

RESEARCH

Open Access



Towards affective touch interaction: predicting mobile user emotion from finger strokes

Sachin Shah, J. Narasimha Teja and Samit Bhattacharya 

Abstract

The role of affect and emotion in interactive system design is an active and recent research area. The aim is to make systems more responsive to user's needs and expectations. The first step towards affective interaction is to recognize user's emotional state. Literature contains many works on emotion recognition. In those works, facial muscle movement, gestures, postures and physiological signals were used for recognition. The methods are computation intensive and require extra hardware (e.g., sensors and wires). In this work, we propose a simpler model to predict the affective state of a touch screen user. The prediction is done based on the user's touch input, namely the finger strokes. We defined seven features based on the strokes. A linear combination of these features is proposed as the predictor, which can predict a user's affective state into one of the three states: *positive* (happy, excited and elated), *negative* (sad, anger, fear, disgust) and *neutral* (calm, relaxed and contented). The model alleviates the need for extra setup as well as extensive computation, making it suitable for implementation on mobile devices with limited resources. The model is developed and validated with empirical data involving 57 participants performing 7 touch input tasks. The validation study demonstrates a high prediction accuracy of 90.47 %. The proposed model and its empirical development and validation are described in this paper.

Keywords: Emotional state, Touch screen, Strike and tap, Features, Linear regression, Empirical study

Introduction

In human-computer interaction (HCI), we are mostly concerned about taking the human factors into account while designing interactive systems. Such factors include cognitive behaviour such as our abilities to learn, reason or solve problems; sensory abilities such as haptics and auditory input; motor actions that include eye movement, finger/hand movement and so on. These factors in turn may get influenced by our emotional state of mind, affecting the way we interact with a computer [22, 38]. Therefore, researchers tried to incorporate affect and emotion in HCI, resulting in interaction styles that are "affective" [28]. The goal of affective interaction is to make systems more natural and responsive to the goals and expectations of the user, so as to improve usability and user experience.

Affective interaction has two main stages. In the first stage, the affective state of the user is *recognized*. This information about the emotional state of mind is used to design interaction in the second stage, which *complements* and/or *change* the user's affective state. For example, the number of steps required to perform a task may be reduced if the user is in a happy state, so that the user experience improves. On the other hand, the same task may be designed to require more number of steps to complete, when the user is excited so as to minimize the possibilities of error. Although the second stage is non-trivial, the more difficult part is the first component; how to recognize the emotional state of an individual?

We see all around us various devices that are operated by touch. The popularity of mobile touch screen devices have seen significant rise in recent years, mainly due to the availability of affordable smart phones and tabs. According to a survey by the Federation of Indian Chambers of Commerce and Industry (FICCI) and KPMG

* Correspondence: samit3k@gmail.com
Department of Computer Science and Engineering, IIT, Guwahati
781039 Assam, India

International¹, 59 million people were using smart phones till 2013 in India and it is expected to reach 265 million by 2016. The statistics is in fact reflective of a global trend. Since these devices are being used by the masses, usability and consequently the HCI issues are very important for these devices. As we have already noted before, emotion influence usability. Therefore, it is necessary to work in the direction of *affective touch interaction* for mobile touch screen devices. In this article, we present the first step towards achieving the objective; namely, detect emotion for touch screen users.

Literature contains many works on emotion recognition. In these works, methods were proposed to recognize emotion from facial expression, gesture, posture and physiological signals. These mostly involved computer vision and image processing techniques, which are computationally expensive. Since we are dealing with mobile devices, we should keep in mind that the devices (particularly the affordable ones) come with limited resources in terms of power backup or processor speed. Moreover, the existing methods often require additional set ups such as cameras or probes and wires to record physiological signals. In the present context, attaching extra hardware may not be feasible considering the mobility aspect. Therefore, the existing emotion recognition approaches are not suitable for the mobile touch screen devices.

For affective interaction to happen on mobile touch screen devices, we therefore require an emotion recognition technique that does not depend on expensive computations or need any extra setup. The method we propose relies on users' touch interaction behaviour (finger strokes) to detect the emotional state of mind. It does not require additional sensors or wires to record the finger strokes. Also, the computations are much less compared to the existing methods. Hence, the proposed approach is expected to be more suitable for mobile touch input devices.

Our proposed approach works based on the categorization of users into three emotional states, namely *positive* (representing the emotions happy, excited and elated), *negative* (representing the emotions sad, anger, fear and disgust) and *neutral* (representing the emotions calm, relaxed and contented). From the finger stroke behaviour of a user, we predict his/her affective state into one of these three categories. We identified ten features based on the user's finger stroke behaviour, to build the predictor. We assume that these features provide *indirect indication* to the user's emotional state. The idea comes from the research findings that touch can tell about emotion types [14]. We first experimented with classification models as the predictor. However, the accuracy of the best classifier turned out to be moderate (72 %). Subsequently, we developed a

linear combination of seven out of the ten features following the linear regression approach, as the proposed predictor. Empirical validation of the regression model demonstrates a prediction accuracy of 90.47 %, making it suitable for practical use. The proposed model along with the empirical data collection and analysis are described in this article. The article is organized as follows.

Works related to emotion detection are discussed in Section Related work. The feature set is described in details in Section The feature Set. The empirical study details are presented in Section Empirical data collection. The proposed model is discussed in Section Proposed model. It contains description of both the classification and regression approach and the comparative results for both. A discussion on the proposed model, including its pros and cons, along with the scope for future works is presented in Section Discussion. Section Conclusion concludes the article.

Related work

In recent years, emotion and its role in computing system design and application has become a much studied area [29]. It is often found to be the driving force behind motivation, positive or negative. Therefore, HCI researchers attempted to integrate the theory of emotions with usable system design [28].

There are broadly two ways of representing emotions: the discrete model [9] of emotion and the continuous model [31]. The former view posits that emotions are discrete, measurable, and are physiologically distinct. According to this model, there are basic emotions such as happy, sad, angry, scared, tender and excited. Any emotional state can be considered to be a sub state of either of the basic emotional states. The continuous model, on the other hand, represents emotion as a point in a two dimensional space of *valence* and *arousal*. The x-axis represents the valence and the y-axis represents the arousal values. Based on these two models, several works are reported on emotion detection. We can broadly classify these techniques into the following categories.

1. Emotion detection from facial expressions
2. Emotion detection from gestures and postures
3. Emotion detection from physiological signals and eye gaze data
4. Other works on affective computing with detection as a component

Emotion detection from facial expression

Ekman et al [10] proposed the theory of emotion by facial expression, which had a significant effect on the works on emotion detection. They correlated facial display of emotion to the facial muscular movement. The facial muscle movements give rise to six basic emotions:

anger, happiness, disgust, fear, sadness and surprise, which are mutually exclusive. Bartneck and Reichenbach [3] reported the influence of *geometric intensity* on an emotional facial expression. They used the intensity information to categorize emotional faces into the basic classes. Katsyri and Sams [18] proposed to use *dynamic* facial expressions as a more accurate detector of emotions. Isbister et al [16] used the continuous model to evaluate emotion from facial expressions. Niewiadomski and Pelachaud [25] proposed embodied conversational agents (ECAs), which could synthesize and display a large number of complex facial expressions. The synthesis of facial expressions was based on a fuzzy method. Using this method, the ECA is able to mask/hide emotions and generate superposed/inhibited/fake expressions.

Emotion detection from gesture and posture

We are familiar with the fact that body movements and gestures play an important and significant role during communication. According to McNeill [23], there are four main types of gestures: *iconics*, *metaphorics*, *beats* and *deictics*, which result from the hand and arm movements. Glowinski et al [12] reported on GEMEP (Geneva multimodal emotion portrayals) related to human upper-body movements and their relation to affect states. They tracked head and hand movement for frontal and lateral view. Postural and dynamic expressive gesture features were identified and analysed. Emotions were classified according to the valence (positive, negative) and arousal (high, low) of the body movements (i.e., following the continuous model of emotion). Bianchi-Berthouze and Kleinsmith [4] formalized a general description of posture based on angles and distances between body joints. They used it to create an affective posture recognition system that maps the set of postural features into affective categories using an associative neural network. Emotion detection from full body movements was reported by Kapur et al [17], in which video sensors were used to record 3D positions of 14 body joints over time. Camurri et al [5] worked on emotion detection from full body movements and expressive gestures. They used *non-propositional body movement qualities* (e.g. amplitude, speed and fluidity) to infer emotion rather than trying to recognize different gesture shapes. The work was based on the continuous model of emotion. A somewhat different approach was taken by Paiva et al [27]. They designed SenToy, a *tangible* doll having a sensing device in it. The toy was connected to a synthetic character in a game. It took user gesture and movements as input, which it used to detect the emotional state of the user. Based on the emotional state, it could change the game.

