E-learning System Based on Neural Networks

ZHANG Linfeng,YU Fei+,SHEN Yue,LIAO Guiping,Chen Ken

***Abstract*—Although the current E-Learning systems have many merits, many of them only treat advanced information technology as simple communication tools, and release some learning contents and exercises in the network. In this paper, a one-class-in-one network for emotion recognition system in E-learning is proposed and implemented in the paper. Using a large database of phoneme-balanced Chinese words read by speakers consciously trying to portray an emotion, we trained and tested this module. We achieved a recognition rate of approximately 55%. The results obtained in this study demonstrate that emotion recognition in speech is feasible, and that neural networks are well suited for this task.**

***Index Terms*—Affective Computing, Speech Emotion recognition, Neural Network, E-Learning.**

I. Introduction

E-Learning uses modern educational technologies to implement an ideal learning environment through integrating the information technology into curriculum, which can embody the learning styles of students' main-body function, reform the traditional teaching structure and the essence of education thoroughly [1].

Although the current E-Learning systems have many merits, many of them only treat advanced information technology as simple communication tools, and release some learning contents and exercises in the network [2]. This kind of movable textbook or electronic textbook is indifferent to the learners, which lacks of the interaction of emotion. Besides, this kind of learning materials without using of the superiority of interactive multimedia technology and displaying the function of network effectively, which leads to the phenomenon of emotion deficiency in the current E-Learning system.

Emotion deficiency refers to the separation among students and teachers, students and students, which make students and teachers, students and students can't carry on face to face communicating promptly like conventional education. Thus, some learning problems of the learners in the learning process can't be solved and perplexity of the psychology can't get help. If students gaze at indifferent computer screens for a long time, they do not feel the interactive pleasure and emotion stimulation, and they may have antipathy emotion.

The network institute of Central China Normal University (CCNU) has carried out an investigation on learning conditions for 203 distance education learners in 2006. They took out two results that are related to emotion deficiency form the investigation results. Two results are shown in table1 and table 2.

From tables 1 and table 2, the influence that is caused from the separation among teachers and students or students and students is obvious (obvious influence and great influence occupy 56.5%. The influence that is caused from the interaction deficiency of teaching contents is obvious too. (Obvious influence and great influence occupy 63.8%).

Table 1. The influence is caused from the separation among teachers and students or students and students.

Influence degree	Great influence	Obvious influence	A little influence	No influence
Student number	33	81	62	27
percent	16.3%	40.2%	30.3%	13.2%

Table 2. The influence is caused from the interaction deficiency of teaching contents.

Influence degree	Great influence	Obvious influence	A little influence	No influence
Student number	42	87	42	32
percent	20.8%	43.0%	20.6%	15.6%

How to measure cognitive emotion of learners in the E-Learning system and realize harmonious emotion interaction becomes an important research topic in the distance education [3]. Aiming at the problem of emotion deficiency in E-Learning, domestic and abroad scholars bring forward some strategies as follows:

(1) Designing the emotional network curriculums. The emotional mark is added to interactive network curriculums.

(2) Implementing exploring and cooperative learning. The learning groups are constructed, which is advantage to the communication among students and teachers.

Manuscript received December 29, 2006.

This work was supported in part by the Guangdong Province Key Lab of Electronic Commerce Market Application Technology , Guangdong University of Business Studies No. 2006gdecof002

ZHANG Linfeng is with School of Computer & Information Engineering, Hunan Agricultural University, Changsha 410128,P.R. China(Email: zlf@hunau.net)

YU Fei is with Provincial Key Laboratory of Computer Information Processing Technology, Suzhou University, Suzhou, 2150063,P.R. China and Guangdong Province Key Lab of Electronic Commerce Market Application Technology , Guangdong University of Business Studies, Guangzhou,510320,,P.R. China (Email: yufei@hunau.edu.cn or hunanyufei@126.com)

SHEN Yue is with School of Computer & Information Engineering, Hunan Agricultural University, Changsha 410128,P.R. China(Email: shenyue@hunau.net)

LIAO Guiping is with School of Computer & Information Engineering, Hunan Agricultural University, Changsha 410128,P.R. China(Email: lgpxf@hunau.net)

Chen Ken is with School of Computer & Information Engineering, Hunan Agricultural University, Changsha 410128,P.R. China(Email: chk@hunau.net)

+To whom correspondence should be addressed. E-mail:yufei@hunau.edu.cn or hunanyufei@126.com

(3) Implementing blended learning. The superiority of traditional learning and E-Learning is combined to display the leading role of teachers and the main-body function of students.

