A Learning Analytics Model Based on Expression Recognition and Affective Computing: Review of Techniques and Survey of Acceptance

Chengliang Wang, Jian Dai, Yu Chen, Xing Zhang, and Liujie Xu(✉)

College of Educational Science and Technology, Zhejiang University of Technology, Liuxia Street, Hangzhou, China
xulj2004@126.com

Abstract. The development of technology informatization and intelligence makes personalized adaptive learning possible, especially the rapid development of artificial intelligence provides practical technical support for intelligent teaching. This paper reviews the current research status of expression recognition and affective computing in the education field, and analyzes the technical basis required to realize affective computing in the online learning environment from facial expression recognition, image pre-processing and feature extraction, and explores the technical feasibility of the learning analysis model based on affective computing is discussed. Through a questionnaire survey, we investigated learners' acceptance of expression recognition and affective computing applied to the field of education and teaching, and found that most learners were willing to use AI-related technologies to collect information when learning in order to improve the efficiency of learning, and some learners were even willing to share this information with teachers.

Keywords: Personalized Learning · Expression Recognition · Affective Computing · Deep Learning · Teaching Evaluation

1 Introduction

In the past, due to the limitation of technology, education evaluation mainly relied on the summative evaluation based on examinations, which has a large limitation and cannot measure the learning effectiveness of learners comprehensively. With the development of artificial intelligence and other technologies, personalized formative assessment has gradually become possible. At the same time, the development of online learning platforms has provided great convenience for the collection of formative data.

In recent years, intelligent education has become a research boom in the field of pedagogy and artificial intelligence (AI); with the development of humanistic theory in pedagogy and the advancement of emotion computing technology in the field of computers, it has become possible to use computers to extract emotions from facial

C. F. Peng et al. (Eds.): EIMT 2022, AHSSEH 3, pp. 169–178, 2023.
https://doi.org/10.2991/978-94-6463-012-1_19

expressions in the field of education. Undoubtedly, such personalized information of learners obtained based on facial expression recognition and emotion computing can help learners understand their learning status, assist teachers to better carry out effective teaching, and achieve personalized teaching methods and deployment of teaching resources, but it also brings privacy and acceptance problems. The aim of this paper includes the following two aspects:

(1) To review key technologies in the field of affective computing and analyze the technical feasibility of their application in education.

(2) To investigate learners' acceptance of affective computing technologies applied in the field of education.

2 Literature Review

2.1 Expression Recognition and Affective Computing

In the learning process, emotion is a very important factor that can largely affect the learning state of the learner, the ability to think and analyze while learning, the ability to make decisions and the feeling of well-being while learning [14]. So there is a desire to have a technological means to understand the emotional state of humans while learning and to record it analytically. The development of artificial intelligence, especially the wide application of deep learning techniques in the field of image recognition, has led to excellent processing tools and solution paradigms for facial expression recognition and affective computing.

The steps of affective computing for a facial expression can be divided into four steps in general, including feature extraction, calculation of feature values, feature selection and emotion recognition, where feature selection and emotion recognition depend on the chosen method strategy and are highly flexible.

In terms of feature selection, in the context of online education, learners' emotions are often conveyed through facial expressions [1]. In the 19th century, the most basic emotions were classified into six categories, namely happiness, sadness, fear, surprise, disgust, and anger, each of which has a matching facial expression feature. In the process of computer recognition and processing of facial expressions, two more major methods of facial expression feature extraction have been developed, namely static picture feature extraction and dynamic sequence feature extraction. For example, Cohen et al. have used principal component analysis to extract static features of face pictures [4], while Zeng et al. have investigated the use of combined units of facial actions as feature values of the emotions expressed by faces [16].

With the continuous progress of machine learning, the original paradigm of expression recognition is constantly updated, and the expression algorithm based on deep learning is to learn by classifying images and eventually train to get a model with strong generalization ability. Nowadays, the common methods are mainly neural network methods and support vector machine methods, which have replaced the original traditional face recognition methods based on principal component analysis, linear discriminant analysis, etc.