Emotion detection from physiological signals and eye gaze data

Researchers also made efforts to use physiological signals for emotion detection. Such signals include electrooculogram (EOG), galvanic skin response (GSR), heart rate (HR), electrocardiogram (ECG) and eye blinking rate (EBR). Takahashi [34] recorded EEG (electroencephalogram) and peripheral physiological signal like pulse and skin conductance for recognition of five basic emotions, although the recognition rate was not high. Koelstra et al [20] recorded multiple physiological signals such as EOG, HR, GSR, EBR and EEG to detect emotions. In Alzoubi et al [1], three physiological signals, namely EMG (Electromyography), ECG and GSR were recorded and used for detection of user's affective states. Hazlett [13] reported the use of EMG signal to measure positive and negative emotional valence during interactive experience. Soleymani et al [32] worked on using eye gaze data for emotion detection. The data included blink data such as blink depth, blinking rate, length of the longest blink and time spent with eye closed, gaze distance and gaze coordinates.

Other related works

We surveyed few more works that are related, though may not be directly relevant, to our work. In these works, emotion detection plays an important role and various methods were proposed for the same. However, the input need not have been taken from the users directly, unlike the works discussed in the previous sections.

Research was done to understand emotion from multimedia content [2, 15, 26]. Xu et al [37] reported work on emotion recognition of multimedia content. The work was aimed at improving *video on demand* services, to let the user pick up their favourite affective content. Use of text conversation system to learn emotion was reported in [6]. There had been work related to embodied agents monitoring emotion [33] and mental states of their users and provide appropriate, affective responses. Educational agents, for example, can monitor student attention and seek to improve it when student engagement decreases. Another work investigated whether and how digitally mediated social touch (remote touch) may influence the sense of connectedness toward a speaker and the emotional experience of what were being communicated [36]. The PAM (Photographic Affect Meter) tool was reported in [30], which was designed to help in the assessment of affect. A continuous and objective evaluation of emotional experience through interactive play environment had been reported by Mandryk et al [24]. Vloed and Berentsen used a non-intrusive sensor (bed sensor) to detect emotion [35].

In this work, we aim to detect emotion for users of mobile touch screen devices. The primary examples of

such devices and the most popular ones are smartphones and tablets. These devices have limited computing powers. In contrast, the major approaches towards emotion detection, as discussed in Section Emotion Detection from Facial Expression, Emotion Detection from Gesture and Posture, Emotion Detection from Physiological Signals and Eye Gaze and Other Related Works require significant computations. Moreover, many of those require additional hardware setup such as video sensors, sensors to measure physiological signals, eye trackers and so on. These extra setups may be costly, may not be convenient to use considering mobility of the targeted devices and may not be supported by the devices at all. As a result, we need to come up with techniques that do not require extra set-up or significant computation.

We propose to use touch interaction characteristics to predict emotion. The touch interaction characteristics are captured in terms of finger strokes. We assume that the strokes are indirect pointers to the user's affective state. Since we are not collecting any other input, no extra set-up is required. We found very few works in this direction. Khanna and Sasikumar [19] have shown that emotion can be recognized by keystroke patterns, especially frequency of some special keys (e.g. spacebar and backspace). Epp *et al* [8] have also tried to identify emotional states using keystrokes dynamics. However, these works were aimed to detect emotion for desktop computer users. In contrast, Coutrax and Mandran [7] attempted to detect affective state from touch inputs. Clearly, the work fits our objective directly. However, they used both 2D screen gestures and 3D motion gestures. Apart from using a large feature set, the approach is constrained by its requirements of large amount of sensory inputs that may not be available with all smart phones, particularly the affordable ones. Lee *et al* [21] also attempted to detect user emotion from indirect cues. Their approach, however, is based on inputs that are available only for text entry tasks (the context of the work was social communication in which text entry plays a key role) and not found in general purpose tasks. In additions, some extra sensory features were also used, which may not be available with all devices. Gao *et al* [11] studied emotion detection from touch information. However, the work was done in the limited context of game playing on iPod. In this work, we propose a predictive model for emotion detection, which works based on finger strokes, requires much less computation compared to the other methods and applicable to general touch interaction scenarios. The model is based on a set of features, which is described next.

Methods and Results

The feature Set

There are three finger actions during a touch interaction: *down*, *up* and *move*. A *down* action signifies the

time instance at which the finger touches the screen. Likewise, *up* action is the time instance when the finger is released. After a *down* action, if the finger moves on the screen without *up* action, we call it a *move* action. These three actions can be used to define two touch interaction characteristics: *strike* and *tap*. A tap is a combination of *down* and *up* actions whereas a strike is a combination of *down*, *up* and *move* actions. In practice, however, we may not be able to have a perfect tap and there will always be some small movement by our finger, although we intend to avoid it. Hence, we differentiate between the two based on the *strike length*. If the length is less than or equal to a specified limit, then we designate it a tap; otherwise a strike. Using the notion of strikes and taps, we define the following ten features.

Deviation in number of strikes

For a given touch interaction task (e.g., setting a reminder, opening a web browser etc), we can define *minimum* number of strikes, as the number required to perform the task without error. Let this value be denoted by S_M . We propose the feature *deviation in number of strikes* (S_{dev}) as the difference between the observed number of strikes (S_O) and S_M , as shown in Eq 1.

$$S_{dev} = S_O - S_M \quad (1)$$

Deviation in number of taps

Similar to S_M , we define the minimum number of taps (T_M) for a task as the ideal number of taps required to perform the task without error. Then, we define the feature *deviation in number of taps* (T_{dev}) as the difference between the observed number of taps (T_O) and T_M , shown in Eq 2.

$$T_{dev} = T_O - T_M \quad (2)$$

Average strike length

For calculating this feature value, we first find the length of each strike. We then sum up these lengths and divide the sum by the total number of strikes, as shown in Eq. 3 where L_i is the length of the i^{th} strike and M is the total number of strikes.

$$L_{avg} = \left(\sum_{i=1}^M L_i \right) / M \quad (3)$$

Average strike speed

The feature *average strike speed* (SP_{avg}) is determined as follows. We first calculate speed of individual strikes by dividing the strike length with the difference between the down and up action times. The sum of the speed of all the strikes is divided by the total number of strikes to

obtain the average strike speed, as shown in Eq. 4 where SP_i is the speed of the i^{th} strike and M is the total number of strikes.

$$SP_{avg} = \left(\sum_{i=1}^M SP_i \right) / M \quad (4)$$

Total delay

Delay is the time lag between the completion of the current touch action (strike or tap) and the starting of the next one. We can determine the delay between two consecutive touch actions by taking the difference between the up action time (i.e., finishing) of the current touch action and the down action time (i.e., starting) of the next touch action. We add up all these values for all the touch actions to get the value of the feature *total delay* (D_{tot}), shown in Eq. 5. D_i is the delay between the i^{th} and $(i+1)^{th}$ touch actions. M and N are the total number of strikes and taps, respectively.

$$D_{tot} = \left(\sum_{i=1}^{M+N-1} D_i \right) \quad (5)$$

Average delay

We compute the *average delay* (D_{avg}) feature value by simply dividing D_{tot} (see Eq 5) by the total number of strikes and taps (Eq. 6).

$$D_{avg} = \left(\sum_{i=1}^{M+N-1} D_i \right) / (M + N - 1) \quad (6)$$

Turnaround time

The feature *turnaround time* ($T_{turnaround}$) is the total time taken to complete a task. It is calculated by subtracting the down time of the first touch action of the task from the up time of the last touch action of the task, shown in Eq. 7. $T_{u(M+N)}$ is the finger up time of $(M+N)^{th}$ (i.e., last) touch action and T_{d1} is the finger down time of first touch action. M and N are the number of strikes and taps made to perform the task, respectively.

$$T_{turnaround} = T_{u(M+N)} - T_{d1} \quad (7)$$

The *mode* features: mode of strike length, strike speed and delay

In addition to the above seven features, we propose three more features based on the notion of *mode*. The idea of *mode* is conceptually similar to the statistical mode. We calculate the *densest region* in the given

information/data. In our work, we calculated mode features in the following way.

