MULTI-ATTRIBUTE DECISION MAKING FOR AFFECTIVE BI-MODAL INTERACTION IN MOBILE DEVICES

Efthymios Alepis, Maria Virvou
Department of Informatics, University of Piraeus
80 Karaoli & Dimitriou St., 18534, Piraeus, Greece

Katerina Kabassi
Department of Ecology and the Environment, Technological Educational Institute of the Ionian Islands
2 Kalvou Sq., 29100 Zakynthos, Greece

Keywords: Mobile devices, m-learning, affective interaction, bi-modal interaction, multi-attribute decision making theories.

Abstract: This paper presents how multi attributes decision making is used for affective interaction in mobile devices. The system bases its inferences about users' emotions on user input evidence from the keyboard and the microphone of the mobile device. The actual combination of evidence from these two modes of interaction has been performed based on an innovative inference mechanism for emotions and a multi-attribute decision making theory. The mechanism that integrates the inferences form the two modes has been based on the results of two empirical studies, with the participation of human experts and possible users of the system.

1 INTRODUCTION

In the fast pace of modern life, students and instructors would appreciate using constructively some spare time. They may have to work on lessons at any place, even when away from offices, classrooms and labs where computers are usually located. At the current state, there are not many mature mobile tutoring systems since the technology of mobile computing is quite recent and has not yet been used to the extent that it could. However, there have been quite a lot of primary attempts to incorporate mobile features to this kind of educational technology and the results so far confirm the great potential of this incorporation. Moreover, in many cases it would be extremely useful to have such facilities in handheld devices, such as mobile phones rather than desktop or portable computers so that additional assets may be gained. Such assets include device independence as well as more independence with respect to time and place in comparison with web-based education using standard PCs.

However, different problems may occur during people's interaction with mobile devices. This especially is the case of novice users who find such an interaction frustrating and difficult. A remedy to such problem may be given by providing adaptive interaction based on the user's emotional state. For this purpose, affective computing may be used. Regardless of the various emotional paradigms, neurologists/psychologists have made progress in demonstrating that emotion is at least as and perhaps even more important than reason in the process of decision making and action deciding (Leon et al., 2007). Moreover, the way people feel may play an important role in their cognitive processes as well (Goleman, 1995).

Indeed, Picard points out that one of the major challenges in affective computing is to try to improve the accuracy of recognizing people's emotions (Picard, 2003). Ideally, evidence from many modes of interaction should be combined by a computer system so that it can generate as valid hypotheses as possible about users' emotions. It is hoped that the multimodal approach may provide not only better performance, but also more robustness (Pantic & Rothkrantz, 2003).

In previous work, the authors of this paper have implemented and evaluated with quite satisfactory results emotion recognition systems, incorporated in educational software applications for computers (Alepis et al. 2007). As a next step we have

Alepis E., Virvou M. and Kabassi K. (2008).
MULTI-ATTRIBUTE DECISION MAKING FOR AFFECTIVE BI-MODAL INTERACTION IN MOBILE DEVICES.
In *Proceedings of the International Conference on Signal Processing and Multimedia Applications*, pages 376-379
DOI: 10.5220/0001935103760379

extended our affective educational system by providing mobile interaction between the users and handheld device. The system is based on mobile technology and incorporates the quite recent theory of Affective Computing.

In view of the above, in this paper we describe a novel mobile educational system that incorporates bi-modal emotion recognition. The proposed system collects evidence from the two modes of interaction and analyses them in terms of some attributes for emotion recognition. Finally the system associates the users' input data through a multi-attribute model and makes final assumptions about the user's emotional state. For the effective application of the multi-attribute decision making model, we conducted an empirical study with the participation of human experts as well as possible users of the system.

2 EMPIRICAL STUDY FOR ATTRIBUTES DETERMINATION

In order to collect evidence about which information could be used for emotion recognition, we conducted an empirical study.

2.1 Settings of the Experiment

The empirical study that we have conducted concerns the audio-lingual emotion recognition, as well as the recognition of emotions through keyboard evidence. The audio-lingual mode of interaction is based on using a mobile device's microphone as input device. The empirical study aimed at identifying common user reactions that express user feelings while they interact with mobile devices. As a next step, we associated these reactions with particular feelings.

Individuals' behaviour while doing something may be affected by several factors related to their personality, age, experience, etc. Therefore, the empirical study involved a total number of 100 male and female users of various educational backgrounds, ages and levels of familiarity with computers.

The participants were asked to use a mobile educational application, which incorporated a user monitoring component. The user monitoring component that we have used can be incorporated in any application, since it works in the background recording each user's input actions. Part of the interaction included knowledge tests, while participants were asked to use oral interaction via their mobile device's microphone. Our aim was not to test the participants' knowledge skills, but to record their oral and written behaviour. Thus, the educational application incorporated the monitoring module that was running unnoticeably in the background. Moreover, users were also video-taped while they interacted with the mobile application.

