Development of a Database of the Frequently Used Emotional Vocabulary of 4~6 Year Olds

Whei-Jane Wei

University of Taipei
1, Aiguo West Road
Taipei 10048, Taiwan, R. O. C.
Tel: 886-2-23113040 Ext.4915, Fax: 886-2-23754823, E-mail: wwj@go.utaipei.edu.tw

Abstract

Young Children Emotional Competency Scale (YCECS) [1] is a standardized tool assessed by teachers. Is it possible to directly measure emotional competency of 4- to 6-year old? Is database of frequently used emotional vocabulary (DFUEV) significantly correlated with YCECS? How reliable and valid are DFUEV? Instruments comprise developing 40 emotional theaters and 5 rating scoring standards. Procedures include selecting 200 4- to 6-year-old and conducting experiments. Findings indicate inter-rater reliability of 96.9% and criterion-related validity reaching significant.

Key words: database, emotional competency, young children

Introduction

In the era of big data challenges, using big data to analyze the past and predict the future is a trend in the study early childhood development. At present, the evaluation of children's emotional ability is mostly understood through long-term observation by teachers. Many studies have indirectly assessed the emotional scores of young children using a rating scale. Other studies have directly evaluated emotional ability based on expressions, voices, and words [2]. In order to solve these two problems, the purpose of this study is to explore scores directly by using artificial intelligence to detect children's emotional ability (DEVYC) and indirectly by using the YCECS. Explore the relationship between DEVYC and YCECS.

A relate study used emotional pictures for expression recognition and found that widely used facial emotion pictures (PoFA; i.e. "Ekman face") and the Radboud Faces database (RaFD) are generally not considered to show real emotions [3].

Therefore, this study used actual shooting of children's films for facial emotion recognition. Expression recognition of 55% of the emotional score is calculated as follows: We use expression recognition technology to analyze 200 children's responses to each question for expressions of seven emotions, including angry, despised, disgusted, happy, not opinion, sad and surprised. We then calculate the number of times each emotion is identified. According to the database, the score is converted into five levels multiplied by the number of occurrences, and finally added to the numerator. The sum of the occurrences of the seven emotions is the denominator, and the weight distribution of each emotion in each question is calculated. According to the five levels of the database, if the percentage reaches 80%~100%, 5 points are given. We use Microsoft's Project Oxford tool to establish expression recognition as the sub-criteria [4].

Previous studies have found it is impossible to achieve satisfactory results by recognizing emotions based on speech or facial expressions [5] [6]. Therefore, this study first established a common emotional vocabulary database (DEVYC) for children and developed emotional theatre experiments. Young children identified facial expressions while enjoying the experiments and produced facial expressions (55% = A). Second, the teachers presenting the emotional theatre asked the children questions and elicited responses. The scores were compared to DEVYC according the phonetic text of the children's answers, producing a speech recognition sentiment score (38% = B). Finally, we asked the children to draw the parts they liked, verbally interpret the speeches and then convert them to text. The scores were then compared to DEVYC to produce a character recognition sentiment score (7% = C). The emotional scores obtained by young children (A + B + C) are related to the standardized test, YCECS (standard), developed by the investigator.

Related study used a set of expressions consisting of four expressions (happy, angry, sad, neutral) [7]. These expressions were then entered into three different deep neural network models, Restricted Boltzmann Machine, Deep Belief Networks, and Stacked Autoencoder with Softmax Function, for comparison. The final comparison shows that the two layers of the Restricted Boltzmann Machine and Deep Belief Networks training models have more than 100 layers, but the accuracy is only 25%. Stacked Autoencoder with Softmax Function has 96% accuracy rate and can capture features well, but the time needed to identify emotions is greatly increased because there are too many parameters.