We first calculate the range of the given data (by subtracting minimum from maximum value). Next, we divide the range into very small sub-ranges called *chunks*. Each and every element in the data must belong to one of the chunks. We then form a *window* with a group of chunks (e.g., window of size 2 indicates two chunks in the window). Once a window is defined, we perform *sliding window* operation over the range of values. In each position of the window (covering a group of chunks equal to its size), we determine the number of elements present in the window as well as the *total sum* of all those element values. The *average window value* is computed by dividing the total sum by the number of elements. For a given window size, we determine the densest region as the one having the largest average value. We repeat this process for various window sizes, starting from 1 to 10 % of the range. We determine the densest region among all this window sizes. The average window value of that region is chosen as the mode value.

The three mode features we used in our work are: *mode of strike length*, *mode of strike speed* and *mode of delay*. For each of these feature, the data range represent the corresponding feature values. We calculated mode of these data ranges using the above procedure.

The ten feature values along with their units of measurement, considering typical mobile touch screen dimensions, are shown in Table 1. The last column of the table indicates the notations we used to denote the features in subsequent discussions.

We are using these features based on the assumption that the emotional state has a role to play in inducing error behaviour, the speed of strokes, the delay between two consecutive touch actions or the total task completion time. For example, a person in an excited state is more likely to make some errors than someone in a calm state. Therefore, we can expect to have different values

Table 1 The ten features with their units of measurement. The last column shows the notations we used for these feature

| Feature | Unit | Notation |
|--------------------------------|--------------|----------|
| Deviation in number of strikes | | F1 |
| Deviation in number of taps | | F2 |
| Mode of strike length | Millimeter | F3 |
| Average strike length | Millimeter | F4 |
| Mode of strike speed | Meter/second | F5 |
| Average strike speed | Meter/second | F6 |
| Mode of delay | Millisecond | F7 |
| Average delay | Millisecond | F8 |
| Total delay | Second | F9 |
| Turnaround time | Second | F10 |

for these features for different emotional states. In other words, the features act as indirect cue (indicator) to the emotional states.

It may be noted that the finger strike or the dragging gesture is manifested in two forms: slower “swipe” and faster “flick/fling.” Generally it is assumed that the gesture depends on the task: some tasks such as web page scrolling are performed with swipe whereas flicks are suitable in certain interaction such as navigating long lists. However, this behaviour may not hold true in affective interaction; a frustrated user may use flicks for web page scrolling while a happy user may prefer to use swipe to leisurely navigate a long list. Hence, we have not considered the task-dependency of the finger strike features (speed and length) in this work. Similarly, it is possible in real usage scenario that the delay between two touch actions is affected by attention shifts and other external factors. However, we are ignoring such effects in our calculation of total delay in this work.

We have experimented with two different approaches for building the predictor: a machine-learning based classification approach and a linear regression approach for curve fitting. Both the approaches were based on empirical data. Details of the empirical data collection are discussed in the following section.

Empirical data collection

We collected touch interaction data from a total of 57 volunteers. Data of 36 participants were used as training data. The rest of the data (from 21 participants) were used for model testing.

Experimental set up

The data were collected using 7” tablets (Akash™), running on the Android OS, version 4.0.3 (Gingerbread). We had developed an Android app for data collection. The app was developed in Eclipse™ using the Android Development Kit (ADK). The app contained seven general tasks, which required finger strikes and taps to

execute. The tasks are shown in Table 2. We also estimated the minimum number of strikes and taps required to perform each task, which are also shown. The tasks were chosen since they represented typical functionalities of touch devices.

Usually during touch interaction, we have to choose the desired app icon from a set of app icons on the screen. Sometimes, the required icon may not be on the current screen. In that case, we have to change screen. In order to mimic this behaviour, we designed our app with four screens. In each screen, 15 icons were shown in grid view with a total of 60 icons. Among them, only seven icons were activated for the tasks and the rest were dummy icons. Few task icons were placed on the first (main) screen itself. However, icons for some of the tasks could be accessed only through screen change. Figure 1 shows the first (main) screen (Fig. 1(a)) and the interface for one of the tasks (draw pattern, Fig. 1(b)). The app captured and stored finger down time, up time and strike length in a log file. The event times were recorded using the function `elapsedRealtime()`. A portion of the log file is shown in Fig. 2, for illustration.

Participants

We selected a group of 57 male participants in the age group of 20 to 26. They were under graduate and post graduate students. The participants were chosen on the basis of their familiarity with the touch devices. We divided them into two groups: *training group* having 36 participants and *testing group* having 21 participants. Participants’ data from the first group was used to train the model. The other group data was used for model testing.

Each group was further divided into three sub groups corresponding to the three emotional states. Participants belonging to a sub group provided data for the corresponding emotion only. The division was equal in the training group. Thus, we had 12 participants each in the three sub groups of the training group. In the testing

Table 2 The seven tasks along with the minimum numbers of strikes and taps

| Serial No | Task | Description | Min no of Strikes | Min no of Taps |
|-----------|--------------------------------|--|-------------------|----------------|
| 1 | Setting a reminder | Set a reminder for a specific date and time. Also type a reminder message ('bday') | 3 | 7 |
| 2 | Play a song | There is a list of 30 songs in alphabetical order and user has to play a particular song, which is 5 th from the last | 2 | 2 |
| 3 | Type keyword on message screen | Type 'hello' word on message screen and select the send button | 1 | 7 |
| 4 | Calling a contact | There is a list of 25 contacts in phone directory and user has to call one of them, which is 5 th from the bottom of the list | 3 | 2 |
| 5 | Drawing task | Draw a pattern 'THE IIT LIFE' on a blank canvas | 29 | 1 |
| 6 | Edit a contact | There is a list of 25 contacts in phone directory and user has to edit one contact which is the second last in the list | 2 | 7 |
| 7 | Type URL | Type 'google' on URL box and select the search button | 2 | 8 |

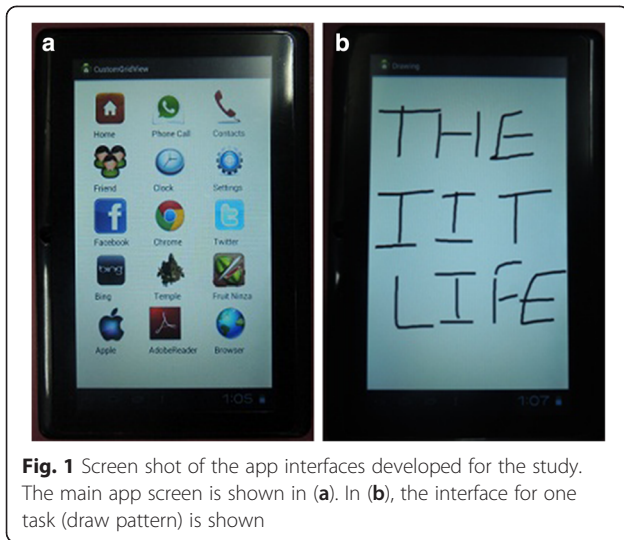


Fig. 1 Screen shot of the app interfaces developed for the study. The main app screen is shown in (a). In (b), the interface for one task (draw pattern) is shown

group, we had 7 participants in each of the positive, negative and neutral sub group.

Procedure

In order to collect data, we disabled network connections in the devices (both mobile data network and wi-fi network). The data collection study was divided into the following five stages.

- a) Training
- b) Intentional emotion changing
- c) Self-assessment questionnaire
- d) Actual data collection
- e) Self-assessment questionnaire

We trained each participant to make him familiar with the device. During the training session, participants’ were familiarized with the app also. The app training included introducing the participants to the active task icons, the steps required to locate those icons in the four screens and the steps for executing the seven tasks. They were given some dummy tasks to perform for the purpose. Training sessions lasted for about 10–15 min. Each participant was provided with a volunteer ready to help at any stage.

In order to collect data for positive and negative states (from those participants whom we put into those sub groups), we used a method to bring a participant to one of these states. Usually, it is difficult to change one’s mental state from positive to negative or vice versa. Changing the mental state from neutral to positive or neutral to negative is much easier. Therefore, we performed an initial screening of the participants and selected those whose mental states were likely to be neutral. We interviewed them about the activities they performed prior to coming for the test. Based on their responses, we made the judgement. For example, if a participant said that he was playing football and scored a difficult goal, he was likely to be in an excited state. Therefore, we decided not to collect data from him at that point of time. On the other hand, if someone reported to be sleeping, we assumed him to be in a neutral state and included him in the study.

