Implementation of Emotional-Aware Computer Systems Using Typical Input Devices

Kaveh Bakhtiyari[1,2], Mona Taghavi[1], and Hafizah Husain[1]

[1] Department of Electrical, Electronics and Systems Engineering,
Faculty of Engineering and Built Environment,
Universiti Kebangsaan Malaysia (The National University of Malaysia)
43600 UKM Bangi, Selangor Darul Ehsan, Malaysia
[2] Department of Computer & Cognitive Science, Faculty of Engineering
University of Duisburg-Essen
47048 Duisburg, North Rhine-Westphalia (NRW), Germany
academic@bakhtiyari.com, mona@siswa.ukm.edu.my,
hafizah@eng.ukm.my

Abstract. Emotions play an important role in human interactions. Human Emotions Recognition (HER - Affective Computing) is an innovative method for detecting user's emotions to determine proper responses and recommendations in Human-Computer Interaction (HCI). This paper discusses an intelligent approach to recognize human emotions by using the usual input devices such as keyboard, mouse and touch screen displays. This research is compared with the other usual methods like processing the facial expressions, human voice, body gestures and digital signal processing in Electroencephalography (EEG) machines for an emotional-aware system. The Emotional Intelligence system is trained in a supervised mode by Artificial Neural Network (ANN) and Support Vector Machine (SVM) techniques. The result shows 93.20% in accuracy which is around 5% more than the existing methods. It is a significant contribution to show new directions of future research in this topical area of emotion recognition, which is useful in recommender systems.

Keywords: human emotion recognition, keyboard keystroke dynamics, mouse movement, touch-screen, human computer interaction, affective computing.

1 Introduction

Emotional Intelligence can recognize human emotions and respond to the user accordingly. Human Emotions Recognition (HER) might be used in different categories such as e-learning, game, adaptive user interfaces, etc. Emotion is one of the features which promotes human-computer interactions, and plays a significant role in making trust. If a recommender system can recognize the user's emotion, it would produce responses relevant to the user's emotional state. Consequently, it attracts the user's attention and loyalty. Emotion is a way of interaction which is about the message owner features. Emotions are discussed by three parameters. The first

N.T. Nguyen et al. (Eds.): ACIIDS 2014, Part I, LNAI 8397, pp. 364–374, 2014.

parameter is *Arousal* which shows the energy of feeling. The second parameter is *Valence*. Valence presents whether the feeling is a pleasure (positive) or displeasure (negative) in case of the energy. The third is *Dominance* which shows the strength of the feelings.

A classification of emotions is proposed by Plutchik, which is used as a standard classification in 8 emotions of *Acceptance, Fear, Surprise, Sadness, Disgust, Anger, Anticipation, Joy* [1].

HER has been done by various methods and techniques to achieve this goal. However, there are some challenges in different areas, which make it an open research topic to work through [1]. The first challenge is achieving a higher accuracy in emotion recognition with a reliable precision (lower false positive rate) [2]. Still the available techniques are not reliable and accurate enough to be employed in real applications. New methods and hypotheses can be applied to gain better results with a higher performance, thus new techniques are being introduced. The second challenge is the real time processing [3-5].

This research tends to perform a solution for human emotions recognition to address those mentioned problems. We have analysed the users' inputs on common input devices such as keyboard keystroke dynamics, mouse movements, and touch-screen interactions. Chang has tracked the individual pattern of keyboard, mouse and mobile device usage [6]. He showed that the users' patterns are unique, and it can be applied for security purposes. In addition, a hybrid analysis by combination of few input devices tries to perform a better performance in HER.

2 Critiques

The first issue is the recognition accuracy. Only facial expression recognition could achieve the highest recognition accuracy of 90% lately in 2012 by using image processing techniques. The other methods have still less performance which are not reliable in business applications. Facial expression recognition gained one of the best accuracies in emotions recognition. However, for real time processing, it works worse than the other methods, because in fact image processing techniques are time and resource consuming. Natural Language Processing (NLP) and common devices can be used for real time applications, but the resulted recognition accuracies in recent research are not satisfying. Some methods such as using EEG machines' signals are expensive and still those machines are not available to be used in a daily usage. There are other methods for HER such as using a microphone, camera and other input devices. However because of the security and privacy issues, many computers may not use microphones and cameras. These challenges cause to have the limited number of applications for facial expression recognition, body gestures recognition and voice processing.

