JKIICE Journal of the Korea Institute of Information and Communication Engineering

한국정보통신학회논문지 Vol. 26, No. 3: 333~340, Mar. 2022

딥러닝의 얼굴 정서 식별 기술 활용-대학생의 심리 건강을 중심으로

리파[1] · 조경덕[2*]

Exploration of deep learning facial motions recognition technology in college students' mental health

Bo-Li[1] · Kyung-Duk Cho[2*]

[1]Doctoral student, Department of Psychological Counseling, Paichai University, Daejeon, 35345 Korea
[2*]Professor, Department of Psychology & Counselling, Paichai University, Daejeon, 35345 Korea

요 약

코로나19는 모두로 하여금 초조하고 불안하게 만들고, 사람들간에는 거리두기가 필요하다. 코로나19로 인해 심리적으로 초조하고 불안 해 지고 거리두기가 필요해졌다. 대학교에서는 학기 초에 정신건강에 대한 단체 평가와 검사가 이루어진다. 본 연구에서는 다층감지기 신경망 모델을 채택하고 훈련시켜 딥러닝을 진행했다. 훈련이 끝난 후, 실제 사진과 동영상을 입력하고, 안면탐지를 진행하고, 표본에 있는 사람의 얼굴 위치를 알아낸 후, 그 감정을 다시 분류하고, 그 표본의 예측한 감정 결과를 그림으로 보여주었다. 결과는 다음과 같다. 테스트 시험에서는 93.2%의 정확도를 얻었고, 실제 사용에서는 95.57%의 정확도를 얻었다. 그중 분노의 식별율은 95%, 혐오의 식별율은 97%, 행복의 식별율은 96%, 공포의 식별율은 96%, 슬픔의 식별율은 97%, 놀라움의 식별율은 95%, 중립의 식별율은 93%이었다. 본 연구의 고효율적 정서 식별 기술은 학생들의 부정적 정서를 포착하는 객관적 데이터를 제공 할 수 있다. 딥러닝의 감정식별 시스템은 심리건강을 향상하기 위한 데이터들을 제공할 수 있다.

ABSTRACT

The COVID-19 has made everyone anxious and people need to keep their distance. It is necessary to conduct collective assessment and screening of college students' mental health in the opening season of every year. This study uses and trains a multi-layer perceptron neural network model for deep learning to identify facial emotions. After the training, real pictures and videos were input for face detection. After detecting the positions of faces in the samples, emotions were classified, and the predicted emotional results of the samples were sent back and displayed on the pictures. The results show that the accuracy is 93.2% in the test set and 95.57% in practice. The recognition rate of Anger is 95%, Disgust is 97%, Happiness is 96%, Fear is 96%, Sadness is 97%, Surprise is 95%, Neutral is 93%, such efficient emotion recognition can provide objective data support for capturing negative. Deep learning emotion recognition system can cooperate with traditional psychological activities to provide more dimensions of psychological indicators for health.

키워드 : 딥러닝, 감정 인식, 대학생의 심리 건강

Keywords : Deep learning, Emotion recognition, Students' mental health

Received 20 October 2021, **Revised** 24 December 2021, **Accepted** 2 February 2022

*** Corresponding Author** Kyung-Duk Cho(E-mail: chokd@pcu.ac.kr Tel:+82-42-520-5880)
Professor, Department of Psychology & Counselling, Paichai University, Daejeon, 35345 Korea

Open Access http://doi.org/10.6109/jkiice.2022.26.3.333 print ISSN: 2234-4772 online ISSN: 2288-4165

Ⅰ. The Introduction

People will intentionally or unintentionally reveal their inner feelings and thoughts through facial expression, voice, language, body movements and so on, so as to establish a good and friendly relationship with each other. Among so many expressions, facial expression is the most intuitive and natural signal when we express emotions. We can express ourselves and recognize the emotions of others through expressions. Through understanding people's emotional information, it can help People's Daily life more delicately and bring people a lot of convenience. If such facial emotion recognition technology is used in the education and evaluation of mental health in colleges and universities, it can provide more dimensions to provide indicators of mental health by constructing the deep learning system of computer automatic emotion recognition with traditional psychological evaluation, interview and classroom activities. People are often affected by emotions, so are psychological teachers or counselors. The emotion recognition system of deep learning can track the emotional state of consultants in real time and give appropriate supporting suggestions. The combination of human and intelligent system will be more objective and comprehensive, better protection for patients and consultants, and work will be more smooth.