(4) Improving learning supporting service system. Comprehensive, prompt, and convenient learning of support service is provided to maintain learners' positive emotion.

The application of above strategies have avoided emotion deficiency in certain degree, but learner's emotion state cannot be tracked accurately, the corresponding emotional encouragement and compensation also cannot be provided according to specific emotion state, which cannot help the learner to solve emotion deficiency fundamentally.

Affective computing is a hot topic in Artificial intelligence, it is computing that related to, arise from, or deliberately influence emotion [4], which is firstly proposed by Professor Picard at MIT in 1997.Affective computing consists of recognition, expression, modeling, communicating and responding to emotion [5].In this components, emotion recognition is one of the most fundamental and important modules. It is always based on facial and audio information.

Basing on it, affective computing is applied in the traditional E-Learning system, while the model of E-Learning system based on affective computing is proposed in the paper. Speech emotion recognition is used to construct affective computing module. Then, the emotion state is judged and understood. The corresponding emotion encouragement and compensation are provided according to the specific emotion state. Teaching strategies and learning behaviors are adjusted according to learners' emotion state. Thus, the system could help the learners to solve emotion deficiency in E-learning system essentially.

II. The Model of E-learning System Based on Affective Computing

In the learning process, language of learners is abundant. When they understand and accept the learning content, their emotion is high and displays as cheerful spoken language. Otherwise, their emotion is low.

These emotional behaviors are quite important feedback signals of learners. We may use these feedback signals effectively and adjust teaching strategies to serve personalized learning. Basing on it, the traditional E-Learning model is added the affective computing module. The model of E-learning system based on affective computing is Fig.1.The system model is composed of several modules as follows:

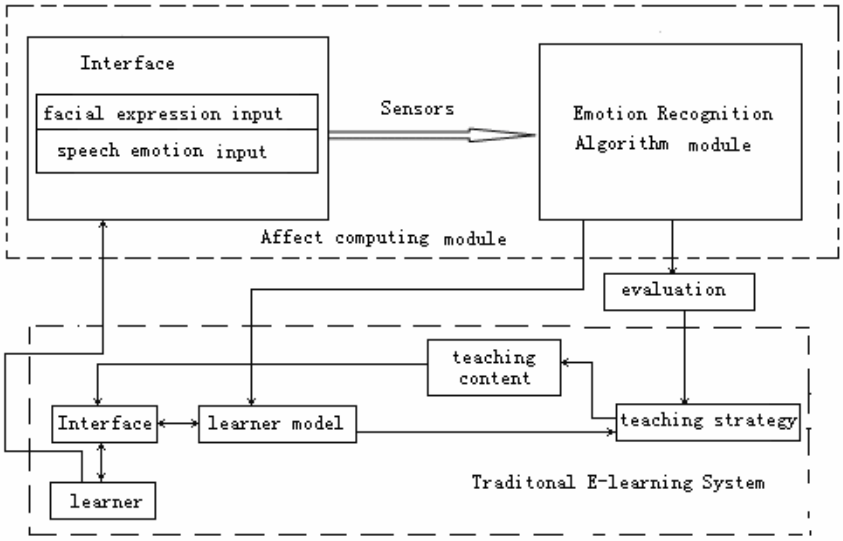

Fig. 1 The model of E-learning system based on affective computing.

Interface module: affective computing input (speech emotion recognition input)is added to human machine interface of traditional E-Learning system, which collects learners' emotion feedback information primarily, thus emotion compensation is realized.

Emotion Recognition module: emotion recognition module is composed of input, pre-processing, feature extraction, feature selection, emotion recognition and output. Basing on it, the appropriate evaluation of learning is obtained.

Evaluation module: it collects evaluation results and transforms into a corresponding evaluation parameters. Simultaneously, Learning records are extracted from learners' model.

Teaching strategy adjusting algorithm module: teaching strategies are adjusted according to the evaluation parameters and learning records of learners' model. Appropriate teaching contents and teaching mode are provided to learners. Simultaneously, emotional encouragement and compensation is provided.

Expanding learner's model: it records personal information mainly, including learning background information, cognitive style information, emotional information, and so on.

