International Journal of Computer Science and Engineering (IJCSE)
ISSN 2278-9960
Vol. 2, Issue 4, Sep 2013, 163-170
© IASET

MOOD-I-COMP

VIDUSHI JAIN & RUSHABH MEHTA

B. E. Student (EXTC), D. J. Sanghvi College of Engineering, Mumbai, Maharashtra, India

ABSTRACT

Computers today have become an integral part of our lives. They are used everywhere, be it at an office or home or a school, and by everyone from a young child to an adult. Although computers have made life simpler in many ways, still one often feels drained after sitting for hours together in front of it. Today a computer is one of the most efficient man-made machines, but it is still only a machine.

Now imagine if the computer with which you spend most of your day could interact with you just as a friend. It could detect your mood, understand how you feel, speak with you and react appropriately.

"Mood-i-Comp" is a technology which takes us a step closer to what you have just imagined. Mood-i-Comp aims at creating smart and interactive computers with extraordinary perceptual abilities, enabling them to detect human emotions and react to them.

KEYWORDS: Active Appearance Model, Audible Contour, Electro-Cardiogram, Emotion Mouse, Mood-i-Comp, Photoplethysmogram, Standard Deviation, Trust Factor

INTRODUCTION

Humans are both conscious and intelligent although it's very easy to imagine one attribute without the other. An intelligent but unconscious being is a "zombie", similarly we can also imagine a conscious but a non-intelligent person.

In this age of technology digital computers will most certainly be intelligent one day, and we have almost reached that stage. But will they ever be conscious?? This is a question which is still to be answered. The possibility of intelligent computers has obsessed mankind since Alan Turing first raised it formally in 1950. Turing was vague about consciousness, which he thought unnecessary to machine intelligence. But artificial consciousness is surely as fascinating as artificial intelligence. According to John Searle digital computers can never be conscious as they are made up of the wrong stuff. He noted that consciousness results from the chemical and physical structure of humans and animals as photosynthesis results from the chemistry in plants. You can't program your laptop to transform carbon dioxide into sugar; computers are made of the wrong stuff for photosynthesis and for consciousness too.

But some thinkers reject the wrong-stuff argument and believe that, once computers and software grow powerful and sophisticated enough, they will be conscious as well as intelligent. They point to a similarity between neurons, the brain's basic component, and transistors, the basic component of computers. Both neurons and transistors transform incoming electrical signals to outgoing signals. Now a single neuron by itself is neither conscious, nor intelligent. But together they form the brain of a conscious and intelligent human. A single transistor seems likewise unpromising. But gather lots together, hook them up right and you will get consciousness, just as you do with neurons.

So let's turn to today's digital computer. It is an ensemble of (1) the processor, which executes (2) the software, which (when it is executed) has the effect of changing the data stored in (3) the memory. The memory stores data in a numerical form, as binary integers or "bits." Software can be understood in basic terms as a series of commands to be executed by the processor, each intended to accomplish one part of a (potentially complex) transformation of data in the memory. In other words: by executing the software, the processor gradually transforms the memory from an *input state* to an *output or result* state. Hence today's computers, though not conscious, ever-faster and more capable computers with ever-more-complex, sophisticated software will eventually be conscious.

Therefore, though intelligence and consciousness can exist independently if we put the two together, the result is obviously more powerful than either alone. Hence in order to interact with your computer in the same way as you interact with your friend, intelligence along with consciousness is what your computer needs. Human cognition depends on the ability to perceive, interpret and integrate any form of sensory (audio, video or touch) information.

The "Mood-i-Comp" technology aims at creating computational machines that have these abilities very close to that of humans. Hence a machine can understand what a user wants, realize his emotional state and take action accordingly. In order to achieve this level of interaction between computer and humans it is necessary that the computer understand the various human emotions. In this paper we have proposed three ways of detecting the six major human emotions of happiness, sadness, fear, surprise, disgust and anger. They are:

- The Emotion Mouse
- Emotion detection via facial expressions
- Emotion detection via speech.

The aim of this technology is to ultimately reach a stage where the computer will react to the emotions expressed by the user. For instance when the user switches on the computer, it greets him by saying, "Good evening sir, how was your day?" Then depending on the reply from the user, the computer will suggest music or videos the user prefers when under a similar emotional state. Further if the user is sad or angry it may suggest some comical videos or open entertaining emails in order to improve the user's mood. The emotional computation of the user is carried out continuously when he or she is using the computer using the three methods mentioned above.

In this way, Mood-i-Comp technology may eventually be able to convert your personal computer into a pleasant companion which could understand you.

PRINCIPLE

Emotion computing is an important aspect in order to have meaningful interaction with your computer. Emotion computing comprises of two parts:

- Emotion detection and
- The ability to express emotions

The techniques used for this purpose in Mood-i-Comp are as described below:

Emotion Mouse

It has been seen that there is a correlation between a person's emotional state and physiological measurement. These physiological measurements include pulse, galvanic skin response, temperature, somatic movement and blood pressure. The physiological measurements are different for people with six different emotions which include anger, fear, sadness, disgust, joy and surprise. Hence a correlation can be drawn between these measurements and emotions to provide emotion detection. This principle for emotion detection is used in the Emotion Mouse.