For taking the user to a particular emotional state, we defined some intentional emotion changing dummy tasks. A participant took around thirty to forty minutes to carry out the tasks. We informed the participants

| Action | DownTime | Action | UpTime | Lenght | OfStrike |
|---------------|----------|---------------|--------|-----------|----------|
| 1382117230319 | | 1382117230708 | | 300.8802 | |
| 1382117231887 | | 1382117232266 | | 216.60188 | |
| 1382117234487 | | 1382117234571 | | 0.0 | |
| 1382117234773 | | 1382117234774 | | 0.0 | |
| 1382117239234 | | 1382117239237 | | 0.0 | |
| 1382117240208 | | 1382117240212 | | 0.0 | |
| 1382117240552 | | 1382117240555 | | 0.0 | |
| 1382117241109 | | 1382117241113 | | 0.0 | |
| 1382117252184 | | 1382117252519 | | 167.0142 | |
| 1382117255332 | | 1382117255555 | | 236.18565 | |
| 1382117257006 | | 1382117257122 | | 0.0 | |
| 1382117265120 | | 1382117265493 | | 162.90895 | |
| 1382117273014 | | 1382117273018 | | 0.0 | |
| 1382117273788 | | 1382117273794 | | 0.0 | |
| 1382117274681 | | 1382117274685 | | 0.0 | |
| 1382117274786 | | 1382117274790 | | 0.0 | |
| 1382117274966 | | 1382117274971 | | 0.0 | |

Fig. 2 A portion of the log file generated during the empirical study

beforehand that the tasks were likely to trigger a change in emotional state and obtained their informed consent.

We brought a participant into a positive emotional state by showing them funny videos, comedy videos and comedy scenes from YouTube. The videos were selected on the basis of viewers' rating. These tasks were expected to induce happy emotion in the participants, which is a part of the positive emotional state. We also had set some SGT puzzles², in which we had chosen the Untangle, to make participants *excited*, another positive emotion. In addition, we made use of some inspirational videos and motivational tasks to trigger the elated emotion in the participants.

We used a *slow response* device to change participants' emotional state to negative. On this device, we had defined a task that took nearly 30 min to complete. Due to its slow response, the participants were expected to become angry and frustrated, triggering a negative emotional state. We also had shown some explicit videos related to poverty, malnutrition and post war trauma, to induce the negative emotions (sadness, anger, fear and disgust) among the participants. Participants were also asked to get 25 points within 60 s playing the SGT puzzle, which were beyond their ability. This was expected to trigger the negative emotional state (anger and frustration).

The previous step was followed by a self-assessment questionnaire. It helped us deciding whether that particular participant had come into the desired emotional state. There were eight Yes/No type questions in the self-assessment questionnaire, which are listed below.

1. Are you happy or excited?
2. Are you enjoying?
3. Do you want to leave this room?
4. Are you interested to do the same thing you did in the last 20 min?
5. Would you like to listen to joke?
6. Would you like to solve puzzle?
7. Would you like to sing a sad song?
8. Are you sad, angry or frustrated?

We determined the effectiveness of the emotion changing tasks by evaluating the answers of the above questions. The questions 1, 2, 4, 5 and 6 focused on positive aspects of mind like happiness, excitement, enjoyment and affinity towards a rewarding thing. If a participant answered YES to all these questions and NO to the other questions, we concluded that he was in the positive emotional state. Similarly, YES to the questions 3, 7 and 8 and NO to the other questions by a participant led us to conclude his emotional state to be negative. We repeated the emotion changing tasks if the self-assessment questionnaire did not indicate a change in

the participant's emotion to the desired state. Most of the participants answered the questions "correctly" in the first attempt itself. For 20 odd participants, we had to repeat the process (once for 13, twice for 4, thrice for the remaining 3 participants). In fact, we had to eliminate 3 volunteers from the experiments altogether, since we couldn't bring them to particular state even after repeated attempts.

In the fourth stage, we asked the users to perform the seven tasks on the tablet in a single session, which took between 3–5 min for a participant to complete. The order of the tasks was counterbalanced (changed for each participant) to take into account the learning effects (if any). After the data collection stage, we administered the same self-assessment questions to the participants, to ensure that the emotional states during data collection were the desired ones. In case the questionnaire indicated a change in emotion, we repeated the whole data collection steps again for the particular participant.

Data

Our app captured the touch interaction data of the participants by logging the down action time (in millisecond), up action time (in millisecond) and the finger stroke length (in millimeter). We determined strike and taps from the stroke lengths as follows. Ideal length of a stroke in case of a tap should be zero. However, due to minor slip of finger while making a tap, there may be slight movement and the stroke length becomes non zero. In this work, we considered a finger stroke to be a strike if its length was greater than 5.0 mm; otherwise, it was considered a tap and its length is set to zero programmatically, as can be seen in the log file sample of Fig. 2. From the strike and tap information, we computed the ten feature values for each participant. The feature values computed from the logged data for the participants in the training group are shown in Tables 3, 4 and 5. In Table 6, the feature values computed for the test group of 21 participants are shown.

We used the data to build a three-class classifier. We compared performance of four classification models and found maximum prediction accuracy with the k-means clustering based classifier (72 %). In order to improve the accuracy further and reduce the computational complexity, we experimented with a linear regression approach and derived a model with 90.47 % accuracy. The two approaches and the final model are described in the following.

Proposed model

The classification approach

In order to obtain a classifier as predictor of emotional state, we compared the results of four classification models: support vector machine (SVM), maximum

Table 3 The feature values in the positive emotion sub group

| Features | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 | F9 | F10 |
|--------------|----|----|--------|--------|------|------|--------|--------|--------|--------|
| Participants | | | | | | | | | | |
| P1 | 19 | 5 | 201.35 | 62.37 | 4.17 | 1.09 | 332.49 | 267.3 | 282.62 | 464.41 |
| P2 | 21 | 5 | 121.57 | 64.22 | 2.38 | 1.67 | 137.5 | 147.86 | 119.62 | 145.94 |
| P3 | 8 | 6 | 205.3 | 78 | 2.41 | 0.97 | 278.58 | 181.76 | 208.94 | 259.62 |
| P4 | 17 | 3 | 129.03 | 88.45 | 4.27 | 2.74 | 272.21 | 309.26 | 220.49 | 252.61 |
| P5 | 27 | 3 | 125.95 | 49.58 | 3.91 | 1.71 | 342.86 | 234.97 | 312 | 459.35 |
| P6 | 10 | 5 | 146.1 | 96.31 | 2.99 | 1.11 | 215.16 | 530.77 | 163.52 | 174.2 |
| P7 | 22 | 3 | 157.49 | 53.7 | 4.47 | 1.94 | 224.07 | 274.35 | 192.7 | 305.74 |
| P8 | 21 | 7 | 134.62 | 56.82 | 4.36 | 1.28 | 278.35 | 311.25 | 247.73 | 283.56 |
| P9 | 7 | 5 | 180.06 | 360.04 | 2.05 | 2.16 | 112.29 | 173.23 | 81.97 | 123.59 |
| P10 | 7 | 11 | 182.38 | 106.3 | 1.87 | 0.62 | 95.55 | 136.75 | 75.49 | 118.91 |
| P11 | 10 | 18 | 165.14 | 98.39 | 1.89 | 1.18 | 94.37 | 157.25 | 83.99 | 113.77 |
| P12 | 8 | 9 | 178.76 | 97.48 | 1.86 | 1.3 | 118.37 | 163.42 | 92.33 | 129.35 |

entropy model, conditional random fields (CRF) and the K-means clustering. We had chosen these four as they are very common approaches used in classification problems. The training data set (of 36 feature vectors) was used to train each of these models. The prediction accuracy was estimated using the 21 test vectors.

We used the WEKA tool to implement the k-means clustering model. As we have mentioned, there are more than one emotional state in each of the three states. Therefore, we decided to use multiple clusters for each state. Since the negative state contains the maximum of four emotions (sad, anger, fear and disgust), we decided to use four clusters in each state, with a total of twelve such clusters. The model works as follows: For each cluster, we determined a centroid. Each centroid is a point in a ten dimensional feature space. When we get a

(new) feature vector (a point in the ten dimensional space), we compute the Euclidean distance of the point from all the centroids. The point belongs to the cluster whose centroid is the *closest* to the input feature vector. Depending on the affective state (positive, negative or neutral) in which the cluster belongs to, we classify the point. The 12 centroids obtained for the model are shown in Table 7.

