An XML-based Implementation of Multimodal Affective Annotation

Fan Xia[1], Hong Wang[1], Xiaolan Fu[2], and Jiaying Zhao[2]

[1] State Key Laboratory of Intelligent Technology and Systems
Department of Computer Science and Technology, Tsinghua University
Beijing, 100084, P.R.China
[2] State Key Laboratory of Brain and Cognitive Science
Institute of Psychology, Chinese Academy of Sciences
Beijing, 100101, P.R.China

Abstract. In simple cases, affective computing is a computational device recognizing and acting upon the emotions of its user or having (or simulating having) emotions of its own in complex cases. Multimodal technology is currently one of the hottest focuses in affective computing research. However, the lack of a large-scale multimodal database limits the research to some respective and scattered fields, such as affective recognition by video or by audio. This paper describes the development and implementation of an XML-based multimodal affective annotation system which is called MAAS (Multimodal Affective Annotation System). MAAS contains a hierarchical affective annotation model based on the 3-dimensional affect space derived from Mehrabian's PAD temperament scale. The final annotation file is formed in XML format in order to interchange the resources with other research groups conveniently.

1 Introduction

In psychology, affect is an emotion or subjectively experienced feeling, or the involvement of such processes in a psychological system or theory, different from mood which is more sustained [11]. Affective computing is an area of computer science research aimed to simulate emotional processes or make use of human emotion in human-computer interaction. In simple cases, affective computing is a computational device recognizing and acting upon the emotions of its user or having (or simulating having) emotions of its own in complex cases.

Multimodal technology is a key issue of emotion recognizing and understanding in affective computing research. The development of multimodal database has attracted the attention of some researchers in the past few years. This is motivated by the fact that the lack of a large-scale multimodal database limits the research of affective computing. Our work is a sub-project of the RACTA (Research on Affective Computing Theory and Approach), a key research project of NSFC (National Nature Science Foundation of China), with the purpose of establishing a large-scale affect database. In this paper, we present an approach

to annotate human emotions precisely and conveniently, as well as its implementations.

In achieving our goal, a series of experiments have been designed. Participants were stimulated in order to induce some real affect. A few problems arised such as how to evaluate the stimulation materials and how to annotate the various affect precisely and conveniently. In the following sections, we will present an annotation system named MAAS with a useful toolkit to solve these problems.

2 Approaches of affective evaluation

There is no exact definition of affect in psychology. Nonetheless, various psychologists have attempted to provide temporary definitions, among which Leverson's theory is the most acceptable - affect is "a transient psychological-physiological behavior, which represents the model that the organism adopted to adapt the variety environment" [1].

Adopting a universal evaluation standard will be greatly helpful to allow the easy interchange of annotation files and the sharing of annotation model among different research groups. There are two different theoretical approaches to the study of affect: categorical and dimensional.

The categorical approach considers that there are several basic types of affect; each type has its special experiential characters, physiological arousal pattern and performance pattern. The different combinations of the basic affect form the entire human affect [2]. According to this approach, commonly proposed members include happiness, surprise, fear, sadness, anger, disgust, contempt, shame, and guilt [12].

Another one is the dimensional approach. This theory shows that the human affect consists of several dimensions. The dimensional theorists consider affect as gradual and smooth transformation. In this approach, similarity and difference between two types of affect is expressed as the distance in the dimensional space. The most acceptable dimensional pattern is a two-dimension space can be described as below: (a) valence or in other word hedonic tone, this dimension is based on negative-positive affect's separation-activation; (b) arousal or activation, this dimension indicates the activated power intensity associated with the affect status. However, this approach doesn't hold good in all cases. For example, it fails to specify the difference between fear and anger [6].

In MAAS, Mehrabian's PAD temperament scale is used [5] [6] [7].The PAD scale determines emotions using a three-dimensional emotion space representing pleasure-displeasure, arousal-nonarousal and dominance-submissiveness. Pleasure-displeasure represents the positive or negative quality of the emotion; arousal-nonarousal refers to a combination of physical activity and mental alertness; and dominance-submissiveness is defined in terms of levels of control [5]. According to Mehrabian, every possible human emotion can be represented as definite points in the three dimensions. The following emotions exist as points in that space: angry, bored, curious, dignified, elated, hungry, inhibited, loved, puzzled, sleepy,

unconcerned, and violent [5], as shown in Figure 1. Violent, for instance, represents a displeased, highly aroused, and highly dominant emotional state.

