Automatic detection of emotions in depressed patients and healthy subjects: a mobile application study

Authors: Min-Shan Li, Yen-Lin Chen, Hsiao-Chi Chu, Yen-Ying Liao, Jung-Hsien Chiang, Galen Chin-Lun Hung

Author Affiliations:

1Department of psychiatry, Taipei City Hospital Zhongxiao Branch, No.87, Tongde Rd., Nangang Dist., Taipei City 115, Taiwan

2Department of Computer Science and Information Engineering, National Cheng Kung University, No.1, Daxue Rd. East Dist., Tainan City 701 Taiwan

3Department of General Psychiatry, Taipei City Psychiatric Center, Taipei City Hospital, No. 309 Songde Road, Xinyi Dist., Taipei City 110, Taiwan

4Institute of Medical Informatics and Department of Computer Science and Information Engineering, National Cheng Kung University, No.1, Daxue Rd. East Dist., Tainan City 701 Taiwan

5Department of Public Health, School of Medicine, National Yang-Ming University,
Running title: Automatic detection of emotions in depressed patients

Sources of financial: This work was supported by the Ministry of Science and Technologies, R.O.C. (MOST 105-2314-B-532-001) and Department of Health, Taipei City Government (104-52, 53, 54). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Previous presentation: not applicable

Acknowledgment: not applicable

Disclosure Statement: The authors have declared that there are no conflicts of interest in relation to the subject of this study.

Corresponding author: Galen Chin-Lun Hung, MD, MPH

Department of General Psychiatry, Taipei City Psychiatric Center, Taipei City Hospital,

No. 309 Songde Road, Xinyi Dst., Taipei, 110, Taiwan

EMAIL: galenhung@tpech.gov.tw

TEL: +886-2-27263141-1270

FAX: +886-2-2728-5059
ABSTRACT

Background

Improving the recognition and management of momentary emotion is crucial for mental health, both in depressed patients and in normal population.

Objective

The objective of this study is to examine the performance of a mobile phone system regarding the detection of real-time emotion.

Methods

A mobile phone application, iHOPE, was developed for detecting emotion. A Bayesian network with 15 nodes was used for inferring momentary emotion based on contextual factors and smartphone use patterns. Five patients with major depressive disorder and seven normal participants were recruited. Participants used the Circumplex model to label their daily emotions for 8 consecutive weeks, which were used for model training and validation.

Results

Depressed patients spent 77% more time with smartphone than healthy subjects. In comparison with accuracy of 25% by random guessing, our detection algorithm achieved an accuracy of 54% in all participants, as demonstrated by 10-fold cross validation. Predictive accuracy was better in patients than in healthy subjects. In
depressed patients, using data 180 minutes prior to emotion tags achieved the best performance, whereas for healthy subjects, the optimal time window was 30 minutes.

We also find that using more recent data (i.e. the past 2 weeks vs. the first 2 weeks or all data) resulted in better performance. The contribution of individual variable on predictive accuracy demonstrated significant inter-subject variability

**Conclusions**

The findings here suggest that, both in depressed patients and in healthy subjects, it is possible to infer momentary emotion with individualized detection algorithm, while identifying personal attributes to improve emotional awareness and design intervention.

**Keywords:** depressive disorder, mobile health, emotion detection
Introduction

Depression is a worldwide challenge for mental health, with an estimated 350 million people affected. It is a major contributor to the global burden of disease, as measure by disability-adjusted life years[1]. Presently, identification of depression relies mainly on subjective reporting, which depends on an individual’s ability to aware and correctly recall her emotional state. Reporting of depressive symptoms mostly occurs in the clinical context, which threatens the ecological validity[2]. Moreover, emotional state is fluid in nature; detection of its change is often postponed or prone to errors, which hinders a time-sensitive intervention. More importantly, in healthy population, undetected negative affect may escalate to clinical depression; prompt recognition and management of momentary emotion thus serves as a mean of primary prevention.

One way to unobtrusively monitor momentary emotion is to passively tracking its contextual correlates and behavioral markers, which reduces the burden for data entry, and is potentially more objective[3, 4]. In recent years, great efforts have been made to harness technology for improving mental health. One major strategy is to leverage the popularity and functionality of mobile phones[5]. The number of mobile phone users is expected to surpass 4.7 billion in 2017[6]. It is a ubiquitous, non-intrusive and
highly personal device, often equipped with sensors, which are capable to collect contextual and behavioral data both continuously and passively. Furthermore, mobile phone ownership in patients with affective disorders can be as high as 86%[5]. Although in its infancy, a growing body of research has explored the potential use of mobile technologies to automatically infer emotions by analyzing environmental information (e.g. weather, location) and behavioral characteristics, such as sleep, conversation and smartphone interaction (e.g. making calls, text-messaging, opening applications).