In the conditional random field model, we used exact inference technique to calculate the *marginals* in each iteration owing to the small data set. We updated the parameters until convergence. The model was used to generate the conditional probability of the possible labels of each data instance in the test data set based on its features. Since we have 3 possible labels per data instance (positive, negative and neutral), we took the highest

Table 4 The feature values for the negative emotion sub group

| Features | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 | F9 | F10 |
|--------------|----|----|--------|--------|------|------|--------|--------|--------|--------|
| Participants | | | | | | | | | | |
| P1 | 6 | 4 | 179.44 | 154.71 | 3.18 | 4.46 | 188.19 | 283.13 | 133.61 | 201.17 |
| P2 | 7 | 3 | 145.12 | 30.09 | 5.48 | 1.61 | 166.69 | 413.72 | 118.35 | 197.55 |
| P3 | 16 | 4 | 163.91 | 60.41 | 2.39 | 1.31 | 230.1 | 193.23 | 157.67 | 186.38 |
| P4 | 7 | 5 | 116.49 | 60.23 | 4.81 | 3.64 | 174.78 | 153.12 | 127.59 | 170.36 |
| P5 | 12 | 3 | 109.26 | 55.48 | 3.6 | 1.22 | 123.31 | 191.37 | 93.72 | 185.87 |
| P6 | 7 | 3 | 174.33 | 89.28 | 2.14 | 2.47 | 150.1 | 173.08 | 106.57 | 152.29 |
| P7 | 6 | 3 | 207.99 | 118.06 | 2.29 | 2.47 | 208.03 | 182.38 | 145.62 | 170.5 |
| P8 | 6 | 4 | 197.38 | 325.94 | 2.76 | 1.23 | 153.63 | 165.21 | 109.78 | 163.48 |
| P9 | 16 | 3 | 180.69 | 271.23 | 4.45 | 1.24 | 129.34 | 393.37 | 103.47 | 194.34 |
| P10 | 17 | 9 | 137.42 | 78.54 | 3.47 | 2.27 | 125.65 | 335.23 | 109.33 | 188.86 |
| P11 | 21 | 9 | 185.68 | 74.76 | 4.84 | 1.57 | 137.24 | 363.27 | 124.89 | 302.05 |
| P12 | 10 | 17 | 162.29 | 97.24 | 2.72 | 2.15 | 161.44 | 141.87 | 142.51 | 163.19 |

Table 5 The feature values for the neutral emotion sub group

| Features | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 | F9 | F10 |
|--------------|----|----|--------|--------|------|------|--------|--------|--------|--------|
| Participants | | | | | | | | | | |
| P1 | 16 | 4 | 174.09 | 87.46 | 3.17 | 1.61 | 170.44 | 193.63 | 138.05 | 229.11 |
| P2 | 13 | 6 | 158.82 | 64.45 | 2.9 | 1.19 | 148.71 | 316.75 | 118.97 | 145.18 |
| P3 | 14 | 11 | 162.18 | 86.5 | 3.69 | 1.41 | 165.22 | 369.27 | 142.09 | 182.02 |
| P4 | 6 | 3 | 159.3 | 143.11 | 2.63 | 1.77 | 161.61 | 181.93 | 113.12 | 140.35 |
| P5 | 9 | 3 | 116.17 | 84.29 | 4.48 | 4.56 | 142.8 | 199.75 | 104.25 | 172.37 |
| P6 | 12 | 7 | 153.56 | 68.64 | 2.89 | 1.38 | 151.32 | 287.36 | 121.06 | 155.36 |
| P7 | 16 | 3 | 175.87 | 303.04 | 3.28 | 1.15 | 202.63 | 332.45 | 132.1 | 164.88 |
| P8 | 8 | 4 | 158.46 | 139.83 | 2.53 | 1.98 | 156.23 | 174.65 | 114.05 | 152.63 |
| P9 | 7 | 16 | 132.01 | 67.59 | 2.85 | 0.63 | 101.61 | 212.14 | 85.35 | 122.31 |
| P10 | 13 | 5 | 148.69 | 74.82 | 3.61 | 4.12 | 120.08 | 175.65 | 94.86 | 190.83 |
| P11 | 13 | 3 | 197.52 | 50.17 | 2.49 | 1.22 | 240.13 | 172.42 | 184.9 | 240.63 |
| P12 | 8 | 7 | 104 | 113.27 | 3.82 | 2.62 | 121.35 | 272.74 | 92.22 | 129.33 |

probability of the three, and assigned the corresponding label to that test data instance. We used the Scikit-Learn tool in Python which provided the libraries for different SVM models. We compared the results of four SVM classifiers including (a) the linear SVM, (b) the kernel linear SVM, (c) the kernel poly SVM and (d) the kernel

RBF. In order to develop the maximum entropy model, we used the Stanford classifier tool.

The results of the application of the different classifiers on the training and test data are summarized in Table 8.

As the table shows, the K-means clustering based classifier gives the highest accuracy figure of 72 %. Therefore,

Table 6 The feature values for the testing group. Each row represents the ten feature values

| Features | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 | F9 | F10 |
|--------------|----|----|--------|--------|------|------|--------|--------|--------|--------|
| Participants | | | | | | | | | | |
| P1 | 15 | 5 | 154 | 80.64 | 2.63 | 1.92 | 141.48 | 312.42 | 114.89 | 181.23 |
| P2 | 9 | 5 | 145.61 | 109.17 | 1.83 | 1.77 | 165.62 | 246.36 | 124.18 | 160.29 |
| P3 | 8 | 8 | 138.02 | 77.24 | 1.82 | 2.1 | 200.25 | 351.34 | 154.04 | 209.81 |
| P4 | 13 | 3 | 189.02 | 173.28 | 2.37 | 1.14 | 152.02 | 190.26 | 124.66 | 192.17 |
| P5 | 12 | 7 | 145.01 | 89.13 | 3.05 | 3.41 | 143.42 | 259.76 | 114.73 | 186.16 |
| P6 | 11 | 5 | 152.23 | 63.21 | 2.75 | 3.47 | 134.66 | 194.93 | 130.71 | 181.19 |
| P7 | 13 | 6 | 146.13 | 88.86 | 3.26 | 3.36 | 141.99 | 267.56 | 116.48 | 179.26 |
| P8 | 10 | 6 | 143.97 | 107.24 | 1.81 | 1.79 | 164.83 | 249.46 | 129.33 | 179.78 |
| P9 | 8 | 3 | 148.23 | 76.71 | 2.22 | 2.52 | 126.66 | 136.61 | 91.19 | 120.18 |
| P10 | 5 | 3 | 136.17 | 112.7 | 2.24 | 2.22 | 137.92 | 144.88 | 95.17 | 120.49 |
| P11 | 11 | 3 | 146.18 | 93.5 | 2.63 | 1.61 | 130.09 | 139.81 | 97.56 | 165.4 |
| P12 | 7 | 4 | 146.71 | 79.14 | 2.17 | 2.56 | 124.98 | 138.52 | 92.93 | 188.57 |
| P13 | 15 | 10 | 153.76 | 90.29 | 1.98 | 1.31 | 97.11 | 108.66 | 83.51 | 135.46 |
| P14 | 15 | 16 | 172.86 | 93.58 | 2.03 | 2.41 | 102.68 | 142.15 | 94.46 | 138.53 |
| P15 | 11 | 3 | 174.37 | 112.48 | 1.88 | 1.18 | 137.54 | 196.25 | 103.15 | 139.78 |
| P16 | 5 | 3 | 139.66 | 74.26 | 2.19 | 2.06 | 85.85 | 129.66 | 59.23 | 121.38 |
| P17 | 9 | 3 | 129.58 | 70.76 | 3.06 | 3.01 | 124.65 | 138.05 | 90.93 | 147.24 |
| P18 | 17 | 3 | 169.58 | 47.85 | 4.65 | 1.83 | 153.5 | 487.71 | 124.33 | 263.15 |
| P19 | 19 | 3 | 211.25 | 380.7 | 2.65 | 0.7 | 247.35 | 60.56 | 205.3 | 227.29 |
| P20 | 20 | 3 | 157.99 | 57.44 | 3.18 | 1.08 | 185.48 | 229.66 | 155.8 | 180.01 |
| P21 | 15 | 4 | 134.13 | 66.35 | 4.32 | 5.24 | 156.22 | 250.23 | 124.97 | 258.37 |