Ⅱ. Research concepts

2.1. Classification of emotions

In the 20th century, Ekmanetal defined six basic emotions on the basis of cross-cultural studies, namely Anger, Disgust, Fear, Happiness, Sadness and Surprise. This study shows that human beings express certain basic emotions in the same way regardless of cultural differences, indicating that these emotions are universal and cross-cultural. In addition to the six types of facial expressions, there are also periods in real life when the mood is relatively calm and with little fluctuation. Therefore, Neutral facial expressions are included as one of the emotions to be identified in this study. Therefore, seven types of facial expressions are identified in total.

2.2. Deep learning

With the rapid improvement of chip processing capacity and the continuous emergence of various neural network structures, many research and applications in different fields begin to use deep learning methods, and deep learning has indeed greatly improved the accuracy of identification. Deep learning networks include multilayer neural network, recursive neural network, long and short duration memory network, convolutional neural network, etc. You can use different network structures based on different application scenarios. Deep learning is a new area in machine learning research, motivated by building, simulating neural networks of the human brain for analytical learning, which interpret data by mimicking human brain mechanisms, such as images, sound, and text.Deep learning forms more abstract high-level features by combining low-level features to discover the distribution laws of the data[1].

Ⅲ. Application of deep learning in life scenarios

Table. 1 Deep learning and Life scenarios

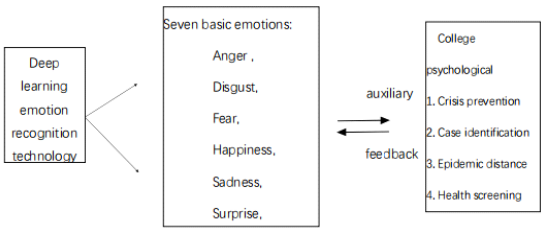
Life scene application	Technology	Content
Network media (Ren, F., Quan, C. 2012)	Customer situation Customer preference	An experience mechanism to improve the rate of emotional recognition by understanding and analyzing the user's internal tendencies
Autism - treatment (Rashidan.2021)	Emotion recognition Face recognition	Studying the emotional expression characteristics of autistic children through emotion recognition technology is more beneficial to the social rehabilitation process of autistic children.
Social cognition in the elderly (2021) Ferreira.	Facial emotion recognition	Deep learning facial emotion recognition is used to assess the changing social cognition of the elderly.

Life scene application	Technology	Content
Emotional monitoring (Khanna, 2011)	Speech recognition Emotion recognition	Through the multi-mode emotion recognition system can better monitor people's emotions, so as to prevent crisis. For example, students' crisis emotion monitoring, patients' negative emotion monitoring, drivers' fatigue emotion monitoring, movie and theater audience emotion monitoring.
Robot doll for children (H.G. Kim. 2021).	Multimodal emo -tion recognition program Emotion recognition Speech recognition	Through the multi-mode emotion recognition program, the dialogue scenes between children and smart toys can be deeply learned and analyzed, and through facial expressions and voice confirmation, children can form social and self-construction through the interaction with toys, and constantly strengthen two-way emotional communication.
Class monitor The attendance management (Harikrishna n et al. 2019)	Face recognition Emotion recognition Speech recognition	By using deep learning, teachers and students' absence management and classroom status can be supervised and managed in real time, and the results can be sent back to the back-end for recording.

As can be seen from the literature in Table -1 above, the life scenes that deep learning technology can be applied to have become increasingly rich. In the past, due to the variation of visual Angle and the monotonous identification network, the detection efficiency was low and the accuracy could not be improved. In addition, the data set was mainly in the laboratory environment, and external factors such as light and shadow would also affect the identification efficiency[2]. The rapid development of neural networks, coupled with a large number of marked data sets of different poses, expressions, and light, can train a detection model for real scenes, which greatly improves the efficiency and accuracy of face recognition.