Most existing emotion recognition techniques rely on the CNN training model. Karen et al. proposed a VGG16 architecture based on the CNN network [8]. VGG16 emphasizes the importance of depth in CNN networks. Compared to CNN, each convolutional layer filter is compressed to 3*3, replacing the larger filter commonly used in CNN. The efficiency of the convolution operation is improved, but the parameters in the fully connected layer of the last layer still account for 90% of the overall network. Therefore, Christian et al. proposed the Inception V3 architecture, which uses Global Average Pooling technology [9]. Averaging the data of the pooling layer can solve both the problem of having too many parameters and features captured by the layer.

Today's CNN-based emotion recognition technology has solved the problem of having many parameters. However, if there are too many filters on the convolution operation layer, the time required for the convolution operation on the convolution operation layer is greatly increased. Thus, Francois et al. proposed the Xception architecture. Through depth-wise separable convolution technology, the speed of convolution operations is further improved and the amount of data is reduced [10]. In the past, the general CNN convolution operation method convolved with each filter for each layer of data. Depth-wise separable convolution is a layer of data that is convoluted with only one filter, and a 1*1 filter is created for each output. Finally, it lets each layer of output data and each filter perform another convolution operation. In this way, the amount of computation of the convolution operation can be reduced from 1/8 to 1/9 of the general CNN.

Related study found that the success of emotional recognition depends on the transmission of information between the computer and the face [11]. The computer should be able to get information about the face instantly, such as emotions and gender. However, because the information of the face is highly complex, there are too many parameters in the process of machine learning. This makes it impossible to achieve instant recognition. Therefore, this study uses a real-time CNN-based emotion recognition program for identification. By using Global Average Pooling in the architecture, the information in each image is compressed while retaining important features. Moreover, the depth-wise separable convolutions technique reduces the amount of computation in the convolution operation, making emotional recognition of images in real-time possible.

Methods

This study attempts to answer the following research question: Is there a significant positive correlation between DFUEV and YCECS? This research method uses in-depth interviews, focus discussion, observation methods and experimental methods. The implementation process includes: (1) Development of emotional theater based on 40 emotional abilities in YCECS scripts, (2) designing 40 key competency questions for children to understand and answer easily, (3) developing five assessments and indicators for each of YCECS's 40 emotional abilities, and (4) designing 40 emotional theater experiments. In the experiments, we (5) randomly selected 200 children aged 4-6 years, (6) performed 40 emotional theater experiments, (7) conducted observer training, (8) analyzed the consistency of four observers, and completed the DFUEV scores. The reliability study (9) analyzes YCECS and DFUEV correlation, establishes the validity of the association, and (10) identifies the sequence of the children's emotional vocabulary database (DFUEV).

This study used Microsoft Azure Bing Speech-to-Text technology to convert children's spoken answers into words. Machine learning uses the Microsoft Azure Text Analytics technology [5] to analyze the emotions contained in the text and to give the child emotional ability scores by analyzing the keywords that appear in the text. The results of the analysis were returned to the database established by this study (DFUEV) for further analysis. In the course of the game, this study records the expressions of children throughout the whole process. This study uses Microsoft Azure Emotion technology [6], which can more accurately analyze seven expressions, namely, anger, contempt, disgust, happiness, no opinion, and sadness. The results of the analysis are stored in the database (DFUEV) established in this study, and reliability and validity analysis of the data in the database are carried out.

This study creates a database to store the results of Microsoft Azure [6] analysis. A UserID is provided in the user profile, which is unique in the database, and each UserID corresponds to a user. The database also records the user's name and age. Through the unique UserID, the user's game record can be queried in the game data table, such as the field after he/she verbally answers the question and converts the text during the game (SpeechToText), text analysis results (TextAnalyticsResult), and finally the emotional data represented by the user's expression (FaceEmotion).

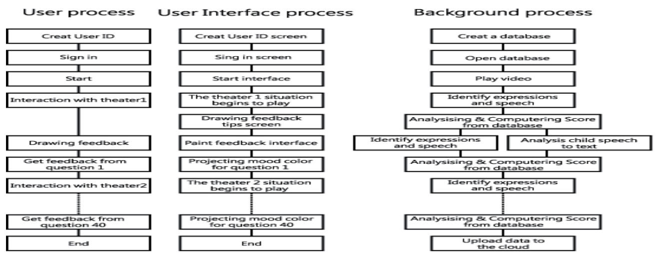

Fig. 1 User process, UI interface process and background process

Start working horizontally from the first step until you complete 40 questions.