Table 7 The four centroid values for each feature in the three affective states

| Emotion State | Centroid | F1 | F2 | F3 | F4 | F5 | F6 | F7 | F8 | F9 | F10 |
|---------------|------------|------|------|-------|-------|------|------|-------|-------|-------|-------|
| Negative | Centroid 1 | 8.5 | 5.8 | 178.2 | 102.2 | 2.49 | 2.54 | 181.9 | 189.8 | 137.4 | 184.7 |
| | Centroid 2 | 13 | 4.5 | 143.1 | 108.8 | 4.25 | 1.58 | 136.2 | 333.4 | 106.2 | 191.6 |
| | Centroid 3 | 7 | 5 | 116.4 | 60.2 | 4.81 | 3.64 | 174.7 | 153.1 | 127.5 | 170.3 |
| | Centroid 4 | 21 | 9 | 185.6 | 74.7 | 4.84 | 1.57 | 137.2 | 363.2 | 124.8 | 302 |
| Positive | Centroid 1 | 7.5 | 7.7 | 186.5 | 166.7 | 2.11 | 1.47 | 151.2 | 163.6 | 114.6 | 157.8 |
| | Centroid 2 | 17.5 | 4.5 | 141.8 | 83.6 | 3.94 | 2.15 | 247.4 | 356.4 | 206.1 | 254 |
| | Centroid 3 | 15.5 | 11.5 | 143.3 | 81.3 | 2.14 | 1.43 | 115.9 | 152.5 | 101.8 | 129.6 |
| | Centroid 4 | 23 | 4 | 163.6 | 55.9 | 4.04 | 1.4 | 337.6 | 251.1 | 297.3 | 161.8 |
| Neutral | Centroid 1 | 10 | 5 | 122.9 | 90.7 | 3.97 | 3.77 | 128 | 215.9 | 97.1 | 164.1 |
| | Centroid 2 | 7 | 7.6 | 149.9 | 116.8 | 2.67 | 1.46 | 139.8 | 189.5 | 104.1 | 138.4 |
| | Centroid 3 | 13 | 3 | 197.5 | 50.1 | 2.49 | 1.22 | 240.1 | 172.4 | 184.9 | 140.6 |
| | Centroid 4 | 14 | 6.2 | 164.9 | 122 | 3.19 | 1.35 | 167.6 | 299.8 | 136.4 | 175.3 |

we believe that the clustering model may be a promising approach in relation to the affective state prediction problem. This may be due to the nature of the data: the data were inherently clustered around the emotional sub states in each broader state. However, the accuracy figure, although reasonably high, is not good enough for practical usage. Therefore, we came up with an alternate model, which is described next.

The linear regression approach

The alternative model we propose is a linear combination of seven out of the ten feature values (obtained by excluding the three mode features from the ten features). In order to develop the model, we assumed that the feature values are mapped along the x-axis and the emotional state of an individual along they-axis in a two-dimensional space. For the three states, we assigned three distinct ranges of values along the y-axis. We then established a linear relationship between the features and the emotional states, using the linear regression technique.

The linear relation assumes the following form (Eq 8), for each feature f_i :

$$y_i = A + B * x_i \tag{8}$$

Where x_i represents the feature value for that particular feature for a given participant in his/her respective emotional state, and y_i represents a unique real number within the range for a specific state. We used Eqs 9 and 10 to compute the constants A and B.

$$B = \frac{\sum_{i=1}^n (x_i - \bar{x}) / (y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \tag{9}$$

$$A = \bar{y} - B * \bar{x} \tag{10}$$

Here, \bar{x} and \bar{y} represent the mean of the data points x_i and y_i respectively. The number of data points is denoted by n.

We propose three different relations for each of the positive, negative and neutral states. These relations constitute the components of the model. The relationships were established with the following approach.

We start by taking individual features one at a time for a particular emotional state. For example, let us consider the feature *deviation in number of strikes* for the affective state positive. For this feature, we plotted the data points and applied Eqs 9 and 10 to determine the unknown constants A and B (see Eq 8). This procedure is repeated for each of the remaining features. In this way, we obtained a set of seven linear equations, one for each feature for each of the emotional state. Next, we combined these seven equations to come up with a single equation, of the form shown in Eq 11, for a particular emotional state.

$$y = K_1 + K_2 * \sum_{i=1}^7 B_i * f_i \tag{11}$$

In Eq 11, K_1 and K_2 are constants and B_i is the weight associated with feature f_i (see Eq 8). In order to obtain

Table 8 Comparative results on prediction accuracy of the seven classifiers

| Classification Model | Accuracy |
|---------------------------|----------|
| K-Means clustering | 72 % |
| Linear SVM | 40 % |
| Kernel Linear SVM | 36 % |
| Kernel Poly SVM | 52 % |
| Kernel RBF SVM | 44 % |
| Conditional Random Fields | 44 % |
| Maximum Entropy | 36 % |

K_1 , we add all the A_i values for each feature (see Eq 8) and the sum is divided by 7, the number of features. K_2 is simply the ratio of (1/7), i.e., the ratio of one and the number of features. We then obtain the final form of the model by adding to the right hand side of the Eq 11 a constant value, which is unique to a particular emotional state. Thus, we obtain three different linear relations, one for each of the three emotional states, as shown in Eqs 12, 13 and 14.

$$y_{pos} \in [POS] = K_1 + K_2 * \sum_{i=1}^7 B_i * f_i + C_{POS} \quad (12)$$

$$y_{neg} \in [NEG] = K_1 + K_2 * \sum_{i=1}^7 B_i * f_i + C_{NEG} \quad (13)$$

$$y_{neut} \in [NEU] = K_1 + K_2 * \sum_{i=1}^7 B_i * f_i + C_{NEU} \quad (14)$$

In the above set of equations, there are several parameters. These include the numeric ranges for each emotional states ([POS] for positive emotion, [NEG] for negative emotion and [NEU] for neutral emotion) and the constants A's, B's and C's. These parameters were determined from the empirical training data of 36 participants.

Our first objective was to identify the numeric ranges that characterize each emotional state, with the following constraint.

- [1]The ranges should not be overlapping.
- [2]The ranges should be such that any feature vector (the set of seven feature values) can map to only one of the three ranges.

We performed several trials and errors with various ranges to fit the empirical data subject to the satisfaction of the constraints. The empirical data consisted of three sets of feature vectors for the three emotional state, each set having 12 vectors. We had chosen the three unique non-overlapping ranges that satisfied the constraints and gave the closet match with the empirical data. The three ranges we obtained through this approach are as follows.

- [1]Positive emotion range: [50, 105]
- [2]Negative emotion range: [1, 12]
- [3]Neutral emotion range: [25, 36]

In order to obtain the other parameters (A's, B's and C's of Eqs. 12, 13 and 14), we used an elaborate assignment based approach. Let us consider the positive emotional state for illustration. We had 12 feature vectors corresponding to 12 participants for this state. The emotion of each of the participants was assigned a number

between [50, 105] (i.e., the positive range). The assignment started with 50, separated by 5 and ended at 105. For example, if one participant is assigned the value 50, the next is assigned 55, the next 60 and so on till 105. With one set of such assignment, we estimated the constants A's, B's and C's, through linear regression. We then reassigned the numbers to the participants (e.g., the participant who was assigned 50 was re-assigned 55 and so on) and re-estimated the constants. The process we repeated for 12 times, corresponding to the 12 participants. The final values of A's, B's and C's were obtained by taking average of all the estimated 12 values. We applied similar procedure to determine the constants for the other two emotional states also. Only minor difference was that the assignment of numbers was separated by 1 in those cases, rather than 5. The final estimated values are shown in Table 9.