3 Methodology

This paper is presenting a methodology based on a software prototype which records the data from user's inputs on mouse, keyboard and touch screen. Following this method, a prototype application has been designed and developed to collect the required data from computer users' interactions. The keyboard keystroke dynamics, mouse movements and touch screen interactions of 50 users with various cultural backgrounds were collected while they were using the system. These users were mostly settled physically in Malaysia, Germany and Iran. Every 4 hours for a month, users were asked to answer a question about their current emotions. Then the collected data were used in RapidMiner to be trained using the SVM technique for classification. For evaluation of the mouse and touch-screen interaction, the methodology presented by Schuller [7] has been used which collects all the mouse movements and mouse keystrokes.

A key question at the beginning was the selection of appropriate emotions which in this study should be considered. First, the seven universal emotions by Paul Ekman [2] have been used as a basis. Then emotions were clustered, and all investigations in this work are concentrated on the following four emotion categories:

- Neutral (includes above all the emotion happiness and as perceived normal mood)
- Fright (Afraid) (includes above all helplessness, confusion and surprise)
- Sadness (primarily sadness, anger and resentment)
- Nervousness (including nervous, fatigue and light-headedness)

3.1 Keyboard

Keystroke dynamics are habitual rhythm patterns by way of typing a word [5]. It has three major parts as *representation*, *extraction* and *classification*. Representation shows the input values as the words. When the user is typing, actually he is representing his identification. The next step is features extraction that the system extracts and defines the features as a fingerprint and records them in a database. The last section is a classification that matches the extracted features of a new user with the existing features in the database to identify him/her. Now, this research is using the similar method but there are differences by using the novel training algorithms to identify the emotions rather than the identification of the users.

There are three major and important features in keystroke dynamics: 1) *Key down-to-down*; 2) *Key down-to-up*; and 3) *Key up-to-down*.

The Keystroke Features were selected from the timing differences of single keystrokes, digraphs (two-letter combinations) and tri-graphs (three-letter combinations).

3.2 Mouse

It seems reasonable to divide the mouse movements in two different sections. The first section is the movement of the mouse without using the left mouse button pressed. The second section is where the mouse button is pressed.

This curve is then transformed into a 2-dimensional coordinate system. The ideal line corresponds to the x-axis of the coordinate system, and therefore y-axis describes a measure of the local variation of the mouse movement from the ideal line. Since these distance values have lost the absolute commitment to its original screen position, it can already measure global properties of the local mouse movement to be made. For example, the sum over all possible distance values states how much the mouse was moving on entire place above or below the ideal line. The properties which were studied are:

- The length of the racing line from start to end point
- The sum over all distance values
- The zero crossings
- Maximum deviation of the values
- Average of the individual values
- Standard deviation
- Variance
- Correlation function of the curve
- First order and second derivatives
- Min. and max. of the values
- Average amount over all values
- Standard deviation,
- Variance
- Autocorrelation function

Time Properties

In parallel with the above discussed features, the time intervals, which register with the result of a new (x,y) point are analysed. It should not be forgotten that only a change in the x or y coordinate of a new data value is read. This elapsed time between two consecutive points together is not only the total time the mouse moves, but the specified values describe information about the movement individually. It explains the time between jerky and slow; and also it can be used to distinguish verse breaks in existing movements very well. However, a complete overview is firstly presented of all the examined given features in Figure 1. This figure shows a possible sequence of values of time intervals, from which the main features are very well seen. This figure presents the time between the pressed keys. For instance, the time between the first and second character is the minimum in comparison with the time between 19th and 20th characters which is the maximum spent time for typing 25 characters. This figure is only a demonstration of a sample registered mouse keystroke time.

