Research on Speech Emotion Recognition System in E-Learning

Ken Chen¹, Guangxue Yue², Fei Yu^{1,3}, Yue Shen¹, and Aiqing Zhu⁴

¹ School of Computer & Information Engineering, Hunan Agricultural University,
Changsha, 410128, China

{shenyue, yufei}@hunau.edu.cn

² College of Information Engineering, Jiaxing University, 314000 China
guangxueyue@yahoo.com.cn

³ Jiangsu Provincial Key Laboratory of Computer Information Processing Technology,
Province, Suzhou University, Suzhou, 2150063, China
hunanyufei@126.com

⁴ College of Urban and Environment Science, Central China Normal
University, Wuhan, China, 430079
zhuaiqin1982@yahoo.com.cn

Abstract. 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.

Keywords: E-learning, Affective computing, Speech emotion, Teaching strategies.

1 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

2 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.

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

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.

3 Emotion Recognition Module

Consciously expressed emotions are easier for humans to recognize, 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.

3.1 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

Neutrality, joy, boredom, Sadness, anger, fear, indignation[3]

Neutrality, happiness, sadness, anger, fear, boredom, disgust[4]

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.

3.2 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.

4 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.

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.

①surprise ②puzzle, bewilderment ③depression, despair ④self-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.

5 Conclusion

The results obtained in this study demonstrate that emotion recognition in speech is feasible, and that neural networks are well suited for this task.

There is still more work that needs to be done in the field of emotion recognition in speech. An examination of the speech features used in this study may allow the number of features to be reduced. In addition, further trials with different topologies of neural networks may also help improve performance. And multi-module emotion recognition including facial, speech, and other features such as gesture will be studied. I wish that this article's work could give some references to certain people.

References

1. Kekang H.:E-Learning Essence- information technology into curriculum. E-education Research(2002)3-4
2. Wang J.J.: Emotion deficiency and compensation in distance learning. Chinese network education(2005) 117-120.
3. Picard R.W.: Affective Computing, Challenges, Cambridge. International Journal of Human Computer studies(2003)55-64.
4. Klasmeyer G, Sendlneier W.: Objective voice parameters to characterize the emotional content in speech. Proceedings of Processing(1995)