Facial Expressions

Facial expressions provide a key mechanism for understanding and conveying emotion. For an ideal human computer interaction facial expressions become very important as they play a major role in conveying one's emotions. Therefore to be able to read the facial expressions and interpret them correctly is another stepping stone towards smarter and conscious computers.

In order to recognise expressions a useful feature scheme and extraction method is chosen. The method incorporates both face and texture information from facial images.[2] The Active Appearance Model (AAM), developed initially by Cootes and Taylor, has shown strong potential in a variety of facial recognition technologies. It has the ability to aid in initial face-search algorithms and in extracting important information from both the shape and texture like wrinkles, nasio-labial lines, etc. of the face that may be useful for communicating emotion.

Speech Recognition

Speech recognition is an effective way by which the mood of a person can be judged. Mood detection is not very easy as human emotions are hard to characterize and categorize. Emotion recognition solutions depend on which emotions we want a machine to recognize and for what purpose.[1] This technology can be used in various ways and in different walks of life. An obvious example is that if through this technology we know that a customer is angry, more preference can be given to him.

Automatic emotion recognition of speech can be viewed as a pattern recognition problem. The results produced by different experiments is characterized by: a) the features that are believed to be correlated with the speaker's emotional state, b) the type of emotions that we are interested in; c) the database used for training and testing the classifier; and d) the type of classifier used in the experiments.

The performance of an emotion classifier relies heavily on the quality of the database. Speech data used for testing emotion recognition can be grouped under the following three categories depending on the way the speech signal is captured.

- The first method uses people to record utterances, where each utterance is spoken with multiple feigned emotions. The people are usually given the time to imagine themselves into a specific situation before speaking.
- The second method uses a program that interacts with the people and drives him into a specific emotion situation and then records his responses.
- The third method, which is hard to obtain, is actual real-world recording of utterances that express emotions.

WORKING

Emotion Mouse

The hand can be used to detect the emotion of a person through a device called an emotion mouse. It provides a relation between the cardiac rhythm, the body temperature, electrical conductivity of the skin and other physiological attributes with the mood. The mouse includes a set of sensors including infrared sensors and temperature sensitive chip.[1] Information is collected by the mouse on the basis of mouse movements, button click frequency as well as the finger pressure when the user presses his or her button. Physiological information is obtained using:

- Electro-cardiogram (used to measure heart rate),
- Photoplethysmogram,

- Thermister (used to measure skin temperature),
- Galvanic skin response (skin electricity) and
- Electro-myogram (electromyographic activity).

Hence once the computer detects the mood the interaction definitely becomes better and more meaningful.[1]

Implementation

- Each of the above mentioned physiological signals are recorded via the sensors and a baseline is calculated.
- The baseline is nothing but the physiological behaviour under normal or no-emotion condition.
- Also the mean value for all of the above mentioned six emotions needs to be pre-calculated.
- All the above calculations are done over a period of five minutes and the mean is taken for each of the emotions as well as the baseline.[5]
- To account for the change in physiology, the difference between the mean-emotion-score and the baseline is calculated.
- During its use if the score for the physiological sensor lies within one and a half standard deviations of the pre-calculated means then the particular emotion is identified.
- Also a note is made of the user's behaviour under various emotions, for example, the music he listens to or the website he browses.
- This helps the system to, in time, offer useful suggestions depending on the user's mood.
- This database of the user's activities under different conditions is always recorded and updated.
- Based on these values of variance from the baseline emotion the appropriate emotion is predicted and the computer interacts with the user accordingly.

Facial Expressions

In order to detect the mood from facial expressions, still images form a preferable input. In addition to the input an optimum method for feature extraction is also needed. [3]Active Appearance Models (AAMs) are chosen for this purpose. The implementation of the technique is described in the following steps:

- Building an appearance model for the AAM entails choosing images to be used as training data and then properly labelling the images using a predefined format.
- The images for the AAM are labelled such that the key expressing features such as the lips, eyes, eye-brows and other facial guidelines that prominently help in expression identification are all covered.
- Facial images and their expressions in the training database are categorised into different emotions based on the emotion clarity and sincerity.
- Emotion clarity ranks the image based on the clarity of the emotion.
- Sincerity is the measure to help determine how well the subject conveys a certain emotion.
- The subject is assigned different scores from zero to one for emotion clarity as well as sincerity for each emotion.

- Sincerity is marked inversely i.e. if a person shows happiness less sincerely he is marked with a high score. This makes it easier to detect those emotions which are less sincerely expressed, yet felt all the same, by the user.
- Hence emotion clarity helps in identification of the emotion while sincerity is used to reinforce the decision.
- The average of the emotion clarity and sincerity is calculated and the resultant is assigned as the final weight.
- During the operation, images of the user are taken regularly at predefined intervals and the mood is predicted depending on the final weight calculated from the expression.
- In cases where final weights are lower than a threshold, no decision is made.
- After the system is trained with a predefined database and also with its user for some time it is ready to operate.
- The system, during its training procedure with the user makes note of the activities of the user under different emotions. This also continues after the training process and the activities are continuously updated.
- Hence in time it is ready to interact with the user and make suggestions depending on the user's mood.