It is similar to the local variation, made and analysed with a number of time delta values. Then first two statements about the time relationships are possible:

✓ Total time of motion by summing over all values
✓ Average time distance between two points or the average required time.

However, when a change occurs to the location coordinates of the mouse movements, averaging is performed on the “Standard deviation of individual values” and “Variance of these values”. Finally, the derived variables of mouse movements and keystrokes are:

- Correlation function
- First derivative
- Second derivative with the corresponding analysis

More precise statements are possible to be described. However, the formation of a distribution function of these values and the derived properties of this distribution function lead to the catalogue of the properties.

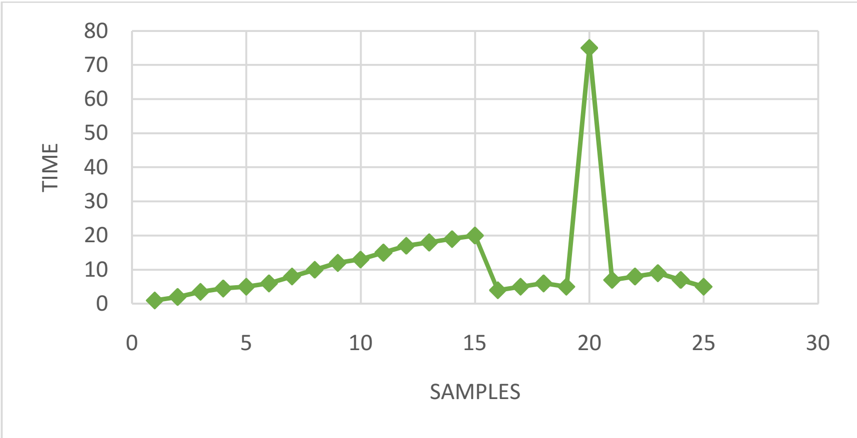

Fig. 1. Elapsed time between the modified coordinates of mouse movements

3.3 Touch Screen

The touch screen is only able to determine points in x, y and z coordinates. User interaction with touch screen monitors only result the changes in these coordinate values [7]. These values along with the time interval of changes are collected and prepared to measure the other important features such as velocity and the movement's details. Some other companies have introduced some new technologies which make the touch-screen monitor to react according to the user's eyes, hands and behaviours. All of these advanced technologies are the combination of image processing with the touch screen displays and AI techniques; and they are not directly related to touch-screen monitors. The most significant expansion was therefore to complement the additional available z-component (pressure strength), which has been evaluated in parallel. Thus, analogous reads the (x,y) coordinates of an initial set of z-values, where they open up a value range between 0 and 255. Straight from the emerging contours of the first and second derivatives as well as the correlation function, some additional values can be used to interpret better. These values include the average, minimum, maximum, standard deviation and variance of the first and second derivatives and the correlation function. By considering this number of features on touch-screen monitors, all the values are obtained from the Cartesian coordinate system. However, a three-dimensional coordinate space is presented. This can also offer a transformation in spherical coordinates (r, α, β).

4 Evaluation

This section demonstrates the diagnosis of the research based on the theories and methods of research methodology.

4.1 Evaluation Criteria

Evaluation of the system is based on the emotions recognition methods and machine learning techniques which have been used in the HER system. There are several criteria to evaluate and measure the performance of the system. These criteria are mainly composed of Classification/Recognition Accuracy, False Positive Rate, and Computational/Process Time. *Classification / Recognition Accuracy* shows how precise a system is able to recognize the emotions. It mostly focuses on the output of machine learning techniques. This criterion is measured by the machine learning classification methods. Generally, for this purpose, from 60% to 80% of the data would be trained, and then the rest of 20% to 40% of the remained data would be tested. *False Positive Alarm* gives some false classified emotions. These emotions are recognized but they are not matched with the actual recorded emotions. *Computational Time* is a classification procedure time to be applied to the collected data set. Different classifiers follow different algorithms, and they have different time complexities.