Table. 2 The application of deep learning in the field of mental health in colleges and universities is supposed to apply the emotion recognition technology of deep learning to mental health activities in colleges and universities on the basis of previous studies. Because the

Table. 2 Application of emotion recognition technology in the field of mental health in colleges and universities



traditional mental health assessment is based on the measurement scale as the only means[3], if the identification data of positive and negative emotions of the subjects are added at the same time, it can help the healthy growth of college students in multiple dimensions and improve the probability of preventing psychological crisis. Due to the long-term existence of COVID-19, people should keep a healthy distance from each other. It would be safer and more convenient to have a dimension that is convenient for measuring emotions. Therefore, this study attempts to explore the application of deep learning emotion recognition technology to help improve the field of mental health assessment in colleges and universities, so as to achieve a more comprehensive capture of negative emotions of college students, hoping to help people in need of help in a more timely manner.

Ⅳ. The research process

4.1. Establishment of the study model

The multi-layer perceptron model is used to carry out deep learning, which firstly propagates forward through the perceptron and then reversely transmits through BP error[4], and adjusts the corresponding connection weights according to the error. Since the training learning sample size is not large, in order to prevent under-fitting, the number of perceptron layers and the number of neurons in each layer are increased. Five perceptron layers are adopted, among which the second and third layers are 257 neurons. Although the network

is more complex, it can achieve better fitting and generalization ability. Generally, activation functions in neural networks include Sigmoid, hyperbolic tangent and improved linear element[5]. In this model, ReLU activation function is used for the hidden layer and Sigmoid activation function is used for the output layer.

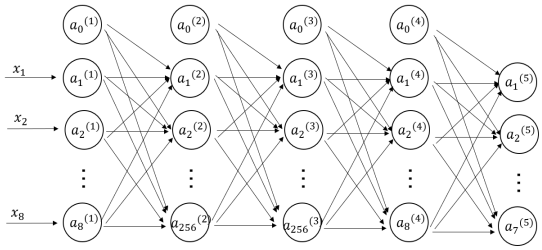


Fig. 1 Multi-layer perceptron neural network model

The x1-x8 is the input variable, the first 8 columns of Table 3 during training, and the 8 values generated by the real-time extracted face coordinates. The superscript number of a neurons represents the network layers, the subscript number represents the few neurons in the current layer, and the connection between neurons is the connection weights.

The first layer is to directly propagate the sample parameters of the training set, as shown in Formula (1) :

$$A_1^{(i)} = x_i \tag{1}$$

The bias value of each layer is initially set as =1 , the first and second layers are full connections, and equation (2) is obtained, where f_h is the activation function and W_{ij} is the connection weight.

$$\alpha_1^{(2)} = f_h\left(Z_1^{(2)}\right) = w_{10}^{(1)}\alpha_0^{(1)} + w_{11}^{(1)}\alpha_1^{(1)} + \ldots + w_{18}^{(1)}\alpha_8^{(1)} \tag{2}$$

$$\vdots$$

$$\alpha_{256}^{(2)} = f_h\left(Z_{256}^{(2)}\right) = w_{256}^{(1)}\alpha_0^{(1)} + w_{2561}^{(1)}\alpha_1^{(1)} + \ldots + w_{2568}^{(1)}\alpha_8^{(1)}$$

Formula (3) can be obtained after generalization to all subsequent neurons:

$$\alpha_i^{(l)} = f_h\left(z_i^{(l)}\right) = \sum_{j=0}^{n_{l-1}} w_{ij}^{(l-1)}\alpha_j^{(l-1)} \tag{3}$$

BP error reverse transmission is adopted, the training sample is compared with the output value of forward propagation, and the error is calculated, and then the error is propagated backward, and then the weight is adjusted[6]. The error backpropagation spreads to the second layer, and the error formula is deduced and formula (4) is obtained:

$$\mu_i^{(l)} = f_h'\left(z_i^{(l)}\right) \cdot \sum_{j=1}^{n_{l+1}} w_j^{(l+1)} w_{ji}^{(l)} \tag{4}$$