III. Emotion Recognition Module

A. Emotions

Emotions conveyed in speech can be grouped into two main categories: consciously expressed emotions; and unconsciously

expressed emotions. Consciously expressed emotions are generally more obvious than unconsciously expressed emotions. For example, when someone raises their voice in speaking, they are often consciously expressing that they are angry. However, in other cases, the only indication of a person trying to conceal their anger might be a slight terseness to their words.

Consciously expressed emotions are easier for humans to recognize[6], and significantly easier to gather data on. Therefore, this study is limited itself to the recognition of emotions that are consciously and purposefully expressed by the speaker. We expect to be able to expand our methodology to unconscious or concealed emotions as well in future work.

B. *Classification of emotions*

How to classify emotions is an interesting but difficult issue. Researchers on emotion recognition differ on the number of categories and the kinds of categories to use. Some classification systems that have been used include[7] [8]:

Neutrality, joy, boredom, Sadness, anger, fear, indignation

Neutrality, happiness, sadness, anger, fear, boredom, disgust

We dealt with four emotional states (anger, sadness, happiness and cheerfulness) in our previous study; based on examining these examples, and on the consideration that increasing the number of recognizable emotional states is effective for achieving interaction between humans and computers, we have selected the following eight emotional states to use in this study: joy, teasing, fear, sadness, disgust, anger, surprise, neutral.

C. *Speaker and Context Independence*

Speaker independence is an important part of speech and emotion recognition. A speaker-dependent system requires a training period for each new speaker before the system is able to function at a reasonable level. On the other hand, a speaker-independent system will tend to have a lower level of accuracy, since it is not finely tuned for each speaker. However, eliminating the need for training sessions with each new speaker seems well worth the resultant loss in accuracy. By carrying out initial training using a number of different speakers, our system has become speaker-independent. Context independence is also a desirable quality in emotion recognition systems. A system that can determine the emotion in an utterance regardless of the context or text of the utterance is considered context-independent. A context-dependent system would require language understanding in order to determine the context. Our system achieves context independence by using a large set of phoneme balanced words as its training set.

Of course, there are other important factors, such as the effect of social and cultural differences. However, these are difficult issues that will require long term research. Therefore, in this research, we deal with emotions contained in the utterances spoken only by Chinese speakers. Under this restriction, we tried to achieve speaker-independent and context independent emotion recognition.

IV. Processing Flow

The processing flow of speech emotion recognition is illustrated in Fig. 2. The process is divided into two main parts: speech processing and emotion recognition.

A speech input (an utterance) is input into the speech processing part. First, the speech features for that utterance are calculated. Next, the utterance is divided into a number of speech periods. Finally, for each speech period the speech features are extracted, and features for the utterance are compiled into a feature vector. The feature vector is then input into the emotion recognition part. In the training stage, the feature vector is used to train the neural network using back propagation. In the recognition stage, the feature vector is applied to the already trained network, and the result is a recognized emotion. These steps are explained further in the following sections.

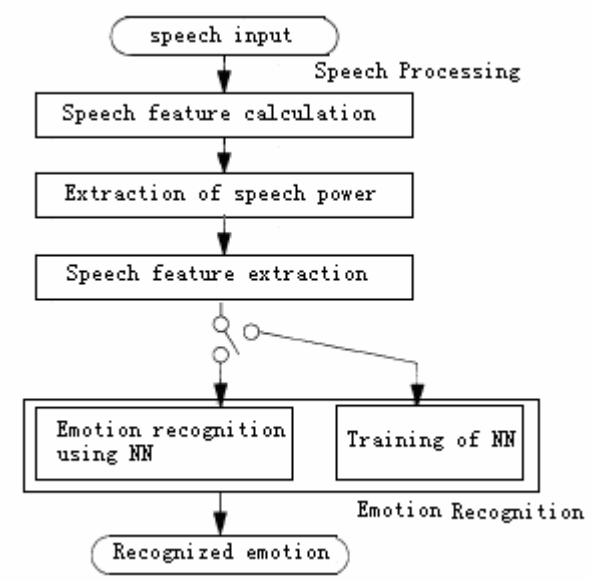

Fig. 2. The processing flow

A. *The choice of speech features*

There are two main types of speech features: phonetic features and prosodic features. Phonetic features deal with the types of sounds involved in speech, such as vowels and consonants, and their pronunciation. Prosodic features deal with the more musical aspects of speech, such as rising or falling tones and accents or stresses.