4.2 Data Analysis

Keyboard

The recognition performance is determined by using Support Vector Machine (SVM) as a classifier in term of classification accuracy and false positive rates. The number of mistakes (backspace + delete key) was calculated. There are many different methods to correct the mistakes, but it was not possible to catch all of the possible correction scenarios as keystrokes were collected from different computer applications. Outliers for all of the features that involved multiple keys were calculated to remove these pauses (e.g. digraph latency). Pauses were removed by considering the mean and standard deviation for all keystroke dynamic features, which they were 12 standard deviations greater than the mean for each individual participant [8].

This process has been considered in the prototype application while recording and collecting the data from users. The Kappa statistic indicates how much the classification rate was a true reflection of the model or how much chance/probability could be attributed to be succeeded.

Table 1. Keyboard keystroke dynamics classification of human emotions

Emotion	Accuracy %	Kappa	False Positive %
Nervousness	85.20	0.67	10.26%
Relaxation / Neutral	79.40	0.55	17.60%
Sadness	87.10	0.76	9.36%
Fright	91.24	0.68	4.35%
Anger	83.90	0.53	12.86%

Table 1 shows the classification results of human emotions based on the keyboard keystroke dynamics with their Kappa values and False Positive Rates. Fright emotion has the strongest classification of 91.24% with the least value of 4.35% for false positive rate.

Mouse

In the first detection process, features were selected. The best result was on Neutral emotion but in the other emotions, the outcome is less than 40%. Then in the second phase, the features were selected according to the Schuller [7] features, and it is far better than the first result as shown in Table 2. Despite in the *fright* (*afraid*) emotion the resulted percentage is weak; but in the other two emotions of *pensive* and *annoyed*, the results are much improved.

The collected data from our volunteers are evaluated and then analysed with our emotions. Table 2 summarizes the evaluation of all collected data set of 4 set of emotions. This evaluation was done based on the 2003 collected data vectors [9, 10]. The overall average of the correctly classified emotions is 0.866 with a mean variance of 0.075.

The correctly classified emotions by RapidMiner are the values at the junction of the same detected emotion with the intended emotion as shown in bold. The other values are called as false positive alarms, which are classified incorrectly. The increased classification rate for neutral emotion can be explained easily. The test subjects were accumulated primarily with expectations of neutral data vectors for the emotion. This probably is the most emotion which is felt to have been distributed over the days.

Table 2. Confusion matrix with the average values for mouse features classification

Intended Emotions	**Detected Emotions**			
	Neutral	Fright	Sadness	Nervousness
Neutral	**0.930**	0.022	0.027	0.023
Fright	0.203	**0.787**	0.040	0.010
Sadness	0.084	0.012	**0.912**	0.015
Nervousness	0.175	0.015	0.065	**0.835**

If a PC user has the emotion of *annoyed*, he moves the mouse usually very fast and also fixed with short presses on the mouse button. The properties of mouse movement are *fast* and *brief*. The system can certainly capture and analyse these features. In the short pressing the mouse pointer doesn't move often. Thus generally no movement is detected during the mouse click.

But a question still remains that why the precision of the detection is still low. Here is the answer by analysing the volunteers. At the time of working in different situations with the computer, they are not sure about their own emotions. When they are asked to input their emotions, they are rather unsure what kind of emotion they have at the moment.

Thus an insecure person presses a little longer and deliberates on the mouse, where they will lead, and the person did not intend slight movement of the cursor. Data analysis of the features is shown for the recognition of emotion. It is not very meaningful, and this is probably one of the reasons for the lower values in the confusion matrix.

Finally, it can be concluded that although the recognition of emotion with a reliable performance works, unfortunately the lack of standard hardware with significant qualities cause a lower accuracy. It would be very important to have several data collection periods to increase the strength of the data. It also brings more clarity about the emotions; and it enables better detection.

Touch Screen

Table 3 shows the final average results over all the test subjects. The overall confusion matrix with an overall average of 0.76 (76%) for the correctly classified emotions values shown in bold font is achieved. After the evaluation of the existing system for the detection of the four emotions, it can be concluded that this system can be used for emotion recognition with a reliable accuracy.