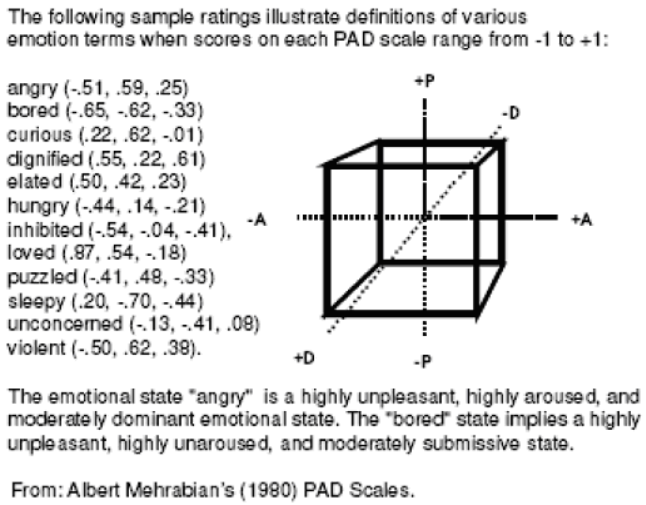

Fig. 1. 3-dimensional affect space [5] [10]

Furthermore, the work of Yuxia Huang and her colleagues based on IAPS (International Affective Picture System) [3] shows that basic affect has a corresponding position in the dimensional affective space. For example, the affect of happiness always has a corresponding position on the aroused and positive dimensions, and the affect of sadness on the nonaroused and negative dimensions. Through the research they also discover that the dimensional affective space has no absolute relation with the basic affect, which means that there is no precise transformation between basic affect and the position in dimensional affect space.

Further analysis shows the correspondence between the two approaches [3]. In some cases, they are similar in some ways. Like the categorical approach, the dimensional approach also includes the existence of basic affect. Moreover, the multidimensional and dynamical nature of affect should be considered while describing the complicated human affect

3 MAAS: an XML-based affective annotation system

3.1 Issues in the design of a multimodal affective annotation system

An effective annotation approach needs to be able to represent the entire human affect. The representation should be precise and convenient for researchers. MAAS assumes that human affect consists of three parts: subjective experience,

performance or in another word expression and physiological arousal. The system should be useful when annotating a person's affect in a particular scene, describing his or her performance and tracking the physiological data.

The implementation of multimodal annotation is another important issue that needs to be considered. Due to the complexity and multi-causality of affect, any single subject like video is not enough for affect recognition. Thus we must integrate the multimodal data: physiological, audio-visual in the annotation process in MAAS.

3.2 The hierarchical annotation model

Considering the convenience of annotation process and the representation of various human affect, MAAS employs a hierarchical structure to describe both affect information and other data sources. There are mainly two top layers in MAAS's hierarchical structure: the affect annotation layer and the data layer.

First in affect annotation MAAS assumes that there are ten basic categories of affect: exuberant, bored, dependent, disdainful, relaxed, anxious, docile, hostile, fearful and disgusted. Each affect in these categories has a value in the form of (P, A, D) in the three dimensional affective space. For example, anger is labeled with a value (-.51,.59,.25) and curious with a value (.22,.62,-.01). During the annotation process, MAAS divides the annotation into two parts: the person's self-evaluation and an objective-evaluation by others. Self-evaluation and objective-evaluation are respectively corresponded to the features of subjective experience and performance of affect. In the following steps, some scattered elements can be added to the annotation, such as the mode which represents how the person performs his or her affect, by expression or gesture; the scene which represents the description of the experiment environment and the person's background, etc..

In the data layer there are now three sub-layers: video, audio and physiology. The data recorded during the experiments will be stored into the data layer in required format. The notable key is that data layer represents not only the physiological arousal but also multimodal factors related to the affect

3.3 MAA: an XML-based affective annotation tool

A demo picture of MAAS's toolkit is given in Figure 2. It shows how to use the MAAS's toolkit to annotate a person's affect. The main window contains a video-window, a wave-window and a thumbnail-window. Before the annotation process, the video and the audio source should be specified respectively. In the center of Figure 2, there is a simple XML-based annotation content editor, which appears when a segment is specified as Figure 3 shows. The MAAS toolkit will automatically track the data to the annotation content editor and save it to the result files.

