorizations divide emotions into four or more categories. For example, in [1], emotions are divided into six categories: anger, disgust, fear, joy, sadness, and surprise. However, it has been found that [2] using this kind of classification, the exact emotional state may not be placed in the appropriate category because there is no suitable tag. Thus, another kind of emotion model appears. It is the Dimensional Emotion Model. It takes emotion as a highly correlated and continuously changing entity [3]. Instead of using discrete values to represent emotions, DES defines emotions using continuous vectors. One of this kind's most widely used models is the Valence-Activation-Dominance Space [4].

2.1.2. Data Type

For the judgment of data, in all emotion recognition techniques, researchers are free to choose the scale, feature construction, and experiment period of the data set without a united conception of data sets [5]. Some studies try to establish a unified dataset quality analysis method to improve the accuracy of sentiment detection [6].

There are two kinds of signals for emotion recognition.

The first one is the physiological signal. To identify emotions, human parameters and physiological signals can be captured through sensors, including Electroencephalography (EEG), Electrocardiography (ECG), Galvanic Skin Response (GSR), Heart Rate Variability (HRV), Respiration Rate Analysis (RR), Skin Temperature Measurements (SKT), Electromyogram (EMG), Facial Expressions (FE), Body Posture (BP), and Gesture Analysis (GA) [5]. Among these signals, the most popular is the EEG signal. Electrical activity in the brain can show detailed changes in emotional states. It is more intuitive and reliable to detect emotional states than other methods [7]. Facial expressions, as relatively accessible physiological signals, have also received great attention in fields such as human-computer interaction, marketing, and healthcare [8]. With the development of smart wearable devices, it is possible to capture all kinds of physiological signals in real-time in the natural environment [9], so affective computing using other physiological signals may receive more attention in the future. Some studies explore less commonly used physiological signals, such as [10]. It discusses how dynamic displays express information better than static images for emotion recognition.

The next one is the external signal. Emotions cause people to express information to the outside world. Therefore, it is feasible to analyze and recognize emotions through text, speech, and other external signals. Speech, as the fastest form of human communication, hides valuable information and has long been an important research field [11,12]. Speech emotion recognition (SER) emphasizes acoustic features associated with emotion. In recent years, end-to-end methods and more multi-dimensional feature learning have gained attention [13]. With the development of large language models and the more frequent use of chatbots, text-emotional detection has become more important. It is seen as a new way to prevent and detect mental health conditions [14]. It is more often combined with audio in a multimodal emotion analysis system [15,16].

2.2. Emotion Analysis

2.2.1. Data Processing

External signals are not as intuitive as physiological signals. For text and speech, relying on context can help better categorize emotions. However, differences in personal identity (i.e., gender, age, living environment, culture, etc.) can still lead to false correlations in the analysis of emotions. Therefore, ruling out false correlations has become one of the problems that researchers are keen on. Some studies choose to reduce the overall error through technical improvements [17]. Other studies distinguish different groups of people to identify differences between certain groups [18].

With the rise of multimodal emotion recognition, how to deal with different kinds of signals to let them play a complementary role and improve models' robustness and accuracy has become an attentive problem [19].

2.2.2. Algorithms

Researchers often choose machine learning or deep learning for emotion recognition.

Machine learning is letting computers learn from old data and analyze patterns to make inferences about new data. The most commonly used machine learning algorithms are support vector machines (SVM) and random forests (RF). RF makes predictions by combining multiple decision trees. SVM predicts by finding the optimal hyperplane in a high-dimensional space. Both of them are suitable for classification tasks and can improve the accuracy and robustness of the results. They also work well with high-dimensional data. SVM's ability to draw clear boundaries between classes has made it the most popular machine-learning algorithm to categorize different emotions [20].

Deep learning uses artificial neural networks with many layers of large-scale data to find better models and representations. In the field of emotion recognition, convolutional neural networks (CNNs), recurrent neural networks (RNN), generative adversarial networks (GAN), Long Short-Term Memory (LSTM), and transformers are very popular. Deep learning can extract abstract features efficiently and often has better detection accuracy than machine learning models [21]. The fusion of the attention mechanism allows weight to be calculated adaptively, and the change of weight can improve the correlation between different modalities, thus optimizing multimodal emotion recognition [21].

2.3. Emotion Generation

After classifying emotions, how to effectively integrate them into content generation is a problem. The object being generated is usually text, image, or speech.