Table 3. Confusion matrix with the average values for touch screen

Selected Emotion	**Detected Emotion** Neutral	Fright	Sadness	Nervousness
Neutral	**0.71**	0.321	0.090	0.022
Fright	0.000	**0.900**	0.073	0.000
Sadness	0.008	0.113	**0.893**	0.000
Nervousness	0.071	0.354	0.122	**0.553**

4.3 Hybrid Analysis

By the combined results of the keyboard, mouse and touchscreen, the accuracy in the *Fright (Afraid)* emotion is the best among the others. Neutral and Nervousness have the lowest result, and these two emotions have the greatest rate of confusion with each other. These results are tabulated in Table 4.

As it can be seen in Table 4, all of the four emotions have been detected more accurately by using all three input devices (Keyboard, Mouse and Touch Screen) analysis methods. Also in some cases, the error has been increased a little bit, but the increase of performance is much higher than the error rates.

Table 4. Confusion matrix with the average values of Keyboard, Mouse and Touchscreen

Selected Emotion	**Detected Emotion** Neutral	Fright	Sadness	Nervousness
Neutral	**0.851**	0.121	0.076	0.022
Fright	0.001	**0.932**	0.082	0.010
Sadness	0.008	0.118	**0.921**	0.004
Nervousness	0.091	0.254	0.122	**0.650**

5 Conclusion

Previously, researchers tried to gain more accurate results on human emotions recognition. This research could gain higher accuracy in comparison with the other researchers who worked with keyboard, mouse and touch screen. Especially by

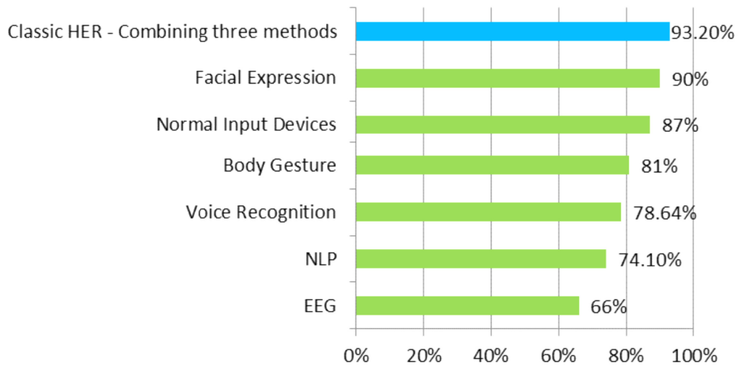

Fig. 2. Comparison of the best accuracies of different methods for HER

combining these three methods together, much better result of maximum 93.20% has been achieved, which is competitive with all other previous methodologies. Figure 2 compares the achieved results with the other methods and accuracies as discussed earlier in research literature.

The final best result of 93.20% among the achieved results of this research has been compared with the best results of the superior research on HER with different methods. In comparison only the best accuracies of the methods are considered. We cannot compare them based on the average accuracy of the emotions because of few problems. The problems are related to existing gap of the acquired accuracies, differences in number of emotions recognized, and different datasets for each research.

HER based on EEG has gained the accuracy of 66% in 2009 and 2010 [11, 12]. In 2008, *Lei et al.* got the result of 74.1% in the emotion of *anger* [13]. Voice processing in HER achieved the 78.64% of accuracy [14, 15].

HER systems based on body gesture recognition has been resulted in 81% by *Gunes & Piccardi* [16]. The most similar methods for this research has been done by *Milanova & Sirakov* and they gained 87% of emotions recognition [17]. The best competitive method is facial expression recognition which has improved a lot recently in 2012.

Among many researchers in facial expression recognition, *Konar et al.* and *Kao & Fahn* got 88.68% and 90% of accuracy respectively [3, 18]. Figure 2 shows that the method in this research has worked nearly from 5 to 6% better than the similar methods and more than 3% better than the superior method in facial expression recognition in 2012.

The evaluation criteria are used to consider for performing the evaluation. There were three main criteria for evaluation.