Speech understanding traditionally uses only phonetic features. One viewpoint on emotion recognition states that, in the light of this, emotion recognition should focus on only prosodic features. Another view is that prosodic features and phonetic features are intertwined in expressing emotion, and that it is impossible to express emotion using only prosodic features. Following the latter perspective, we examine both prosodic features and phonetic features in this study.

The features that we examine are

Prosodic features:

Speech power (P)

Pitch (p)

Phonetic features:

12 LPC parameters ($c_1, c_2, c_3 \cdots c_{12}$)

Delta LPC parameter (d)

Here, the LPC (Linear Predictive Coding) parameters are obtained in the LPC analysis of the speech, and the delta LPC parameter calculated from the LPC parameters expresses a time variable feature of the speech spectrum.

B. Speech feature extraction

The speech features must be extracted from each utterance for the emotion recognition training or testing. Figure3 illustrates this process.

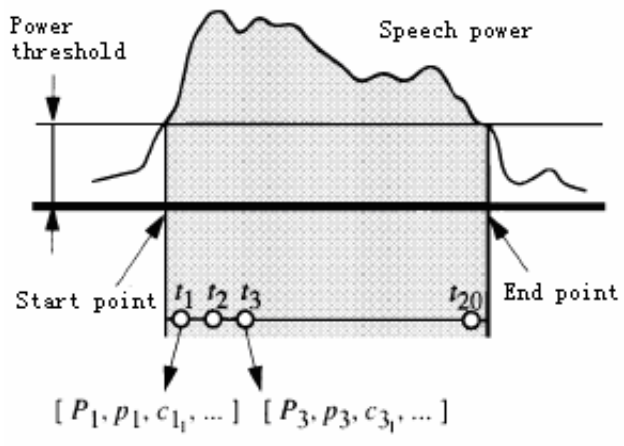

Fig. 3. Speech feature extraction

The first step is to determine the beginning add end points of an utterance. Speech power is compared with a predetermined power threshold. When the speech power first exceeds this threshold for several consecutive frames, this marks the beginning of the utterance. The utterance ends when the speech power drops below this threshold for several frames. By requiring that the speech power be above or below the threshold for several frames before beginning or ending the utterance, short fluctuations due to random noise are neglected.

Once the start and end points of an utterance have been determined, the utterance is divided into 20 intervals of an equal length in time. Let these 20 intervals be expressed as the vectors $f_1, f_2, f_3 \cdots f_{20}$.

Each of these 20 vectors in turn consists of the 15 speech feature parameters for that interval, $(P, p, c_1, c_2, \cdots c_{12}, d)$. The 20 speech vectors of 15 parameters compose a feature matrix of 300 values, or can be flattened out into a 300-length feature vector:

$$F_V = [f_1, f_2, f_3 \cdots f_{20}] \quad (1)$$

$$f_n = [P_n, p_n, c_{n1}, c_{n2}, \cdots c_{c12}, d_n] \quad (2)$$

This feature vector F_v is then used as input for the emotion recognition stage.

C. Neural Network Architecture

The emotion recognition stage of the processing flow is shown in Fig.4. The network is actually composed of eight sub-neural networks, with one network for each of the eight emotions that are examined. This type of network is called a One-Class-in-One Neural network (OCON). The feature vector (300 speech features for each utterance) is input into each of the eight sub-neural networks. The output from each sub-network is a value ($v_1, v_2, v_3 \cdots v_8$), representing the likelihood that the utterance corresponds to that sub-network's emotion. Decision logic selects the best emotion based on these values.

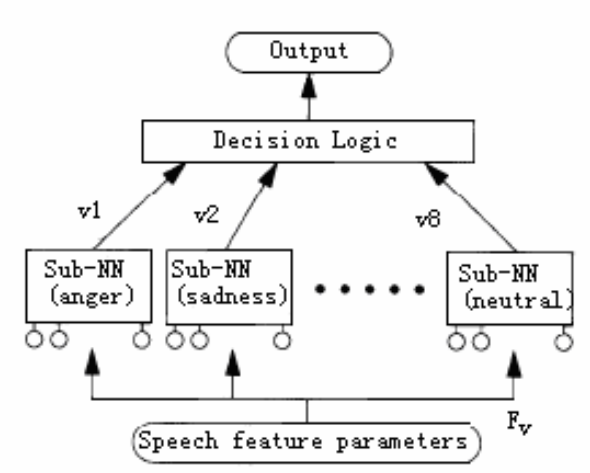

Fig. 4 Emotion recognition stage.