For the generation of dialogue, how to make it produce an empathic response has attracted much attention. The research focuses on capturing the emotional relevance of past conversations. After generating text, it can also be converted to speech, which is about speech generation. Speech

generation focuses on how to make the generated speech natural and expressive. Besides the understanding of contextual emotion, the analysis and imitation of human speech characteristics are also very important. Image generation is always paired with speech generation since static images alone are difficult to fully correlate with emotion. So the research focus is to explore how to generate head animation that is consistent with the generated speech. This means that the generated head must have a natural and matching facial expression and the right lip movement [22].

3. Applications

As emotion is an indispensable part of human beings, with the rapid development of emotional computing, it has been widely used in games, education, medical care, security, and other fields.

3.1. Education

Intelligent Tutoring Systems (ITS) aims to personalize educators' and students' experiences. Adding affective computing lets computers recognize and respond to emotions, allowing for better-customized learning schemes [23]. It can track students' emotional states in real time and let educators know how students are currently learning. On this basis, educators can give emotional support and adjust teaching strategies [24].

The combination of the two can also be seen in the field of special education. For example, in [25], a social robot that can recognize emotions can act as a learning partner to reduce students' social isolation and help children with autism spectrum disorders receive an education as good as anyone else.

Emotion recognition based on facial expression, voice, or text is popular due to its easy real-time acquisition. In the field of facial recognition, changes in lighting, head posture, etc., pose challenges. In speech analysis, choosing which features to separate from the background noise is a problem. In the text emotion recognition, how to face the subjectivity of people and the result of the conflict remain to be solved [23].

3.2. Gaming

Through feedback from game users, producers can improve the game experience. Therefore, analyzing user emotions is an integral part. In addition, emotion generation can also help adjust game content to induce users' emotions correctly. For example, in [26], the research team used convolutional neural networks and data captured by head-mounted displays (HMD) to predict emotions, pioneering the emotion analysis of game users in virtual reality (VR) games.

In addition to being used to analyze user feedback, affective computing can also be used directly in games. For example, in [27], facial emotion recognition can generate avatars of virtual characters in future meta-universe and video games. Many developers try incorporating social robots into their games to create more lifelike non-player characters (NPCs). In the player's communication with them, emotional recognition of text or speech is necessary.

Emotional computing is more relevant to serious games (games used in teaching, medicine, etc.) than entertainment games. For these games to transfer knowledge, sensing the emotions of "students" and teaching them based on emotions can greatly improve the teaching effect. For example, in the experiment [28], the research team trained a lightweight CNN model to predict seven emotions. The predicted results will be used in a serious game that helps children with autism learn how to express their emotions.

3.3. Healthcare

Emotion recognition has received more and more attention in the medical field. Researchers are working to develop systems that can accurately identify emotions to monitor patient health and respond quickly.

Among them, the accuracy of multi-mode emotion recognition is higher, and the speed of single-mode emotion recognition is faster. Due to more powerful feature extraction capabilities, deep learning is getting more attention than machine learning. The study [29] used deep learning and multimodal data (including physiological data such as facial expressions and EMG) to build a framework for monitoring patients' emotional health. A soft attention algorithm retrieves the most important physiological signals. However, the experimental results show that multimodal emotion recognition is still difficult to reach, and also the framework has not been tested in a real-time environment. However, although the research status is not yet able to achieve the goal, emotion recognition in the medical field theoretically still has a lot of space for future development. In addition to being used to monitor health status in real-time and suggest countermeasures, it can also be used to improve communication between patients and doctors and to facilitate the diagnosis and treatment of mental health problems [30].

3.4. Security

In terms of security, whether in a public or a home environment, it is more convenient to obtain external signals. Therefore, the use of voice and facial emotion recognition is predominant. By perceiving negative emotions such as fear, dangerous events can be prevented in the public environment. At home, the indoor environment can be changed by acknowledging bad emotions to improve the comfort level of residents.

However, the use of voice or facial data, while bringing personal safety and improved personal comfort, also raises privacy concerns. How to balance protecting personal privacy and public safety [31].

4. Ethical Problems

4.1. Classification of Emotion

Most types of emotion classification classify emotions into fixed categories and then focus on signal analysis into those categories. Some emotional classification methods try to show continuous changes between different categories of emotion. However, these categories ignore the complexity of emotions.

The first point is the directivity of emotion, that is, the reason behind the emotion. The third is whether your emotions are controlled. Uncontrolled emotions are often difficult to regulate and may not even be perceived by the individual. Reacting to these emotions may not work and may even be seen as offensive.

The second point is the mixture of emotions. Human emotions are often mixed, consisting of several emotions. Current emotion classification can only recognize one emotion or a continuous switch between two emotions, and cannot perceive multiple emotions at the same time.

The third is whether your emotions are controlled. Controlled emotions are very different from the uncontrolled version [32].