Innovations in facial expression recognition techniques, especially deep learning, have been widely used in the field of expression recognition in recent years, resulting in great improvements in both the techniques of recognition and the accuracy of classification. For example, to address the problems of small number of existing expression datasets, insufficient comprehensiveness of information, and incomplete classification, Lopes et al. have performed complex processing of training samples, and then used transfer learning and multi-view neural networks to train the model, achieving the goal of still achieving better recognition results on small samples [12]. Moreover, Choi et al. have started by improving the convolutional layer in the model and greatly improved the efficiency of expression recognition by adjusting the number of nodes and reducing the parameters of training [3].

2.2 Affective Computing in Online Education

Since 2012, with the continuous development of online education companies and educational institutions, as well as the deepening research on online education modes such as MOOCs, the research on online education has also set off a new climax. However, there is a huge drawback of online education, that is, teachers cannot pay attention to learners' status in real time during their learning process, and learners are more likely to be lax and inattentive due to the lack of supervision, thus easily causing a disconnect between teaching and learning.

Emotion is an important factor that affects learners' learning and can indirectly influence learning outcomes by affecting their intrinsic motivation to learn and the implementation of their learning plans [15]. Based on the special education environment of online education, learners undoubtedly need to pay more attention to their emotional state to avoid affecting their learning status and cognitive ability by appearing disgusted, bored, or frustrated when they encounter frustration.

Regarding the role of affective computing in the teaching field, Chen has added a series of interactive attributes to instructional videos while tracking learners' facial expressions in real time, and analyzed the collected information to obtain learners' measurement indicators when learning and give matching learning instructions [2]. Drawing on its research ideas, the application of AI as a teaching aid in the field of education, especially in the emerging field of online education, is bound to play a role in replacing teachers for learning assistance to a certain extent. Although it needs to be acknowledged that AI may not be able to match the delicacy and relevance of teachers from the perspective of human education, it has its own unique advantages, such as more comprehensive, full-time data collection and real-time processing, one-to-one personalized guidance and service, which are critical factors that are superior to traditional teaching.

In the era of education informatization, emotional interaction is a necessary path for the development of this field of computer-assisted teaching, and the ability of computers to understand human emotions is the basis for their ability to serve humans better. Education needs emotion and sentiment, a cold machine that cannot recognize emotion is always outside the essence of education. "Teaching" and "education" are two complementary parts of education, and the research and development of educational technology cannot be separated from either one of them. One of the core concepts in the development of technology for education service is to make technology humanized, especially in the distance online learning environment which lacks emotional interaction, it is more necessary to pay attention to the change of learners' emotional state, give learners real-time feedback and stage evaluation, and personalize learning strategies and personalized guidance on the use of resources based on real-time feedback and evaluation reports to meet learners' personalized learning needs. We can build a "learner-centered" teaching service model, achieve a balance between "teaching" and "education", and meet learners' emotional needs with high quality and efficiency. We will achieve our teaching goals with high quality and efficiency while meeting the emotional needs of learners.

3 Technological Bases

3.1 Technological Bases for Face Detection

Face detection refers to recognizing and locating the location and size of a face in a given photo and picture, and belongs to a branch in the field of computer picture recognition. In recent years, due to the rapid development of e-commerce, face detection technology has increasingly become a research hotspot. At present, face detection technology has become mature at the application level, especially the emergence of a series of face detection methods based on deep learning, which to a certain extent have replaced the originally used Edge detection methods based on feature analysis and detection methods based on template matching.

Face detection methods based on deep learning mainly use neural networks to deeply learn the features of a large number of face images and train with gradient descent and back-propagation algorithms to generate a binary classifier that can determine the presence of a face in an image or video, which is currently more widely used as cascaded convolutional neural networks. Li et al. have performed face detection with several cascaded convolutional neural networks and found that cascaded neural networks are more efficient and high quality compared to a single convolutional neural network [10]. Moreover, a multi-task convolutional neural network model specifically for face detection has been proposed by the Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences [8].

Cascaded convolutional neural networks have become the most popular method in the field of face detection because of their high efficiency and accuracy, and the existing research results are sufficient to support the face detection needs required by online learning environments.

3.2 Technological Bases for Facial Expression Recognition

Recognition of facial expressions is mainly achieved by extracting and analyzing facial features and motion changes of human faces. However, due to the changes in the environment of graphic information collection, there will be problems such as insignificant face features. Therefore, some pre-processing of image data is needed to eliminate the interference caused by the change of environment light and dark, shooting angle, different background, etc.