After completing the interaction with the educational application, participants were asked to watch the video clips concerning exclusively their personal interaction and to determine in which situations they where experiencing changes in their emotional state.

As the next step, the collected transcripts were given to 20 human expert-observers who were asked to perform audio emotion recognition with regard to the six emotional states, namely happiness, sadness, surprise, anger, disgust and neutral. All human expert-observers possessed a first and/or higher degree in Psychology and, to analyze the data corresponding to the audio-lingual input only, they were asked to listen to the video tapes without seeing them. They were also given what the user had said in printed form from the computer audio recorder. The human expert-observers were asked to justify the recognition of an emotion by indicating the weights of the attributes that they had used in terms of specific words and exclamations, pitch of voice and changes in the volume of speech.

2.2 Analysis of the Results

The analysis of the data collected by both the human experts and the monitoring component, revealed some statistical results that associated user input actions through the mobile keyboard and microphone with possible emotional states of the users. More specifically, considering the keyboard we have the following categories of user actions: a) user types normally b) user types quickly (speed higher than the usual speed of the particular user) c) user types slowly (speed lower than the usual speed of the particular user) d) user uses the "delete" key of his/her mobile device often e) user presses unrelated keys on the keyboard f) user does not use the keyboard.

Considering the users' basic input actions through the mobile device's microphone we have 7 cases: a) user speaks using strong language b) users uses exclamations c) user speaks with a high voice volume (higher than the average recorded level) d) user speaks with a low voice volume (low than the average recorded level) e) user speaks in a normal voice volume f) user speaks words from a specific list of words showing an emotion g) user does not say anything.

Therefore, each moment the system records a vector of input actions through the keyboard (k1, k2, k3, k4, k5, k6) and a vector of input actions through the microphone (m1, m2, m3, m4, m5, m6, m7).

All the above mentioned attributes are used as Boolean variables. In each moment the system takes data from the bi-modal interface and translates them in terms of keyboard and microphone actions. If an action has occurred the corresponding attribute takes the value 1, otherwise its value is set to 0. Therefore, for a user that speaks with a high voice volume and types quickly the two vectors that are recorded by the system are: k= (0, 1, 0, 0, 0, 0) and m= (0, 0, 1, 0, 0, 0, 0). These data are further processed by the multi-attribute model for determining the emotion of the user.

3 EMPIRICAL STUDY FOR WEIGHT CALCULATION

The previous empirical study revealed the attributes that are taken into account when evaluating different emotions. However, these attributes are not equally important for evaluating different emotions. For this purpose, the human experts who participated in the first empirical study and selected the final set of attributes were also asked to rank the 13 attributes with respect to how important they are in their reasoning process.

Human experts resulted that one input action does not have the same weight while evaluating different emotions. Therefore, the weights of the attributes (input actions) were calculated for some stereotypes of different emotions were designed.

Therefore, each human expert was asked to share 21 points into the 6 different attributes with respect to the keyboard input for each emotion.

As soon as the scores of all human experts were collected, they were used to calculate the weights of attributes. The scores assigned to each attribute by all human experts were summed up and then divided to the sum of scores of all attributes (21 points assigned to all attributes by each human expert * 20 human experts = 420 points assigned to all attributes by all human experts). In this way the sum of all weights could be equal to 1.

As a result, there was a set of weights for the attributes that correspond to the keyboards' input actions for each different emotion.

Then each human expert was asked to share 28 points into the 7 different attributes with respect to the microphone input for each emotion. As soon as the scores of all human experts were collected, they were used to calculate the weights of attributes. The scores assigned to each attribute by all human experts were summed up and then divided to the sum of scores of all attributes (28 points assigned to all attributes by each human expert * 20 human experts = 560 points assigned to all attributes by all human experts). In this way the sum of all weights could be equal to 1.

As a result, there was a set of weights for the attributes that correspond to the microphones' input actions for each different emotion.

4 APPLICATION OF THE MULTI-ATTRIBUTES MODEL

For the evaluation of each alternative emotion the system uses SAW (Fishburn, 1967, Hwang & Yoon, 1981) for a particular category of users. According to SAW, the multi-attribute utility function for each emotion in each mode is estimated as a linear combination of the values of the attributes that correspond to that mode.

The SAW approach consists of translating a decision problem into the optimisation of some multi-attribute utility function U defined on A. The decision maker estimates the value of function $U(X_j)$ for every alternative X_j and selects the one with the highest value. The multi-attribute utility function U can be calculated in the SAW method as a linear combination of the values of the n attributes:

$$U(X_j) = \sum_{i=1}^{n} w_i x_{ij}$$

where X_j is one alternative and x_{ij} is the value of the i attribute for the X_j alternative.