App system flow

A. Test process

The child's identification ID number is used to store the child's photo, drawing and database. The child presses "Enter" to jump to the photo taking step. After 5 seconds, the image is automatically captured and the photo is accessed based on the child's identification ID number, such as A123 photo file: A123_date-photo.jpg. After 5 seconds of photographing, the system automatically jumps to unit category classification.

Press the corresponding question number to enter the corresponding page. The program plays the corresponding movie according to the corresponding number. After the movie is finished, it automatically jumps to the system to ask questions.

When the system asks questions, the animation of the teacher is played in the upper right corner; the top left corner shows the big pictures taken by the children.

When the system background, microphone preparation and Google speech recognition system are in progress, a red microphone icon is displayed on the screen.

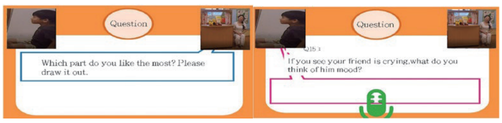

Fig. 2 Q&A

Fig. 2 System questioning problem indicted that drawing preparation stage. When the system asks questions, the animation of the teacher is played in the upper right corner. The system background and canvas preparation are in progress.

B. Expression Score

TABLE 1
EXPRESSION DATABASE

Expression database – score ratio 55%		
Question number	Emoticon	Give points by color
#1	Expression	Average scores for

	File001	Expressions on different timelines.
#2	Expression File002	Average scores for Expressions on different timelines.
#40	Expression File040	Average scores for Expressions on different timelines.
Average score of 40 questions		

TABLE 2
VOICE DATABASE

Voice database – score ratio 38%			
Question number	Voice to Text file	Database score	Other"answers are not in the database"
#1	Text file 001	score	Manual or machine learning judgment score
#2	Text file 002	score	Manual or machine learning judgment score
#40	Text file 040	score	Manual or machine learning judgment score
Average score of 40 questions			

TABLE 3
PICTURE (TEXT) DATABASE

Picture(text) database – score ratio 7%			
Question number	Voice to Text file	Picture file	Score
#1	Text file 001	Picture file 001	Manual or machine learning judgment score
#2	Text file 002	Picture file 002	Manual or machine learning judgment score
#40	Text file 040	Picture file 040	Manual or machine learning judgment score
Average score of 40 questions			

The child's emotional ability score is: expression score x 55 % + speech score x 38 % + picture (text) score x 7 %

Result

Table 3 shows that α of DFUEV is 0.75, and Table 4 shows that α of YCECS is .98. Table 4 shows that there is a significant correlation between DFUEV and YCECS, r = .41 (p = .026 *). The conclusions of the study suggest there is a significant positive correlation between DFUEV and YCECS. This study found that DFUEV is suitable for machine measurement. With regard to the emotional ability of 4~6 year olds children, when using DFUEV and machine learning, the child's emotional development norm will be constantly updated automatically.

A. Reliability Analysis of DFUEV and YCECS

The reliability analysis evaluates the reliability of the whole scale. The reliability evaluation of this study is based on “developing a common emotional vocabulary database for children” to perform Cronbach's Alpha detection. The DFUEV reliability is estimated as follows:

TABLE 4
RELIABILITY STATISTICS

Cronbach's Alpha Value	According to standard items Cronbach's Alpha Value	Number of items
.747	.710	37

Table 4 shows that the DFUEV reliability coefficient Alpha is Cronbach's α 747, and the normalized reliability coefficient is .710. The normalized α indicates the influence of the unequal volatility of each subject, and the corrected coefficient.