3.2.1 Image Preprocessing of Facial Expressions

There are many distinct features on human faces, such as color features, histogram features, contour features, transform domain features, etc. In order to better extract feature information of faces, a series of pre-processing is needed to standardize the features and thus reduce the difficulty of extracting them. The main techniques used include face alignment and image enhancement.

Face alignment is a technique used to locate the position of a face in a picture and to perform alignment operations, with the aim of avoiding adverse effects on recognition results due to uncentered faces. The principle of the technique is mainly through the positioning and alignment of facial features and other prominent features.

Image enhancement techniques are mainly used to deal with unfavorable factors caused by the background of the image and the shooting environment, such as problems like overly bright, dark and reflective pictures caused by lighting factors, blurred pictures and saturation differences. All these factors can lead to degradation of picture quality and affect the accuracy of the face recognition algorithm. Facing the problem of similar brightness of background and foreground, the contrast of the image can generally be enhanced by combining histogram equalization and illumination normalization to make the image subject and background distinguishable; for low-resolution pictures, active noise cancellation can generally be achieved by removing motion blur and other methods.

3.2.2 Feature Extraction of Facial Expressions

The extraction and classification of facial expression features is an important part of facial expression recognition, compared to traditional face expression recognition methods which also need to rely on classifiers to get the results of expression classification. In recent years, various techniques based on deep learning have provided more intelligent methods for facial expression recognition, especially the development of convolutional neural networks, which have shown higher accuracy and stronger robustness in the field of face recognition.

Convolutional neural network is a special kind of artificial neural network for image classification problems with large amount of data. Compared with general artificial neural networks, the hidden layers of convolutional neural networks are replaced by a cross-stacked structure of convolutional, pooling and fully connected layers, which give the convolutional neural network image translation, scaling and rotation invariance.

In emotion recognition, picture data is used as input data for convolutional neural networks, often in the form of a pixel matrix. In the whole network, the structure formed by the combined stack of convolutional and pooling layers is mainly responsible for the extraction of the picture information, and the fully connected layers mainly classify the collected information.

A typical neural network often contains five parts: an input layer, a convolutional layer, a pooling layer, a fully connected layer and an output layer. After the mapping operations in the convolution, pooling and fully connected layers, the final result of the desired classification is obtained. The main function of the convolution layer is feature extraction. Convolution, which is the linear transformation of each position of the image, eventually extracts local features, and the training process is to fit the data by solving the weight method of the local mapping. The function of the pooling layer is mainly to reduce the dimensionality of the output of the features from the convolutional layer. The main purpose of this process is to reduce the number of parameters of the network, reduce the overfitting phenomenon of the model, and thus enhance the fault tolerance of the model. The pooling layer is very important because the process of continuous convolution will make the feature dimensionality grow rapidly, while generating a large amount of redundant information, which will undoubtedly affect the efficiency of model training. Common pooling operations include mean pooling and maximum pooling. Different pooling operations often have different purposes, and the appropriate pooling type is usually chosen according to the research needs, such as using mean pooling to get smoother features, but using maximum pooling to get more significant features of the image. The selection of features will often affect the degree of generalization of the model, for example, maximum pooling is more likely to lead to overfitting of the model. The role of the fully-connected layer is mainly to perform mapping relations and dimensional transformations of features. In addition, its final layer often needs to act as a "classifier" to present the final classification results.

Since the shooting angle is very fixed in the online learning environment, an OpenCV-based face detection method can also be considered. The advantage of this method is that it is simple in its construction and can detect face sizes of different proportions, especially under relatively stable shooting conditions and in real time when the face target is more visible as a larger part of the picture. However, using this method makes multiple assumptions about the face, sometimes it is not possible to locate the face accurately, and it is not possible to identify non-frontal faces, which is prone to false detection, and in summary, it can be used as an alternate method for face detection.

3.2.3 Public Expression Datasets

In model training, a large number of expression datasets are often needed, and manual acquisition of face photos is time-consuming and labor-intensive, while each photo needs to be labeled with an emotion tag, which is costly. Therefore, it is more appropriate to integrate the existing open-source expression datasets to form a set of datasets that meet the research needs for experimentation. The current open-source expression datasets are listed in Table 1.