Table. 3 DLIB68 feature training data

feature value / sample	X1	X2	X3	X4	X5	X6	X7	X8	f(x1-x8)
1	0.364341	0.015504	0.062016	0.44031	0.083333	-0.18605	0.465116	0.682171	6
2	0.403226	0.040323	0.077419	0.448387	0.068548	-0.20161	0.475806	0.653226	2
3	0.481203	0.067669	0.173684	0.443609	0.06391	-0.21053	0.473684	0.661654	3
4	0.335766	0.043796	0.064964	0.448175	0.071168	-0.24818	0.459854	0.649635	5
5	0.328571	0.042857	0.067143	0.431429	0.076786	-0.22857	0.442857	0.685714	5
6	0.424	0.048	0.0856	0.4528	0.076	-0.2	0.48	0.64	2
7	0.333333	0.043478	0.081159	0.418841	0.070652	-0.19565	0.427536	0.702899	0
8	0.347222	0.013889	0.175694	0.4125	0.065972	-0.22222	0.444444	0.680556	1
9	0.348837	0.100775	0.089922	0.406202	0.031008	-0.17829	0.465116	0.705426	1
10	0.323077	0.015385	0.174615	0.433846	0.059615	-0.19231	0.438462	0.707692	0
11	0.333333	0.083333	0.083333	0.40303	0.077652	-0.19697	0.484848	0.643939	6
12	0.348837	0.100775	0.081395	0.412403	0.079457	-0.1938	0.472868	0.666667	6
13	0.343511	0.083969	0.08626	0.410687	0.082061	-0.20611	0.465649	0.664122	6
14	0.304348	0.210145	0.171014	0.408696	0.083333	-0.15217	0.442029	0.630435	6

4.2. The production of the training data

By making training data of character pictures, the pictures were first classified and labeled, which were divided into 7 expressions of normal, happy, angry, disgusted, fear, sad and surprise[7]. The 7 expressions were put into different folders respectively.

68 feature points on human face were obtained through DLIB, from which the data of face size, eyebrows, eyes, mouth, chin and so on were extracted. Due to the different shooting distances of photos, these data values were all based on the size of face.

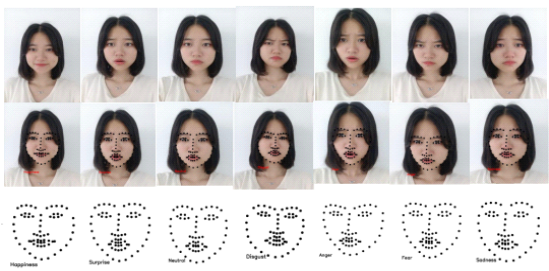
After extracting the data, store it in CVS file, and generate 0-6 values in the last column of attributes according to the classification of expressions, and scramble the order.

4.3. Stochastic training of the model

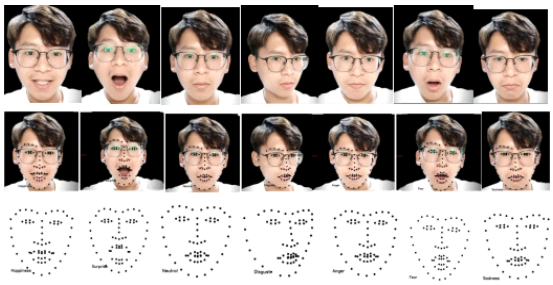
Google's TensorFlow framework was used to build a model for training[8]. During the training, 90% of data sets were randomly extracted as training sets and 10% as test sets. MLP network was used for training and supervised learning. When loss no longer decreased, the training ended and the prediction model was obtained.

4.4. Results of the model test

Through the test set, the accuracy rate is 93.2%, which has met our requirements. After the sample is put into the prediction model, the emotion classification is output. FIG. 2 shows the video test of a girl and a boy respectively. When each frame of picture is saved during the test, it can be seen that their mood state can be well recognized, even wearing glasses does not affect the recognition effect, as Shown in FIG. 2(b) :



a.Girl



b. Boy

Fig. 2 Examples of Test Subjects and Respective Facial Landmarks

The first line shows the original image collected from the test subject. The second line shows the original image of the face after emotion was recognized in the real interaction. The third row shows the region of the face after emotional feature shave been extracted, denoted by black dots.