In view of the above, for the evaluation of each emotion taking into account the information provided by the keyboard is done using formula 1.

$$em_{ke_1} = w_{e_1k1}k_1 + w_{e_1k2}k_2 + w_{e_1k3}k_3 + w_{e_1k4}k_4 + w_{e_1k5}k_5 + w_{e_1k6}k_6 \quad (1)$$

Similarly, for the evaluation of each emotion taking into account the information provided by the other mode (microphone) is done using formula 2.

$$em_{me_1} = w_{e_1m1}m_1 + w_{e_1m2}m_2 + w_{e_1m3}m_3 + w_{e_1m4}m_4 + w_{e_1m5}m_5 + w_{e_1m6}m_6 + w_{e_1m7}m_7 \quad (2)$$

em_{ke_1} is the probability that an emotion has occurred based on the keyboard actions and em_{me_1} is the probability that refers to an emotional state using the users' input from the microphone em_{ke_1} and em_{me_1} take their values in [0,1].

In formula 1 the *k*'s from *k1* to *k6* refer to the six attributes that correspond to the keyboard. In formula 2 the *m*'s from *m1* to *m7* refer to the seven attributes that correspond to the microphone. The *w*'s represent the weights. These weights correspond to a specific emotion and to a specific input action and were calculated in the previous empirical study.

In cases where both modals (keyboard and microphone) indicate the same emotion then the probability that this emotion has occurred increases significantly. Otherwise, the mean of the values that have occurred by the evaluation of each emotion using formulae 1 and 2 is calculated.

$$\boxed{\frac{em_{ke_1} + em_{me_1}}{2}}$$

The system compares the values from all the different emotions and selects the one with the highest value of the multi-attribute utility function. The emotion that maximises this function is selected as the user's emotion.

5 CONCLUSIONS AND FUTURE WORK

In this paper we have described how multi-attribute decision making could be used for affective interaction in mobile devices. More specifically, we describe the implementation of an affective educational application for mobile devices that recognizes students' emotions based on their keyboard and microphone actions. The educational application employs a bi-modal user interface.

A similar approach to the proposed one has previously been used in a learning environment operating over the web (Alepis et al., 2007). However, the main difference of the approach described in this paper is that the interaction provided by mobile devices differentiates from the human-computer interaction in many ways. The keyboard and the screen are very different as well as the places where a user may interact with them. A user may interact with a mobile device in the places that s/he can interact with a PC as well as other places, such as a station, a bus or the beach. In such places the users' mood may be affected by other factors. Therefore, the need for affective interaction in mobile devices may be even more essential than in normal computers.

It is among our future plans to incorporate user modelling techniques such as stereotypes in combination with the multi-attribute decision making in order to personalise interaction with each individual user interacting with the mobile device. Furthermore, we intend to enrich multi-modal interaction by incorporating a third mode of interaction, visual this time (Stathopoulou & Tsihrintzis, 2005).

ACKNOWLEDGEMENTS

Support for this work was provided by the General Secretariat of Research and Technology, Greece, under the auspices of the PENED-2003 program. Travel funds to present this work were provided by the University of Piraeus Research Center.

REFERENCES

Alepis, E., Virvou, M., Kabassi, K., 2007. Knowledge Engineering for Affective Bi-modal Human-Computer Interaction, SIGMAP.

Fishburn, P.C., 1967. Additive Utilities with Incomplete Product Set: Applications to Priorities and Assignments, Operations Research.

Goleman, D., 1995. Emotional Intelligence, Bantam Books, New York .

Hwang, C.L., Yoon, K., 1981. Multiple Attribute Decision Making: Methods and Applications. Lecture Notes in Economics and Mathematical Systems 186, Springer, Berlin/Heidelberg/New York.

Leon, E., Clarke, G., Gallaghan, V., Sepulveda, F., 2007. A user-independent real-time emotion recognition system for software agents in domestic environments. Engineering applications of artificial intelligence, 20 (3): 337-345.

Pantic, M., Rothkrantz, L.J.M., 2003. Toward an affect-sensitive multimodal human-cumputer interaction. Vol. 91, Proceedings of the IEEE, Institute of Electrical and Electronics Engineers, pp. 1370-1390.

Picard, R.W., 2003. Affective Computing: Challenges. Int. Journal of Human-Computer Studies, Vol. 59, Issues 1-2, pp. 55-64.

Stathopoulou, I.O., Tsihrintzis, G.A., 2005. Detection and Expression Classification System for Face Images (FADECS), *IEEE Workshop on Signal Processing Systems*, Athens, Greece.