Observing the datasets of common expressions, we can see that the datasets in the laboratory environment are generally smaller in number, but the pictures are neat and can be used with only simple adjustments in the pre-processing procedure; the data collected in the Internet environment are generally larger in number, but the disadvantage is that some of them are not set with labels, which is not suitable for the need of supervised learning in this experiment, while the pictures have different degrees of lightness and darkness and saturation of the faces, which require a lot of pre-processing of the data.

Table 1. Common expression datasets and their basic properties

Dataset Name	Recording Environment	Amount of Data	Emotion Classification
FER2013 [7]	Internet	35,886 static images	Angry, disgusted, afraid, happy, sad, surprised, neutral
JAFFE [13]	Laboratory	213 static images	Anger, disgusted, afraid, happy, sad, frightened, neutral
AFEW 7.0 [5]	Film Clip	1,809 videos	Angry, disgusted, afraid, happy, afraid, frightened, neutral
RaFD [9]	Laboratory	1,608 images	Angry, contemptuous, disgusted, afraid, happy, sad, surprised, neutral
Qulu-CASIA [17]	Laboratory	2,880 image sequences	Angry, disgusted, afraid, happy, sad, surprised
ExpW [11]	Internet	91,793 static images	Angry, disgusted, afraid, happy, sad, surprised, neutral
EmotioNet [6]	Internet	1 million static images	No label

4 Questionnaire Survey

The application of affective computing in education often requires access to information related to learners' learning, which involves issues of learner acceptance and privacy, and therefore requires relevant social research before being promoted.

The questionnaire was designed with 12 questions, including 3 questions to investigate the basic information of the respondents, 3 questions to investigate the respondents' understanding of AI, 5 questions to investigate the respondents' acceptance of AI in the field of education, and 1 test question (anti-forgery question) to screen out invalid questionnaires. The scores were used to quantify the two dimensions of respondents' understanding of AI and respondents' acceptance of AI in education, which facilitated the subsequent data processing. The questionnaires were mainly targeted at college students, and 82 questionnaires were collected through the WJX platform. After screening, excluding the questionnaires with too short a filling time and wrong answers to the anti-forgery questions, the final number of valid questionnaires was 78, with a valid questionnaire recovery rate of 92.3%.

The questionnaire data were analyzed using SPSS 26.0 software. Monofactor analysis was used to test whether there was a significant effect of AI understanding on AI acceptance in education and profession on AI acceptance in education, and the results

Table 2. Significance test of understanding of ai and profession on ai acceptance in education

Independent Variable	N	Mean	df	F	P
Understanding of AI	78	33.83	13	0.536	0.892
Profession	78	33.83	2	0.763	0.470

of the analysis are shown in Table 2. As can be seen from the data in the table, the significance coefficients of both are much greater than 0.05, indicating that neither the learners' understanding of AI nor the learners' profession has a significant effect on the acceptance of AI in the field of education. This tendency is also felt in real life, especially among those who know less about AI, and tends to go to two different extremes. Those who advocate for technology to change lives will welcome AI in education to assist in teaching, while those who oppose technology will feel that cold technology does more harm than good to education. There is still a tendency among those who understand AI, but the situation has improved more significantly than among those who do not understand AI.

Next, an independent samples t-test was used to explore whether gender has a significant effect on the acceptance of AI in the field of education. The analysis revealed that the results of Levene's test showed homogeneity of variance ($F = 0.849$, $p = 0.360$), and the value of the significance coefficient (Sig.) is 0.246, which is still greater than 0.05, indicating that gender does not have a significant effect on the acceptance of artificial intelligence in the field of education.

By further analyzing the internal questions of the questionnaire, in the dimension of acceptance of AI, a question of the questionnaire asked "Do you agree with the use of AI technology for collecting information from learners in online learning", and it was found that 80.77% of the students were willing to use AI-related technology to collect information for learning to assist their learning activities. Among them, 42.31% were willing to share the data with their teachers.

The above data indicate that gender, major, and understanding of AI do not significantly affect the acceptance of AI. Careful analysis revealed that most college learners are willing to use AI technologies to assist learning activities, and there exist some learners who are willing to share their data with teachers so that teachers can be more deeply involved in the teaching and learning aspects of online learning. The above social research provides a realistic basis for the promotion of AI-related technologies among the college students.

5 Conclusions

The application of affective computing in education often requires access to information related to learners' learning, which involves issues of learner acceptance and privacy, and therefore requires relevant social research before being promoted.