TABLE 5
RELIABILITY STATISTICS

Cronbach's Alpha Value	According to standard items Cronbach's Alpha Value	Number of items
.976	.976	40

Table 5 shows that the YCECS reliability coefficient Alpha is Cronbach's α .976, and the normalized reliability coefficient is .976. The normalized α represents the influence of the unequal volatility of each subject, and the corrected coefficient.

B. Correlation Analysis between DFUEV and YCECS

TABLE 6.
DESCRIPTIVE STATISTICS

	Average value	Standard deviation	Number
YCECS	35.6000	8.37731	30
DFUEV	42.6667	6.13263	30

Table 6 shows that the sample averages of DFUEV and YCECS are 35.6 and 42.7 respectively.

TABLE 7
CORRELATION ANALYSIS

		YCECS	DFUEV
YCECS	r	1.000	-.406"
	Signficance	.	.026*
	N	30	30
	r	-.406"	1.000
	Signficance	.026	.
DFUEV	N	30	30

*The relevant significant level is 0.05 (two-tailed).

Table 7 shows that Spearman's rho coefficient reaches up to .406*. (p = .026), which is significant, indicating that DFUEV and YCECS are significant. When it comes to correlation, YCECS [1] is a standardized test with high reliability and validity, so the DFUEV developed in this study also has high reliability and validity.

C. Database of Children's Frequently Used Emotional Vocabulary

TABLE 8
TOP 5 FREQUENTLY USED EMOTIONAL VOCABULARY DATABASES

5.Passed: Say two kinds of emotional words 4. Partially passed: Say one emotional words		3. Non-emotional word	2. Wrong answer	1.Didn't answer
Sad(25%)	Happy(60%)	Very uncomfortable(13%)	Thank you(50%)	Did not answer
Unhappy(25%)	Well(13%)	Thank you(13%)	I feel very happy eating a lollipop(50%)	
grieved / bad / worried / not very good(13%)	joyful / a little happy(7%)	stomachache(13%)		
Uncomfortable(13%)	Fulfilled(3%)	Too much(13%)		
	Glad(3%)	Want to go home to sleep(13%)		
	Good(3%)	Want to go to the toilet(13%)		
		not good(13%)		
		Crying(13%Thank you)		

D. Database Construction Section

1. Child's Personal Information Sheet

The purpose is to record the basic information of each child, recording the child's identity card number, name, birthday, headshot electronic file, and age, for subsequent use.

2. Children's Emotional Five-Level Data Sheet

The purpose is to record the emotional ability of students when answering questions, including recording what students are asking, thinking, doing, on commenting, which areas to answer, which questions to answer, and the emotions to be identified when answering questions. The expression recognition score (55%), speech recognition score (38%) and language recognition score (7%) are partially used as the scoring basis, and then converted into the five-level emotional results according to the scores for subsequent language consistency. Expression consistency analysis is conducted.

E. Emotional Recognition Results (Using Microsoft Azure Api to Identify Images)

The expression recognition software used in this study has seven emotions, and the image recognition test is performed using Microsoft Azure Api [6].

F. Emotional Recognition Results (Using Pen Lens Recognition)

This research program design uses an open source real-time face detection and emotion program, combined with OpenCV and CNN training models for real-time image emotion recognition. The training data set is fer2013. CNN goes through multiple convolutions, pooling, and fully-connected steps to determine what emotion the current expression is [7].

In the beginning, when the characteristics of the feature are unknown, the user needs to compare the matching features of all the regions in the image. The mechanism used in this step is convolution, which fixes the pixels of each block in the fixed size. After performing a convolution operation on the feature detector to be detected, the result value of the degree of coincidence of the block with the feature is obtained.

Pooling is a method of compressing images and saving image features. Select a fixed square range (EX: 2X2, 3X3, ...) on the image, and take the maximum or average value of the convolution operation in the range. As a feature of the range, it is recompressed into a picture that retains only the features of each range. Although the number of pixels is reduced, the degree of matching between each range and each feature is retained; that is, the information after the pooling can focus more on whether there are matching features in the image, not where the features exist in the image.