4.2. Composition of The Data Set

How to build data sets directly affects the accuracy of emotion recognition results. In recent years, multimodal data sets have taken into account different signals, which has gradually improved the accuracy, but there are still many problems to consider.

The first is that large datasets use simulated emotions rather than naturally occurring ones, and the differences need to be considered. Secondly, how to show the emotional background is a problem. In addition, the expression of emotions is different from culture to culture and from individual to individual, and this difference needs to be taken into account.

4.3. Practical Application

Since emotional computing can have many benefits for people, it is morally understandable to try to use emotional computing. However, there are many issues to consider when applying emotional computing to reality.

The first is how to strike a balance between giving positive emotions and reality, that is, to give users a good emotional experience, but also to ensure that the user can realize the reality. For systems that give feedback directly to the user, for example, some teaching systems, failing to maintain this balance can hinder the achievement of the application's goals or give people unrealistic expectations about the emotional computing system or the real environment. The purpose of some applications is to provide emotional support. However, such positive emotions may not only be mismatched with negative reality, leading to isolation from reality but also may cause people to become over-dependent and give up personal development because of the temporary happiness they bring. It is also worth debating who bears the blame when these situations arise [32].

Systems that use emotional computing to give information are often used in fields such as medical monitoring. In these fields, it is a question of how to maintain equality and help those who don't have a happy life to reach the level of the majority, rather than just helping people having a good quality of life to live better.

Some apps help to interfere with users' emotions. They are also often used in the fields of therapy and education. The biggest problem in this area is that users' collaboration with these applications is based on trust. In situations where they are not subject to human moral control and do not fully conform to humans, these applications gain trust by imitating humans. This kind of trust is built on a fake reality, which is arguably, ethically troubling.

In addition to problems related to methods of use, there are also general ethical issues. One is the risk of technology being manipulated by power. Excessive collection of other people's emotional information will form emotional surveillance and aggravate the disrespect for ordinary people. Emotions, as part of irrationality, can even play the role of manipulating the opinions of the masses when guided by control. This power creates differences not only between those in power and the masses but also between mainstream groups and minorities [32]. Systems generated by large amounts of data can be overly biased towards mainstream expressions of emotion and unable to cope with other forms of expression.

5. Conclusion

With the attention of scholars and the development of science and technology, affective computing has become a new research field. Affective computing is a very complex research field, involving cognitive science, computer science, and other fields of knowledge. This paper provides an overview of the development of existing techniques for affective computing, including the acquisition of affective features, affective recognition, and affective generation. The application of affective computing in the fields of education, games, and medicine is summarized. It also includes the ethical

issues that emotional computing can raise. The article has several limitations. First, because of the complexity of the field, problems may arise that do not allow for a comprehensive description of the current state of development. Second, due to the selection of recent papers, the conclusions may be too biased or absolute. Third, this article only summarizes the research status, and may not touch on the more essential thinking of each part. Affective computing should pay more attention to the simulation of the human cognitive process in the construction of an emotion model. In terms of technology, with the rapid development of algorithms and devices, we can expect richer and more accurate data sets and suitable algorithms to emerge. In terms of application, there are a large number of systems and ideas that are in their infancy and are expected to be realized in the future, but at the same time, it is important to consider ethical issues. From this point of view, there are still many problems and directions in this field that need to be solved and realized. This article is intended as a starting point for those who wish to enter this area.