Undoubtedly, from the review of the existing technologies in this field, the research on facial expression recognition has been fruitful, and it has been more mature in terms of localization, image preprocessing and feature extraction, and a large number of publicly available expression datasets have been constructed. However, there are still many technical points that can be improved in the field of details, such as overfitting easily due to the significant feature anisotropy, which leads to the generalization ability that cannot meet the demand of practical use. At the same time, there is still much room for innovation in research on affective computing because of its high inherent complexity and the need to consider the hidden emotions in a particular state.

As for the question of the acceptance of affective computing in the field of education, the answer was given by a questionnaire survey. The data analysis showed that learners' understanding of AI, profession, and gender did not have a significant impact on the acceptance of AI in education, and most learners were willing to use AI-related technologies to collect information to assist their learning activities, and more than 40% of them were willing to share their data with teachers.

Acknowledgements. This research was supported by the National University Student Innovation and Entrepreneurship Training Project of China under Grant 201910337031, 202210337027, 202210337003.

References

1. Cacioppo, J.T., and L.G. Tassinary. 1990. Inferring psychological significance from physiological signals. *American Psychologist* 45: 16.
2. Chen, H.R. 2012. Assessment of learners' attention to e-learning by monitoring facial expressions for computer network courses. *Journal of Educational Computing Research* 47: 371–385.
3. Choi, I.K., H.E. Ahn, and J. Yoo. 2018. Facial expression classification using deep convolutional neural network. *Journal of Electrical Engineering & Technology* 13: 485–492.
4. Cohen, I., N. Sebe, A. Garg, L.S. Chen, and T.S. Huang. 2003. Facial expression recognition from video sequences: Temporal and static modeling. *Computer Vision and Image Understanding* 91: 160–187.
5. Dhall, A., R. Goecke, S. Ghosh, J. Joshi, J. Hoey, and T. Gedeon. 2017. From individual to group-level emotion recognition: Emotiw 5.0. In *Proceedings of the 19th ACM international conference on multimodal interaction.* ACM.
6. Fabian Benitez-Quiroz, C., R. Srinivasan, A.M. Martinez. 2016. EmotioNet: an accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition.* IEEE.
7. Goodfellow, I.J., D. Erhan, P.L. Carrier, A. Courville, M. Mirza, B. Hamner, and Y. Bengio. 2013. Challenges in representation learning: A report on three machine learning contests. In *International conference on neural information processing*. Cham: Springer. https://doi.org/10.1007/978-3-642-42051-1_16
8. Kumar, A., R. Chellappa. 2018. Disentangling 3D pose in a dendritic cnn for unconstrained 2d face alignment. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.* IEEE.
9. Langner, O., R. Dotsch, G. Bijlstra, D.H. Wigboldus, S.T. Hawk, and A.D. Van Knippenberg. 2010. Presentation and validation of the Radboud faces database. *Cognition Emotion* 24: 1377–1388.

10. Li, H., Z. Lin, X. Shen, J. Brandt, and G. Hua. 2015. A convolutional neural network cascade for face detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. IEEE.
11. Li, S., W. Deng. 2020. Deep facial expression recognition: A survey. *IEEE Transactions on Affective Computing*, 1.
12. Lopes, A.T., E. De Aguiar, and T. Oliveira-Santos. 2015. A facial expression recognition system using convolutional networks. In *2015 28th SIBGRAPI conference on graphics, patterns and images*. IEEE.
13. Lyons, M., S. Akamatsu, M. Kamachi, J. Gyoba. 1998, April. Coding facial expressions with Gabor wavelets. In *Proceedings third IEEE international conference on automatic face and gesture recognition*. IEEE.
14. Morrish, L., N. Rickard, T.C. Chin, and D.A. Vella-Brodrick. 2018. Emotion regulation in adolescent well-being and positive education. *Journal of Happiness Studies* 19: 1543–1564.
15. Sandanayake, T.C., A.P. Madurapperuma, and D. Dias. 2011. Affective E learning model for recognising learner emotions. *International Journal of Information Education and Technology* 1: 315.
16. Zeng, Z., M. Pantic, G.I. Roisman, and T.S. Huang. 2008. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis* 31: 39–58.
17. Zhao, G., X. Huang, M. Taini, S.Z. Li, and M. PietikäInen. 2011. Facial expression recognition from near-infrared videos. *Image and Vision Computing* 29: 607–619.