For example, in patients with depression and bipolar disorder, geospatial variability may be associated with depressive state, so as physical activity and phone call features[3, 7]. Comparatively more studies were conducted in healthy population and found that location variability, environmental features (e.g. ambient light and sound), physical activity, sleep duration and phone use patterns are correlated with depressive symptoms[8-12]. Emotion-aware mobile applications have been developed using complex machine learning methods [11, 13-15]. Currently it is unclear which type of information is universally crucial for detecting emotion; alternatively, there is evidence that the detection algorithm must be personalized[9, 11]. Moreover, certain machine-learning methods function like a 'black box', such that contributors to an emotional detection can't be identified. Lastly, prior studies mostly recruited only
patients or healthy subjects, therefore, it is unclear if the detection performance of a
given system in patients in transferrable to healthy individuals, and vise versa.

Accordingly, in this exploratory study, we examined a mobile application, iHOPE, which detected momentary emotion based on analyzing contextual information and smartphone use patterns. Bayesian network was chosen to construct a personalized detection algorithm, with an intention to identify specific attributes. Detection accuracy was compared in depressed patients and healthy subjects, and between models using data from different time windows.

METHODS

System architecture

Figure 1 shows the system architecture of iHOPE. We used participants’ mobile phone to collect 3 types of data (emotion labels, contextual information, smartphone use patterns) and uploaded them to a cloud database. Data was then processed and discretized for constructing the nodes of Bayesian network. Meanwhile, data from different time slots (30, 60, 120 and 180 minutes) were compared regarding its detection accuracy. This personal model was trained on the cloud with continuously updated data, and then fed back to the mobile phone. By adapting the model, the mobile phone system was able to locally infer users’ emotion without a connection with the cloud.
Emotion labeling

We adapted the circumplex model of emotion developed by James Russell[16] to label momentary emotion in 2 dimensions: valence (i.e. positive or negative emotion) and arousal (i.e. intensity of emotional activation). Visual analogues scale was used to indicate the degree of valence and arousal for a reported emotional state.

Adopting the signal contingent method of experience sampling[17], our system notified the users to label their emotions at least 3 times a day, with a random interval of at least 2 hours.

Contextual information

The following contextual information were collected:

a) Working and non-working days: participants reported, of 7 days in a week, which days they worked and which they didn’t
b) Time of a day, as indicated by the clock function of the mobile phone

c) Quality of sleep: participants reported, in a 5-point Likert scale, their

   quality of sleep, ranging from very good to very bad.

d) Weather: data regarding current weather was collected from two

   websites ‘Yahoo! Weather’ and ‘forecast.io’.

Smartphone use patterns

   We included application usage and call states as patterns of smartphone use.

According to the functionality, applications were categorized to “Economic


   “Health”, and “Shopping” (see Appendix A for details). “Total Time” and “Total

   Frequency” were used to illustrate the cumulative usage time and frequencies of all

   foreground applications during a given time window.

   We defined three call states of users’ smartphone. “Received Call” indicated

   that the user picked up an incoming call; “Missed Call” referred to an unanswered

   call; “Dial Out” indicated that the user made calls. The total amount of time elapsed

   and frequencies of all three states were recorded.

Bayesian Network

   Because inferring emotional states involves uncertainty, we choose Bayesian

   network to construct the detection algorithm, for its capability to predict with

   uncertain clues and incomplete information. Moreover, Bayesian Network provides
full representations of probability distributions over their variables[18]. Therefore, they can be conditioned upon any subset of the variables and that support any direction of reasoning, such as bottom-up direction reasoning, top-down direction reasoning, and inter-causal reasoning that addresses the mutual causes of a common effect, which provides the interpretability of our system. We can examine the assumptions that our Bayesian Network makes as well as the decision process the network uses.

In our system, we used Netica[19] to construct a Bayesian network for inferring momentary emotion. Figure 2 depicted the Bayesian network of one of the participants, with 15 nodes. Name of each node was shown, along with its discretized values and data distribution. For example, “Time” had 3 values: morning, afternoon and night, and 20.6% of this participant’s data were recorded as “morning”.

Figure 2: Nodes and data structure of the Bayesian network
Because Bayesian network is primarily oriented towards handling discrete variables, we discretized all continuous variables to two discrete values, high (“H”) and low (“L”). Therefore, an emotional state was categorized according to the valence and arousal values as one of the 4 combinations: “H (valence) H (arousal)”, “LH”, “LL”, and “HL”.

For other continuous variables in our Bayesian network, values were discretized into two intervals according to their own cut point. The cut points was determined by two points, the mean of the biggest $n$ values and the mean of the smallest $n$ values from the ordered feature values. The cut point was defined as the average of these two points. Note that these $2n$ values were individually chosen based on the distribution of each participant’s data. Accordingly, for the variables “Sleep quality”, “Total time”,...
and “Total frequency”, they were discretized into “H” and “L”. For each application category, we combined discretized usage time and frequency to four attributes: “HH”, “LH”, “LL”, and “HL”.