When the large-scale affective resource database is established, the most imperative problem is how to extract a particular piece of data from a vast database rapidly and efficiently. MAAS's toolkit provides an extract wizard for

Fig. 2. MAAS-toolkit's main window

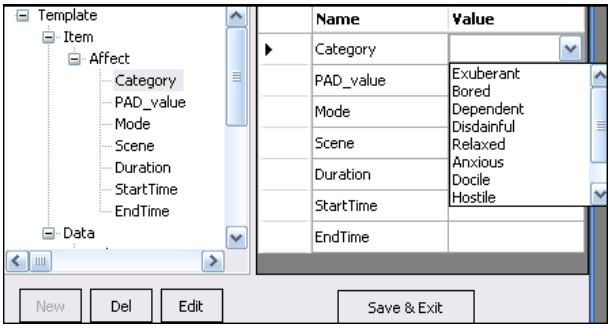

Fig. 3. Annotation content editor

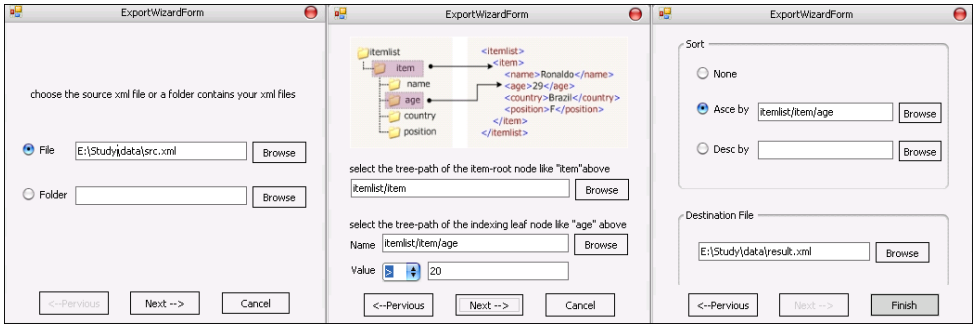

Fig. 4. Extract specified item from annotation file(s)

researchers. This tool in MAAS can extract the data with features specified by user from a great capacity of annotation files and sort the result in several ways, as Figure 4 shows. For example, users can extract only the relaxed affect in a series of experiment order by experiment time.

Here is a simple example of the annotation file:

```
<?xml version="1.0" encoding="utf-8"?>
<affect>
    <Category>Relaxed</Category>
    <PAD_value>(.22, .51, -.10)</PAD_value>
    <mode>expression</mode>
    <scene>in the dormitory</scene>
    <Duration unit=ms>1226</Duration>
    <Other />
</affect>

<data>
    <video> </video>
    <audio> </audio>
    <physiology>    </physiology>
</data>
```

4 Conclusion

A multimodal affective annotation system named MAAS is described in this paper. MAAS and its toolkit are convenient for researchers to annotate common affect and track the data collected in the experiment. It takes the most popular and standard XML format to store the annotation results, which is easy for users to read and to modify. The hierarchical annotation structure which MAAS adopts is flexible enough to expand to be compliant with other particular research.

Acknowledgements

This work is a part of the Research on Affective Computing Theory and Approach, which is a key research project of National Natural Science Foundation of China (NSFC). We would like to thank Prof.Lianhong Cai and other members of the Project Group for their valuable discussion and help.

References

1. R.W.Leverson: The Nature of Emotions [M]. New York, OxfordUniversity Press. (1994) 123-126

2. P.Ekman: An argument for basic emotions [J]. Cognition and Emotion. **6** (1992) 169-200
3. Huang, Y., Luo, J.: Experimental Research based on IAPS in China. Psychology and Sanitation Journal of China. **18** (2004) 631-634
4. J.Broekens, D.DeGroot: Scalable and Flexible Appraisal Models for Virtual Agents. CGAIDE. (2004).
5. A.Mehrabian: Basic Dimensions for a General Psychological Theory. Cambridge, OGH Publishers. (1980).
6. Li, X., Zhou, H., Song, S., Ran, T.,Fu, X.: The reliability and validity of the Chinese version of abbreviated PAD emotion scales. In: Proceedings of the First International Conference on Affective Computing and Intelligent Interaction. (2005)
7. Zhao, J., Fu, X.: An Evaluation-based Hybrid Annotation Scheme for Emotion Using Multimodal Data in Task-oriented Conditions. Submitted to ACM CHI 2006
8. R.W.Picard: Affective Computing. MIT Press, London, England. (1997)
9. E.C.Chang, L.J.Sanna: Optimism, Pessimism, and Positive and Negative Affectivity in Middle-Aged Adults: A Test of a Cognitive-Affective Model of Psychological Adjustment. Psychology and Aging. **16** (2001) 524-531
10. G.A.Hollinger, Z.A.Pezzementi, B.Mitchell, A.Flurie, B.A.Maxwell: Design of a Social Mobile Robot Using Emotion-Based Decision Mechanisms. http://www.engin.swarthmore.edu/ ghollin1/e28/lab5/SocialRobot.pdf
11. http://www.wikipedia.org/wiki/affect
12. M.Pagel: The Evolution of Human Emotions. The Oxford Encyclopedia of Evolution. **1** (2002) 296-299