References

- [1] Ekman, P. (1992). *An argument for basic emotions*. *Cognition & Emotion*, 6(3-4), 169-200.
- [2] PS, S., & Mahalakshmi, G. (2017). *Emotion models: A review*. *International Journal of Control Theory and Applications*, 10(8), 651-657.
- [3] Xiong, Y. (2024). *A review of dimensional emotion models*. *Advances in Psychology*, 14, 270.
- [4] Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., et al. (2001). *Emotion recognition in human-computer interaction*. *IEEE Signal Processing Magazine*, 18(1), 32-80.
- [5] Dzedzickis, A., Kaklauskas, A., & Bucinskas, V. (2020). *Human emotion recognition: Review of sensors and methods*. *Sensors*, 20(3), 592.
- [6] Languré, A. D. L., & Zareei, M. (2024). *Improving text emotion detection through comprehensive dataset quality analysis*. *IEEE Access*.
- [7] Rezaee, K. (2024). *An evolutionary convolutional neural network architecture for recognizing emotions from EEG signals*. In *Recent Advances in Machine Learning Techniques and Sensor Applications for Human Emotion, Activity Recognition and Support* (pp. 103-138). Springer Nature Switzerland.
- [8] Kirbiz, S. (2024). *Facial emotion recognition using residual neural networks*. *Electrica*, 24(3).
- [9] Ba, S., & Hu, X. (2023). *Measuring emotions in education using wearable devices: A systematic review*. *Computers & Education*, 200, 104797.
- [10] Krumhuber, E. G., Skora, L. I., Hill, H. C. H., et al. (2023). *The role of facial movements in emotion recognition*. *Nature Reviews Psychology*, 2(5), 283-296.
- [11] Koolagudi, S. G., & Rao, K. S. (2012). *Emotion recognition from speech: A review*. *International Journal of Speech Technology*, 15, 99-117.
- [12] Al-Dujaili, M. J., & Ebrahimi-Moghadam, A. (2023). *Speech emotion recognition: A comprehensive survey*. *Wireless Personal Communications*, 129(4), 2525-2561.
- [13] Pan, L., & Wang, Q. (2024). *GFRN-SEA: Global-aware feature representation network for speech emotion analysis*. *IEEE Access*.
- [14] Saxena, R. R. (2024). *Applications of natural language processing in the domain of mental health*. *Authorea Preprints*.
- [15] Kyung, J., Heo, S., & Chang, J. H. (2024). *Enhancing multimodal emotion recognition through ASR error compensation and LLM fine-tuning*. In *Proceedings of Interspeech 2024* (pp. 4683-4687).
- [16] Kaneko, T. (2024). *Enhancing emotion recognition in spoken dialogue systems through multimodal integration and personalization*. In *Proceedings of the 20th Workshop of Young Researchers' Roundtable on Spoken Dialogue Systems* (pp. 5-7).
- [17] Yang, D., Chen, Z., Wang, Y., et al. (2023). *Context de-confounded emotion recognition*. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 19005-19015).
- [18] Nagata, M., & Okajima, K. (2024). *Effect of observer's cultural background and masking condition of target face on facial expression recognition for machine-learning dataset*. *PloS One*, 19(10), e0313029.
- [19] Li, Y., Wang, Y., & Cui, Z. (2023). *Decoupled multimodal distilling for emotion recognition*. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 6631-6640).
- [20] Abdumalikov, S., Kim, J., & Yoon, Y. (2024). *Performance analysis and improvement of machine learning with various feature selection methods for EEG-based emotion classification*. *Applied Sciences*, 14(22), 10511.
- [21] Ahmed, N., Al Aghbari, Z., & Girija, S. (2023). *A systematic survey on multimodal emotion recognition using learning algorithms*. *Intelligent Systems with Applications*, 17, 200171.

- [22] Hu, D. (2024). DragGAN-based emotion image generation and analysis for animated faces.
- [23] Tasoulas, T., Troussas, C., Mylonas, P., et al. (2024). Affective computing in intelligent tutoring systems: Exploring insights and innovations. In *Proceedings of the 9th South-East Europe Design Automation, Computer Engineering, Computer Networks and Social Media Conference (SEEDA-CECNSM)* (pp. 91-97). IEEE.
- [24] Sharma, D., & Chawla, S. (2024). Emotion recognition AI in online learning: Enhancing engagement and personalizing educational experiences.
- [25] Schiavo, F., Campitiello, L., Todino, M. D., et al. (2024). Educational robots, emotion recognition, and ASD: New horizon in special education. *Education Sciences*, 14(3), 258.
- [26] Dehghani, F., & Zaman, L. (2023). Facial emotion recognition in VR games. In *2023 IEEE Conference on Games (CoG)* (pp. 1-4). IEEE.
- [27] Bellenger, D., Chen, M., & Xu, Z. (2024). Facial emotion recognition with a reduced feature set for video game and metaverse avatars. *Computer Animation and Virtual Worlds*, 35(2), e2230.
- [28] Anto-Chavez, C., Maguiña-Bernuy, R., & Ugarte, W. (2024). Real-time CNN-based facial emotion recognition model for a mobile serious game. In *ICT4AWE* (pp. 84-92).
- [29] Islam, M. M., Nooruddin, S., Karray, F., et al. (2024). Enhanced multimodal emotion recognition in healthcare analytics: A deep learning-based model-level fusion approach. *Biomedical Signal Processing and Control*, 94, 106241.
- [30] Kumar, D., & Narzary, D. G. (2024). Exploring the utility of emotion recognition systems in healthcare. In *Using Machine Learning to Detect Emotions and Predict Human Psychology* (pp. 245-271). IGI Global.
- [31] Ortiz-Clavijo, L. F., Gallego-Duque, C. J., David-Diaz, J. C., et al. (2023). Implications of emotion recognition technologies: Balancing privacy and public safety. *IEEE Technology and Society Magazine*, 42(3), 69-75.
- [32] Devillers, L., & Cowie, R. (2023). Ethical considerations on affective computing: An overview. *Proceedings of the IEEE*.