After repeating the steps of convolution and pooling, the steps of the full connection are entered, the pooled result matrix is transformed into one dimension, and the values in the matrix are voted on in the result. The result with the highest number of votes is the current result of the identification. Different values have different degrees of discrimination for different results; for example, some values can better judge the emotion of happiness. In such a case, this value can enter more votes than other values, and all values have different options. The number of votes cast will be expressed in terms of weight or connection strength. For the computer, the image captured by the lens is composed of multiple images. By learning and recognizing each image through the CNN training model, the recognition result of the current expression can be obtained.

Discussion

A. Correlation between DFUEV and YCECS

To bring human-computer interaction close to people, studies have made robots have emotional abilities similar to humans, giving them the ability to generate and express emotions. A previous proposed that in the expression of human emotions, language accounts for 7%, sound makes up 38%, while the remaining 55% comes from facial expression [12]. Therefore, it is very important to be able to correctly express the emotional expressions of the face. As long as a facial expression can be correctly expressed, most of the expression of human emotion can be accomplished.

Other studies have also pointed out that human expression involves the muscle movement of the face. A specific expression can be assigned to each action unit according to the intensity of the contraction. Therefore, we can detect a facial expression to observe the emotion of the subject. However, research has shown that this method is very time consuming, so machine must be used to speed up the process [8].

B. Emotional Vocabulary Database for Children

The feelings in this study can effectively improve the emotional ability of young children, and support the emotional theater experiments. The proposed approach offers a feasible and effective design to improve children's emotional ability. The results were same as those of a related study [9], in which teachers used the thematic teaching. The teacher in the study told the children that the puppet was frustrated, and then asked the children to close their eyes and listen to the teacher telling a story and talking to the puppet. The related study also indicated that using seven effective strategies to improve children's vocabulary had a significant relationship with students' social emotional vocabulary scores [10].

References

[1] W. J. Wei, *Emotional Competency Scale for Young Children*, Taipei: Psychological Publishing Co., Ltd., 2011.

[2] K. E. Darling-Churchill and L. Lippman, Early childhood social and emotional development: Advancing The Field of Measurement, *J. Appl. Dev. Psychol,* 45, 2016, pp. 1-7.

[3] A. Dawel, et. al., Perceived Emotion Genuineness: Normative Ratings for Popular Facial Expression Stimuli and the Development of Perceived-As-Genuine and Perceived-As-Fakesets, *Behav Res Methods,* 49(4), pp. 1539–1562, 2017.

[4] M. Zhao, et. al., Emotion Recognition Using Wireless Signals, *Commun ACM,* 61(9), pp. 91-100, 2018.

[5] M. A. Yaxiong, et. al., Audio-Visual Emotion Fusion: A Deep Efficient Weighted Approach, Inform Fusion, 46, pp. 184-192, 2019.

[6] M. Srikanth, et. al., Azure Cognitive Services, In: *Developing Bots with Microsoft Bots Framework*. Berkeley, Apress, pp. 233-260, 2018. https://azure.microsoft.com/en-us/try/cognitive-services/

[7] P. Cunha, *README.md*, GitHub, Inc., 2018. https://github.com/haslab/Electrum/blob/master/README.md

[8] S. Delplanque, A Comment on Prescott's Call for Prudence and Rigor When Measuring Emotions. *Food Qual Prefer,* 62, 2017,

pp. 372-373.
[9] G. E. Joseph, et al., Enhancing Emotional Vocabulary in Young Children, *Young Exceptional Children,* 6(4), pp. 18-26, 2003.
[10] L. S. Poventud, et al., Developing Social-Emotional Vocabulary to Support Self-Regulation for Young Children at Risk for Emotional and Behavioral Problems. *Int J Sch Cog Psychol,* 2(143), 2015.
[11] D. Mehta, et al., Facial Emotion Recognition: A Survey and Real-World User Experiences in Mixed Reality. Sensors*,* 18(2), 416, 2018.
[12] A. Mehrabian, Nonverbal Communication, Routledge, 2017.

ISBN: 978-981-14-2064-1