4.5. Evaluation methods

In this study, seven categories of 7*7 Confusion Matrix were used. In order to know the classification efficiency of the model, Accuracy and Confusion Matrix were used to analyze the overall and respective emotional classification results. The accuracy of the confusion matrix is shown in Table-4 below:

In Table. 4, 100 photo samples of college students are selected to test the recognition effect. The sample is a series of photographs taken by each person. Each group included Anger, Disgust, Fear, Happiness, Sadness, Surprise and Neutral expressions. The results showed that the recognition rate of Anger was 95%, Disgust 97%, Happiness 96%, Fear 96%, Sadness 97%, Surprise 95%, Neutral 93%. Through the confusion matrix, we can get the quantity or probability of each emotion category, and better understand the prediction ability of the model in each emotion. High recognition rate can be achieved if this recognition method is used to identify real-time continuous videos. It can recognize the facial expressions of the characters in each frame of the video, and obtain the current mood state after statistics and analysis, which can be used for auxiliary psychological

Table. 4 Accuracy rate of emotion recognition mixture matrix

EMOTION DETECTED / EMOTIONS EXPRESSED	Anger	Disgust	Happiness	Fear	Sadness	Surprise	Neutral
Anger	95	3	0	0	2	0	0
Disgust	0	97	0	3	0	0	0
Happiness	0	0	96	0	0	3	1
Fear	1	0	0	96	0	3	0
Sadness	3	0	0	0	97	0	0
Surprise	0	0	2	2	0	95	1
Neutral	3	0	2	0	2	0	93

intervention and treatment.

At present, the traditional mental health assessment is mainly based on self-assessment questionnaire and structured interview, which obtains data information from the interviewees through face-to-face interpersonal interaction mode, and then evaluates the mental health status of the interviewees[9]. This kind of traditional mental health evaluation method obtains the mental health status of the subjects subjectively from the perspective of scale, but there are some problems in the implementation process and evaluation work, such as the social approval response bias, the real-time evaluation is not high, and the evaluation work is passive. In view of the problems existing in the traditional way for mental health assessment, and the continuous improvement of facial emotion recognition accuracy and convenience, we will mood ecological momentary assessment theory applied to the evaluation of mental health, the extraction of college students' negative emotions in online psychological evaluation data characteristics, so as to realize the multidimensional evaluation of mental health status. Multidimensional evaluation refers to data evaluation that includes two or more different forms or sources. The test text and facial expression are the external manifestation of human mind and psychology, which can reflect individual psychological state[10].

With the highly developed Internet and the persistence of COVID-19, people are gradually adapting to the way of online office life and work. Data of different dimensions complement each other and can provide more explanatory information. By integrating and interpreting the multidimensional mental health assessment data, a more comprehensive and systematic analysis and assessment of students' mental health status can be realized.

Ⅴ. Conclusion and prospect

5.1. Improve the effectiveness of psychological problems identification and timely prevention

This study focuses on daily mental health management and psychological crisis detection and prevention. Human emotions can be reflected in facial expressions, and with appropriate computer recognition modeling, tools such as computer vision and DNN can be used to identify human emotions. Through the construction of deep learning system of computer automatic emotion recognition and the traditional psychological evaluation, interview and classroom activities, we provide more dimensions to provide indicators of mental health. Firstly, in the aspect of psychological crisis prevention, emotion identification combined with quantitative questionnaire survey can further improve the identification degree of psychological crisis of college students, find high-risk groups more quickly and accurately, and help them; Secondly, in terms of individual case identification and timely intervention, the emotion recognition system can monitor the interviewees' emotional state from time to time. Emotional data combined with the rhythm of psychological interview can better grasp the key points of interview intervention, which is conducive to early discovery, intervention and

solution of psychological problems.

5.2. Explore the multi-dimensional construction approach of mental health online platform

With the increasing awareness of public mental health, the management function of mental health will gradually receive more attention. Online management platform can be used to obtain real-time indicators of mental health, and can be given real-time health advice. In this way, the participation of artificial intelligence can be greatly increased and the current situation of insufficient mental health teachers in colleges and universities can be alleviated. Moreover, people are often affected by emotions, as psychological teachers or counselors are also the same. Deep learning emotion recognition system can also track the emotional state of consultants in real time and give appropriate supporting suggestions. The combination of human and intelligent system will be more objective and comprehensive, better protection for patients and consultants, and work will be more smooth.