Speech Recognition

In each of the methods mentioned for speech capturing, each utterance is split into frames of size 384 samples and a window step of 192 samples where the sampling rate is 22.05 KHz. 37 prosody features related to pitch (fundamental frequency), loudness (energy), and segments (audible durational) as follows.[1]

- **Fundamental Frequency Features**

Pitch contour for the input utterance is obtained. A number of pitch features have to be calculated. Minimum, maximum, mean and standard deviation is obtained for the pitch contour. Minimum, maximum, mean and standard deviation is obtained for the first derivative of the pitch contour. The pitch values are then provided as input to a filter g and then normalized by the frame pitch value.

- **Energy Features**

Energy contour for input utterance is obtained. The energy for each frame in the utterance is a function of time. A number of energy features have to be calculated. Minimum, maximum, mean and standard deviation is obtained for the energy contour. Minimum, maximum, mean and standard deviation is obtained for the first derivative of the energy contour.

- **Audible Durational Features**

A set of features related to audible segments per utterance is obtained. To identify audible segments in a speech utterance, we need to first obtain the maximum frame energy in the utterance and then consider any frame whose energy level is below a threshold (percentage of the maximum energy) as an inaudible frame; else it is audible. The audible/inaudible contour is then filtered using a low pass filter to smooth the segments. As a result, an audible and inaudible segments contour is produced. Inaudible segments are related to pauses in the speaker utterance. This feature can be very useful in detecting the mood of a person.

Hence by using features extracted from pitch contour, energy contour, and audible segment duration contours we can achieve a high degree of accuracy in distinguishing certain emotions.

Trust Factor

In order to apply the above techniques in perfect sync, a trust factor is assigned to the decision made by each of them individually. The trust factor can vary according to the ambiguity in the recognition. For example an image which is ambiguous in expression detection due to bad light or improper expression of user it is given a low trust score, or if the various physiological signals obtained from the emotion mouse contradict each other the trust factor drops. Hence, a detection of “happy” with a trust factor of 0.2 by the speech recognition will be superseded by a detection “surprise” by facial expression having a trust factor 0.9. However, if the identified emotions of two systems agree with one another the trust factors for that emotion are added and if they disagree the trust factors are subtracted. In case all the systems detect a different emotion, or the final trust factor has a value below a set threshold, the process fails and is repeated.

APPLICATIONS

“Mood-i-Comp” technology can be used in different ways and in different places. It can be used in places like:

- Smart computers –Using this technology the computer acts just like your friend, where it talks to you, responds to you and understands you. Thus acting as a stress buster.
- Call centres-The recorded speech of the caller can be analyzed and thus his/her emotion can be detected in order to categorize calls of greater urgency.
- Software games – Now maybe we will not need a second player with you to enjoy your game even further as you now have your computer to do just the same.
- Talking toys – Kids can now ask questions to their dolls and soft toys and expect answers for the same.
- Fraud detection –During interrogation the changing moods of the culprit can be analyzed to draw conclusions about him.
- Customer Relation management – Giving services to the customers depending on their mood can definitely help to build up the customer management relationship.

LIMITATIONS

The proposed technology has immense scope for improvement and sophistication in the future. However in its current state it has the following limitations:

- It becomes extremely difficult and tedious to make a database as it is very vast and has to take into account a large number of moods of different sorts of people.
- The readings obtained for moods like sadness and boredom are quite similar and hence can lead to confusion.
- Facial recognition for people wearing glasses or caps or in uneven lighting becomes difficult. 4) A person who seems to be happy through his facial expressions does not necessarily mean that he is happy. Hence the system and database have to be extremely accurate.

CONCLUSIONS

In this paper we have presented a way in which we can take a step forward towards making our computers smarter. This is a technology has an immense scope in the future. We have illustrated in detail the principle and explained various ways of making our computers friendlier by teaching them how to detect one's emotions and react accordingly. This technology has the potential to explore a new dimension of human relationship with machines.

ACKNOWLEDGEMENTS

Every project is a result of the hard work put in by many. Our project is also not different in this respect. So we hereby take this opportunity to thank all the professors and others who helped in making this project successful.

REFERENCES

1. Binyamin M.(July 2010) "Blue Eyes"
2. Carlos Busso- "Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information"
3. Spiros V. Ioannou, Amaryllis T. Raouzaiou, Vasilis A. Tzouvaras- "Emotion recognition through facial expression analysis based on a neurofuzzy network"
4. Matthew S. Ratliff, Eric Patterson- "Emotion Recognition Using Facial Expressions With Active Appearance Models"
5. Wendy Ark D. Christopher Dryer Davia J. Lu- "The Emotion Mouse"