V. TRACKING THE CHANGE OF EMOTION STATE

In the E-Learning environment, tracking psychological state change of cognition process is always neglected. When the learners cannot understand and accept the learning contents, they will generate worried, disagreeable, fearsome emotion. Therefore, it is necessary to track emotion state in the E-Learning environment. Thus, when learner's state is not fine, the learning process can also carry on under healthy state though adjusting teaching strategies and intervening emotion.

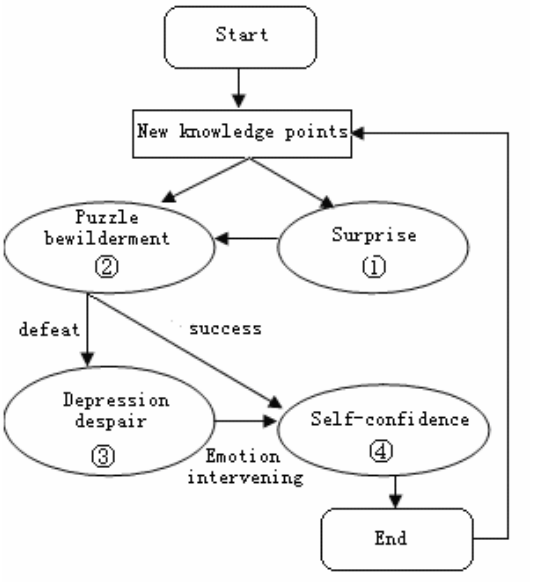

Fig. 5 The change of emotion state

The change of learners' emotion is more complex in the learning process. The learner contacts with a new knowledge point, two kinds of attitudes such as interested state ① and indifferent state may appear. Interested state is to explain the change process of learner's emotion, and learner's emotion state is divided into four categories approximately in the paper. The change of emotion state is Fig.5
①surprise ②puzzle, bewilderment ③depression, despair ④ self-confidence

With the learning process advancing, learner's emotion is changed. For instance, the learner has an idea to solve this problem. If the learner defeated repeatedly, he will suspect himself and changes into the state ③, this kind of state is disadvantageous to learning. The system should apperceive this kind of emotional change and carry on emotion intervening, which let him turn to self-confident state ④. When he contacts with a new knowledge point, the emotion turns to state ① and ② too.

VI. Adjusting Teaching Strategies

Teaching strategies are adjusted according to student's emotion dynamically though production rules. We neglect other influencing factors and take the learning style as the example.

The production rules: IF condition THEN movement

IF < initial emotion value>= confidence AND < new emotion value>= digest AND < initial learning style >=Text

THEN <Entering video learning>

IF < initial emotion value>= confidence AND < new emotion value>= puzzle AND < initial learning style >=Text

THEN <Entering dynamic interaction learning>

VII. Conclusion

Aiming at emotion deficiency in present E-Learning system, a lot of negative effects were analyzed and corresponding countermeasures were proposed. Basing on it, we combined affective computing with the traditional E-Learning system. The model of E-Learning system based on affective computing was constructed by using speech emotion, which took speech feature as input data. Our simulation experiment results showed that neural networks was effective in emotion recognition, and we achieve a recognition rate of approximately 50% when testing eight emotions .besides, other key techniques of realizing the system such as tracking the change of emotion state and adjusting teaching strategies were also introduced.

References

[1] He Kekang. E-Learning Essence- information technology into curriculum. E-education Research, 2002,105(1):3-4.
[2] Wang Jijun. Emotion deficiency and compensation in distance learning. Chinese network education, 2005
[3] Ma Xirong. Research on harmonious man-machine interaction model. Computer science, 2005.
[4] R.W.Picard. Affective Computing, Challenges. Cambridge. International Journal of Human Computer studies, 2003,59(1): 55-64.
[5] Murray IR, Arnott JL. Toward the simulation of emotion in synthetic speech, A review of the literature on human vocal emotion. J Acoustical Soc Am, 1993,93(2):1097–1108.
[6] Fushun Guo, Xuehai Zhang, Tuian Cheng.The realization scheme of Mozziaconacci S. Pitch variations and emotions in speech. Proceedings of ICPhS 95, 1995
[7] Klasmeyer G, Sendlneier WF. Objective voice parameters to characterize the emotional content in speech. Proceedings of ICPhS 95, 1995.
[8] Q.Cai, A.Mitche and J.K.Aggarwal .Track Human Motion in an Indoor Environment. Proceedings of 2nd International Conference of Image, 2003.