5.3. Exploring a new mode of mental health management in the post-COVID-19 era

The COVID-19 pandemic requires a healthy distance between people. In this era of artificial intelligence, the emotion recognition system combined with deep learning can evaluate the emotional state of college students anytime and anywhere. In this way, it is more beneficial for college students to understand their own mental state and conduct proper mental health management. Instead of waiting for serious psychological problems to see a doctor, users can easily know their emotional state at ordinary times. The traditional mental health screening in colleges and universities is usually a unified evaluation at the time of admission, but in this way, only the results of questionnaire evaluation can be seen after the fact, and the emotional state of the parties at that time is unknown. Therefore, the emotion recognition system of deep learning just makes up for this gap. It can not only better understand and evaluate students' emotional state at that time, but also achieve real-time monitoring.

5.4. Epilogue

This paper fully considers the structure and the relationship between convolutional neural network and activation function and the deep learning model of face emotion recognition, designs corresponding models based on ReLU and sigmoid activation functions, verifies the effectiveness of the model in detail through experiments, and has certain reference value for the model performance of deep learning emotion recognition. However, due to the uneven distribution of expression data, insufficient test set, the selection of activation function and other factors, it has a certain impact on the experimental process, which should be improved in subsequent studies.

REFERENCES

[1] K. H. Kyu and K . H. Sung, "Research on complex Emotion Expression scheme of children's interactive intelligent toy--Emotion analysis of human expression and voice," *Journal of Korean Design Culture Society*, 2021.

[2] K. D. Cho, "Physical Features of Faces and Personal Personality Impression," *Science of Emotion and Sensibility*, vol. 7, iss. 2 , pp. 195-201, 2004.

[3] D. Jane and Z. Kim, "The Historical Landscape of Mental Health Evaluation and Treatment," *Physician Assistant Clinics*, vol. 6, no. 3, pp. 361-369, 2021

[4] J. Harikrishnan, A. Sudarsan, A. Sadashiv, and R. A.S. Ajai, "Vision-Face Recognition Attendance Monitoring System for Surveillance using Deep Learning Technology and Computer Vision," in *International Conference on Vision Towards Emerging Trends in Communication and Networking*, pp. 1-5, 2019.

[5] G. Yang and J. Saniie, "Indoor Navigation for Visually Impaired Using AR Markers," in *Electro Information Technology (EIT), 2017 IEEE International Conference on*, Lincoln, 2017.

[6] Ian Goodfellow, Yoshua Bengio, and Aaron Courville, *Deep Learning*, The MIT Press, pp. 199-208, 2016.

[7] Q. Mao, X. Pan, Y. Zhan, and X. Shen, "Using Kinect for real-time emotion recognition via facial expressions," *Frontiers of Information Technology&Electronic Engineering*, vol. 12, pp. 272-282, Apr. 2015

[8] S. H. Lee, "Facial Data Visualization for Emotion

Recognition of Deep Learning," *Journal of Information Science Theory and Practice*, 2019.

[9] L. Jing, L. Dexi, W. Chanxuan, L. Xiping, Q. Xiangqing, B. Liping, and Z. Tingshao, "A review on Automatic Assessment of Mental Health for Social Network Users," *Journal of Chinese Information Processing*, vol. 35, no. 2, pp. 20-29, Feb. 2021

[10] T. Perron, T. Jakubowski, and C. Razzi, "Physical Health Assessment of the Frequent Visitors in the School Setting-Part 1: An Overview," *NASN School Nurse*, vol. 36, no. 4, Mar. 2021.

리파(BO LI)

2005년 중국타이위안사범학원 (이학사)
2011년 중국절강사범대학 교육학원(심리학석사)
2019년 배재대학교 심리상담학과(심리학박사)
2011년~현재: 중국한산사범학원 (강사)
※관심분야 : Counseling Psycholog, Cognitive psychology, Career planning, Expressive art therapy.

조경덕(Kyung-Duk Cho)

1990년 경북대학교(문학사)
1993년 일본 치바(千葉)대학교 인문과학대학원(심리학석사)
1998년 동경도립대학교 인문과학대학원(심리학박사)
2003년 ~ 현재 배재대학교 심리상담학과 교수
※관심분야 : Counseling Psychology, cognitive psychology, Interpersonal Relation Psychology, psychological test.