The constant K_1 is sum of all the A values of Table 8 divided by 7, which is 13.91. The constant K_2 is the ratio 1/7 or 0.14. The constants C's of Eqs. 12, 13 and 14 were estimated as 77.5, 6.5 and 30.5 for positive, negative and neutral emotions respectively. It may be noted in Table 8 that the B_i values for the features average delay, total delay and turnaround time are close to zero. Hence, we can ignore the corresponding product terms (B_5f_5 , B_6f_6 and B_7f_7) in Eqs. 12, 13 and 14. Thus, we obtained Eqs. 15, 16 and 17 as our final proposed model, where B_i values are taken from Table 8 for the corresponding f_i .

$$y_{pos} = 13.91 + 0.14 * \sum_{i=1}^4 B_i * f_i + 77.5 \quad (15)$$

$$y_{neg} = 13.91 + 0.14 * \sum_{i=1}^4 B_i * f_i + 6.5 \quad (16)$$

$$y_{neut} = 13.91 + 0.14 * \sum_{i=1}^4 B_i * f_i + 30.5 \quad (17)$$

After we obtained Eqs 15, 16 and 17, we recomputed the y values for the given feature vectors and refined the emotion ranges. The refined ranges we obtained are as follows.

Table 9 The A and B values for the regression model

| Feature | Notation | A | B |
|---------------------------|----------|--------|-------|
| Deviation in strokes | f_1 | 34.58 | 0.25 |
| Deviation in taps | f_2 | 36.26 | 0.54 |
| Average length of strokes | f_3 | 21.23 | -0.02 |
| Average speed of strokes | f_4 | -26.54 | 91.9 |
| Average delay | f_5 | 20.13 | ≈0.0 |
| Total delay | f_6 | 15.59 | ≈0.0 |
| Turnaround time | f_7 | 15.59 | ≈0.0 |

- [1]Positive emotion range: [97.38, 102.21]
- [2]Negative emotion range: [25.38, 30.77]
- [3]Neutral emotion range: [48.99 to 53.89]

Therefore, our proposed model consists of Eqs 15, 16 and 17 along with the three refined ranges mentioned above.

In order to ascertain the validity of the proposed model, we used the test data set of 21 participants. We used the model to predict the state of the corresponding participant. We matched this data with the (known) emotional state of the participant. The results obtained are shown in Table 10. As can be seen, the proposed model is able to correctly predict 19 out of 21 cases. Hence the accuracy of the model was found to be 90.47 %.

Discussion

In this work, we attempted to propose a model to predict the emotional state of a touchscreen user. We attempted to do that from a set of ten features. These features, in turn, are based on the finger strokes on a touch screen. We assumed that the finger strokes are indicators to the affective state of the user. Hence, the stroke information can be used to infer emotion. We considered the classification approach and the regression

analysis approach to obtain the model. Although we considered many popular classification models, we obtained models with very low or moderate prediction accuracy. The results indicate that the classification approach may not be suitable for developing an affective state predictor of mobile touch screen users. On the other hand, the regression approach resulted in a model with high accuracy (90.47 %). The high accuracy indicates the suitability of the model for practical use. Therefore, we propose to use the linear regression model for predicting emotion of touch screen users.

There are many ways in which the model can be used. Once we are able to detect user’s emotional state from his interaction with the device, we can change the look and feel of the interface, to complement the emotional state. We can also make changes in the way tasks are performed depending on the current state of user emotion. This may lead to “polite” interfaces, which are “empathic” or “sympathetic”. Such qualities, in turn, are expected to improve usability and enhance user experience. We plan to work in this direction in future.

As our literature survey shows, most of the works on emotion detection rely on additional set-up and significant computations. In contrast, recognition of users’ affective state from indirect cues does not need extra hardware or increased computation. We are primarily

Table 10 Results of model validation

| Sl no. | Emotional state observed | Predicted y value | Emotional state predicted |
|--------|--------------------------|-------------------|---------------------------|
| 1. | Positive | 99.1169 | Positive |
| 2. | Positive | 98.0373 | Positive |
| 3. | Positive | 97.2808 | None |
| 4. | Positive | 100.2060 | Positive |
| 5. | Positive | 101.6639 | Positive |
| 6. | Positive | 99.6846 | Positive |
| 7. | Positive | 98.2239 | Positive |
| 8. | Negative | 27.3548 | Negative |
| 9. | Negative | 29.6156 | Negative |
| 10. | Negative | 29.2849 | Negative |
| 11. | Negative | 28.8817 | Negative |
| 12. | Negative | 26.5327 | Negative |
| 13. | Negative | 26.8970 | Negative |
| 14. | Negative | 26.8970 | Negative |
| 15. | Neutral | 50.0243 | Neutral |
| 16. | Neutral | 49.7985 | Neutral |
| 17. | Neutral | 53.8127 | Neutral |
| 18. | Neutral | 54.2420 | None |
| 19. | Neutral | 49.1450 | Neutral |
| 20. | Neutral | 52.9539 | Neutral |
| 21. | Neutral | 51.4049 | Neutral |

interested in recognizing the emotional state of a mobile handheld touch screen user. For those devices, extra hardware or increased computational complexity may not be feasible, preferable or affordable. Hence, indirect cue based emotion detection is the desirable solution for such devices. We found few works in the literature that attempted to detect emotion from indirect cues. However, those were either meant for desktop interaction or designed for a limited context of use. Our work, on the other hand, does not assume any specific usage context. The proposed model works for any touch interaction task. Hence, our proposed model provides a novel and generalized solution to the problem of emotion detection for touch screen users.

While developing the model, we assumed single finger/touch gestures. As we know, many touch interactions are multi-gesture. Although such gestures also contain finger strikes in addition to taps, it is not clear if our proposed model will work in such scenario. We would like to work on this aspect of the model in future. In addition, the model is based on some more assumptions. First of all, we are assuming that the features are indirect cues to the emotional state. The high accuracy observed during the empirical validation indicates the validity of this assumption. We also started our work with ten features, assuming all are necessary for computing user's emotional state. However, comparison between the results of classification and linear regression approach indicate to the contrary. With all the ten features, we achieved the best classification results at 72 %. In contrast, we obtained above 90 % accuracy with lesser (seven) features. We removed the mode features since those were difficult to compute in comparison to the other seven features. The results demonstrate that reduction in feature set not necessarily leads to poor performance in terms of prediction. However, since the features are very simple to compute, any further reduction in feature set may not lead to significant reduction in computation. Hence, we didn't perform any further study to decide if the same prediction results can be obtained with even lesser number of features. A crucial assumption we made is that all emotional states can be clustered into three broader states (positive, negative and neutral). Although there are various models for representing emotions, as mentioned in the related works section, we think it is reasonable to consider emotions as belonging to any of the three classes. This assumption helps in the development of the simpler model of prediction, as compared to the more complex models that are found in the literature. The validation results give an indirect justification for this assumption. We also assumed that we can induce emotions following the method we used in the empirical study. This is another unique approach we adopted in our work. The model

parameters were developed from the empirical data collected following the steps that include the emotion evoking tasks and the self-assessment questionnaires. Since the validation results show high accuracy of prediction, we believe the particular approach is justified.

In summary, we believe that the assumptions are reasonable based on the empirical validation of the model. It may also be noted that the model was built by first obtaining the numeric ranges for each emotional state through trial and error. This is followed by regression analysis assuming those ranges to be correct. The initial ranges were finalized by applying the model on the training vectors. Empirical validation results demonstrate that the final ranges and the model obtained through this process are reliable predictors with high degree of accuracy. However, it may be necessary to base the justifications on more theoretical foundation before we can reach to any conclusion about the validity of those assumptions and the method we adopted for data analysis. We plan to work in this direction in future.

We achieved an accuracy of 90.47 % for our proposed model. Although we believe the model accuracy is reasonably high for practical use, we feel the accuracy can be improved further. It may be noted that the model was developed from empirical data. The data were collected from participants who represent homogeneous group (male, post graduate students with the age group of 22–26). Moreover, the number of participants was also modest (57). Training the model with data from users with varied profiles in terms of gender, age, cultural and educational background, familiarity with touch devices and so on may improve versatility of the model. A larger data set may also lead to higher accuracy. Therefore, we plan to perform more such studies in future.

Conclusion

We proposed an empirically derived linear model to predict the affective state of touch input users. The model is able to predict the user's affective state as one of the three classes: positive negative and neutral. The model parameters were estimated from empirical data. Empirical validation indicates high prediction accuracy of the proposed model.

In order to refine and improve the model further, we plan to work on the following problems in future.

- Theoretical justification for the various assumptions made in the model.
- Model refinement and validation with more empirical data from larger heterogeneous group of users.
- Design interface and interaction that “complements” user's emotional state.

Endnotes

¹The Economic times, May 2012.