- The first criterion was recognition accuracy which is the most important item in the evaluation. The proposed methods of this research have been evaluated in terms of classification/recognition accuracy. Then at the end in Figure 2, they are compared with the similar research areas in measuring human emotions recognition accuracy.

- The second criterion was false positive rate which has been shown in every confusion table. However, the lack of enough information in the previous research papers, comparing the results of this study with the similar works was not possible.
- And the third criterion was computational/processing time. This is only related to the classification methods and the number of extracted features.

References

1. Cowie, R., Douglas-Cowie, E., Tsapatsoulis, N., Votsis, G., Kollias, S., Fellenz, W., Taylor, J.G.: Emotion recognition in human-computer interaction. IEEE Signal Processing Magazine 18(1), 32–80 (2001)
2. Ekman, P., Friesen, W.V.: Unmasking the face: A guide to recognizing emotions from facial clues. Malor Books (2003)
3. Konar, A., Chakraborty, A., Halder, A., Mandal, R., Janarthanan, R.: Interval Type-2 Fuzzy Model for Emotion Recognition from Facial Expression. In: Kundu, M.K., Mitra, S., Mazumdar, D., Pal, S.K. (eds.) PerMIn 2012. LNCS, vol. 7143, pp. 114–121. Springer, Heidelberg (2012)
4. Huang, T.: Audio-visual human computer interface. In: IEEE International Symposium on Consumer Electronics. IEEE, University of Illinois (2008)
5. Monrose, F., Rubin, A.D.: Keystroke dynamics as a biometric for authentication. Future Generation Computer Systems 16(4), 351–359 (2000)
6. Chang, M.: Iowa State engineers use keyboard, mouse and mobile device 'fingerprints' to protect data (November 18, 2013), `http://www.news.iastate.edu/news/2013/11/18/fingerprints`
7. Schuller, B., Rigoll, G., Lang, M.: Emotion recognition in the manual interaction with graphical user interfaces. In: IEEE International Conference on Multimedia and Expo. IEEE (2004)
8. Epp, C., Lippold, M., Mandryk, R.L.: Identifying emotional states using keystroke dynamics. In: Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems, pp. 715–724. ACM, Vancouver (2011)
9. Kemp, F.: Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. Journal of the Royal Statistical Society: Series D (The Statistician) 52(4), 691–691 (2003)
10. Cohen, J.: Applied multiple regression/correlation analysis for the behavioral sciences. Lawrence Erlbaum Associates, Hillsdale (2003)
11. Liu, Y., Sourina, O., Nguyen, M.K.: Real-time EEG-based Human Emotion Recognition and Visualization. In: International Conference on Cyberworlds. IEEE, Alberta (2010)
12. Schaaff, K., Schultz, T.: Towards emotion recognition from electroencephalographic signals. In: Affective Computing and Intelligent Interaction and Workshops (ACII). IEEE, Memphis (2009)
13. Li, H., Pang, N., Guo, S., Wang, H.: Research on textual emotion recognition incorporating personality factor. In: IEEE International Conference on Robotics and Biomimetics. IEEE, Sanya (2008)
14. Amarakeerthi, S., Ranaweera, R., Cohen, M.: Speech-Based Emotion Characterization Using Postures and Gestures in CVEs. In: International Conference on Cyberworlds. IEEE, Alberta (2010)

15. Xiao, Z., Dellandrea, E., Dou, W., Chen, L.: Automatic hierarchical classification of emotional speech. In: Ninth IEEE International Symposium on Multimedia Workshops (2007)
16. Gunes, H., Piccardi, M.: Fusing face and body gesture for machine recognition of emotions. In: IEEE International Workshop on Robots and Human Interactive Communication. IEEE (2005)
17. Milanova, M., Sirakov, N.: Recognition of Emotional states in Natural Human-Computer Interaction. In: IEEE International Symposium on Signal Processing and Information Technology (ISSPIT). IEEE, Ho Chi Minh City (2008)
18. Kao, C.Y., Fahn, C.S.: A Design of Face Detection and Facial Expression Recognition Techniques Based on Boosting Schema. Applied Mechanics and Materials 121, 617–621 (2012)