²Simon Tatham's portable puzzle collection. <http://www.chiark.greenend.org.uk/~sgtatham/puzzles/>

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

SS carried out the literature survey part. SS and JNT contributed to the empirical data collection and analysis part. They also contributed to the development of the classification approach. Overall problem formulation, experiment design, data analysis and paper writing was done by SB. All authors read and approved the final manuscript.

Acknowledgment

We are indebted to the participants who agreed to provide empirical data. We are also thankful to the volunteers who helped us during data collection.

Declaration

I confirm that I have read SpringerOpen's guidance on competing interests and would like to confirm that none of the authors have any competing interests in the manuscript.

Received: 28 May 2015 Accepted: 13 October 2015

Published online: 11 November 2015

References

- AlZoubi, O., D'Mello, S., & Calvo, R. (2012). Detecting Naturalistic Expressions of Nonbasic Affect using Physiological Signals. *IEEE Transactions on Affective Computing*, 3(3), 298–310.
- Ballano, S., Hupont, I., Cerezo, E. & Baldassarri, S. (2011). Recognizing Emotions from Video in a Continuous 2D Space. In *Proc. INTERACT 2011*, Part IV, LNCS6949, 600–603.
- Bartneck, C., & Reichenbach, J. (2005). Subtle Emotional Expressions of Synthetic Characters. *International Journal of Human-Computer Studies*, 62, 179–192.
- Bianchi-Berthouze, N., & Kleinsmith, A. (2003). A categorical Approach to Affective Gesture Recognition. *Connection Science*, 15(4), 259–269.
- Camurri, A., Lagerlof, I., & Volpe, G. (2003). Recognizing Emotion from Dance Movement: Comparison of Spectator Recognition and Automated Techniques. *International Journal of Human-Computer Studies*, 59(1), 213–225.
- Cearreta, I. & Garay, N. (2011). Applying the *Affinto* Ontology to Develop a Text Based Emotional Conversation System. In *Proc. INTERACT 2011*, Part IV, LNCS6949, 479–482.
- Coutrix, C. & Mandran, N. Identifying emotions expressed by mobile users through 2D surface and 3D motion gestures. *Proc. 2012 ACM Conference on Ubiquitous Computing (UbiComp 12)*, pp 311–320, 2012
- Epp, C., Lippold, M., Mandryk, R.L. (2011). Identifying Emotional States using Keystroke Dynamics. In *Proc. CHI2011*, ACM Press, 715–724.
- Ekman, P. (1992). An Argument for Basic Emotions. *Cognition and Emotion*, 6(3/4), 169–200.
- Ekman, P., Sorenson, E. R., & Friesen, W. V. (1969). Pan-cultural Elements in Facial Displays of Emotion. *Science*, 164(3875), 86–88.
- Gao, Y., Bianchi-Berthouze, N. & Meng, H. (2012). What Does Touch Tell Us about Emotions in Touchscreen-based Gameplay? *ACM Transactions on Computer-Human Interaction (TOCHI)*, 19(4).
- Glowinski, D., Dael, N., Camurri, A., Volpe, G., Mortillaro, M., & Scherer, K. (2011). Toward a Minimal Representation of Affective Gestures. *IEEE Transactions on Affective Computing*, 2(2), 106–118.
- Hazlett, R.L. (2006). Measuring Emotional Valence during Interactive Experiences: Boys at Video game play. In *Proc. CHI2006*. ACM Press, 1023–1026.
- Hertenstein, M. J., Holmes, R., McCullough, M., & Keltner, D. (2009). The Communication of Emotion via Touch. *Emotion*, 9(4), 566.
- Hiraga, R. & Kato, N. (2006). Understanding Emotion through Multimedia. In *Proc. ASSETS 2006*. ACM Press, 141–148.
- Isbister, K., Hook, K., Sharp, M. & Laakso, J. (2006). The Sensual Evaluation Instrument: Developing an Affective Evaluation Tool. In *ProcCHI 2006*, 1163–1172. ACM.
- Kapur, A., Kapur, A., Virji-Babul, N., Tzanetakis, G. Driessen, P.F. (2005). Gesture-based Affective Computing on Motion Capture Data. *Affective Computing and Intelligent Interaction*, LNCS 3784, 1–7, Springer.
- Katsyri, J., & Sams, M. (2008). The Effect of Dynamics on Identifying Basic Emotions from Synthetic and Natural Faces. *International Journal of Human-Computer Studies*, 66(4), 233–242.
- Khanna, P. & Sasikumar, M. (2010). Recognising Emotions from Keyboard Stroke Pattern. *International Journal of Computer Applications*, 11(9).
- Koelstra, S., Yazdani, A., Soleymani, M., Muhl, C., Lee, J., Nijholt, A., Pun, T., Ebrahimi, T. & Patras, I. (2010). Single Trial Classification of EEG and Peripheral Physiological Signals for Recognition of Emotions Induced by Music Videos. Springer-Verlag Berlin, Heidelberg: *Brain Informatics*, 89–100. Springer.
- Lee, H., Choi, Y. S., Lee, S. & Park, I. P. Towards Unobtrusive Emotion Recognition for Affective Social Communication. *Proc. 2012 IEEE Consumer Communications and Networking Conference (CCNC '12)*. DOI: 10.1109/CCNC.2012.6181098
- Lottridge, D., Chignell, M., & Jovicic, A. (2011). Affective Interaction: Understanding, Evaluating, and Designing for Human Emotion. *Human Factors and Ergonomics*, 7(1), 197–217.
- McNeill, D. *Hand and Mind: What Gestures Reveal about Thought*. University of Chicago Press, 1992
- Mandryk, R.L., Atkins, M.S. & Inkpen, K.M. (2006). A Continuous and Objective Evaluation of Emotional Experience with Interactive Play Environments. In *Proc. CHI2006*. ACM Press, 1027–1036.
- Niewiadomski, R. L., & Pelachaud, C. (2010). Affect Expression in ECAs: Application to Politeness Displays. *International Journal of Human-Computer Studies*, 68(11), 851–871.
- Oliveira, E., Benovoy, M., Ribiero, N., Chambe, T. (2011). Towards Emotional Interaction: Using Movies to Automatically Learn Users' Emotional States. In *Proc. INTERACT 2011*, Part I, LNCS6946, 152–161.
- Paiva, A., Costa, M., Chaves, R., Piedade, M., Mourao, D., Sobral, D., et al. (2003). Sentoy: an Affective Sympathetic Interface. *International Journal of Human-Computer Studies*, 59(1), 227–235.
- Peter, C. & Beale, R. (eds) (2008). *Affect and Emotion in Human Computer Interaction: From Theory to Application*. LNCS 4868, Springer.
- Piccard, R. W. (1997). *Affective Computing*. Cambridge: MIT Press.
- Pollak, J.P., Adams, P., Gay, G. (2011). PAM: A Photographic Affect Meter for Frequent, In Situ Measurement of Affect. In *Proc. CHI2011*, ACM Press, 725–734.
- Russell, J. A. (1980). A Circumplex Model of Affect. *Journal of Personality and Social Psychology*, 39(6), 1161–1178.
- Soleymani, M., Pantic, M., & Pun, T. (2012). Multimodal Emotion Recognition in Response to Videos. *IEEE Transactions on Affective Computing*, 3(2), 211–223.
- Szafir, D. & Mutlu, B. (2012). Pay Attention! Designing Adaptive Agents that Monitor and Improve User Engagement. In *Proc. CHI2012*, ACM Press, 11–20.
- Takahashi, K. (2004). Remarks on Emotion Recognition from Bio-Potential Signals. In *Proc. Second Int'l Conf. Autonomous Robots and Agents*, 186–191.
- Vloed, G. & Berentsen, J. (2009). Measuring Emotional Wellbeing with a Non-intrusive Bed Sensor. In *Proc. INTERACT 2009*, Part II, LNCS5727, 908–911.
- Wang, R., Quek, F., Tatar, D., Teh, J.K.S., Cheok, A.D. (2012). Keep in Touch: Channel, Expectation and Experience. In *Proc. CHI2012*, ACM Press, 139–148.
- Xu, M., Xu, C., He, X., Jin, J. S., Luo, S. & Rui, Y. (2012). Hierarchical Affective Content Analysis in Arousal and Valence Dimensions. *Signal Processing*, 93(8), 2140–2150.
- The Journal of Human Factors and Ergonomics Society. <http://www.hfes.org/web/pubpages/TopicList-HumanFactorsJournal.pdf>. Accessed 11 February 2015.