Experimental and Evaluation methods

Outpatients with depressive disorder were recruited from Taipei City Hospital, Songde Branch, with healthy subjects recruiting from National Cheng-Kung University. This study conformed with the Declaration of Helsinki and received proper Institutional Review Board approval. Written informed consent was obtained from each participant prior to study initiation. Demographic and clinical characteristics and perceived attributes of momentary emotion were collected at baseline. We then helped them to download iHOPE to their Android mobile phones and requested them to use it continuously for 8 weeks. Initially, data collected in the first 2 weeks was used to train the Bayesian network, which predicted participants’ emotion in the following week. Subsequently, data collected from different time frames (i.e. the first 2 weeks, the recent 2 weeks and all available data) was used to infer momentary emotion in a particular week, and the predictive accuracy was compared. To evaluate the performance of our approach, we applied ten-fold cross validation in which the original sample is randomly divided into approximately 10 equal-sized subsamples[20].
RESULTS

<table>
<thead>
<tr>
<th>ID</th>
<th>Age</th>
<th>Gender</th>
<th>Occupation</th>
<th>Psychotropic use</th>
<th>Use of other emotion-related apps</th>
<th>CGI-S</th>
<th>BDI</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>51</td>
<td>M</td>
<td>Unemployed</td>
<td>Y</td>
<td>N</td>
<td>5</td>
<td>14</td>
</tr>
<tr>
<td>P2</td>
<td>23</td>
<td>F</td>
<td>College student</td>
<td>Y</td>
<td>N</td>
<td>5</td>
<td>37</td>
</tr>
<tr>
<td>P3</td>
<td>24</td>
<td>F</td>
<td>College student</td>
<td>N</td>
<td>N</td>
<td>4</td>
<td>26</td>
</tr>
<tr>
<td>P4</td>
<td>31</td>
<td>F</td>
<td>Unemployed</td>
<td>N</td>
<td>N</td>
<td>3</td>
<td>27</td>
</tr>
<tr>
<td>P5</td>
<td>22</td>
<td>M</td>
<td>College student</td>
<td>N</td>
<td>N</td>
<td>3</td>
<td>8</td>
</tr>
<tr>
<td>Q1</td>
<td>25</td>
<td>F</td>
<td>College student</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Q2</td>
<td>24</td>
<td>M</td>
<td>College student</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Q3</td>
<td>27</td>
<td>F</td>
<td>Assistant</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Q4</td>
<td>27</td>
<td>M</td>
<td>IT specialist</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Q5</td>
<td>24</td>
<td>M</td>
<td>College student</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Q6</td>
<td>23</td>
<td>M</td>
<td>College student</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>Q7</td>
<td>24</td>
<td>F</td>
<td>College student</td>
<td>N</td>
<td>N</td>
<td>N/A</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Participant characteristics

In this pilot study, we recruited 5 patients with depressive disorder and 7 healthy subjects. Their demographic and clinical characteristics were depicted in Table 1. Overall, participants were younger, with an equal distribution of males and females. Only 2 patients took psychotropic, and none of them used other emotion-related mobile applications during the study period. Our patients endorsed substantial depressive symptoms, as indicated by scores of Beck Depression Inventory, with clinically moderate to severe depression, according to scores of clinical global impression.

Table 1: Demographic and clinical characteristics of study participants
In our participants, we asked beforehand if the variables examined in the Bayesian network would affect momentary emotions. Their subjective reporting was shown in Table 2. Except a universal endorsement of sleep as an attribute of emotion, marked individual variability was evident that each participant had a unique set of perceived contributors. Note that in participants P3, P4, P5 and Q7, many attributes were left undetermined.
Table 2. Subjective reporting of potential attributes of emotion in study participants

<table>
<thead>
<tr>
<th>Questions</th>
<th>1. Does the weather affect your emotions?</th>
<th>2. Does talking on the phone frequently or for a long time affect your emotions?</th>
<th>3. Does sleep quality affect your emotions?</th>
<th>4. Does using some mobile applications affect your emotions?</th>
<th>5. Does being at work affect your emotions?</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>P2</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
</tr>
<tr>
<td>P3</td>
<td>U</td>
<td>U</td>
<td>Y</td>
<td>U</td>
<td>Y</td>
</tr>
<tr>
<td>P4</td>
<td>U</td>
<td>U</td>
<td>Y</td>
<td>U</td>
<td>N</td>
</tr>
<tr>
<td>P5</td>
<td>U</td>
<td>U</td>
<td>Y</td>
<td>U</td>
<td>Y</td>
</tr>
<tr>
<td>Q1</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>N</td>
</tr>
<tr>
<td>Q2</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Q3</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Q4</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Q5</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>N</td>
<td>N</td>
</tr>
<tr>
<td>Q6</td>
<td>Y</td>
<td>N</td>
<td>Y</td>
<td>Y</td>
<td>Y</td>
</tr>
<tr>
<td>Q7</td>
<td>U</td>
<td>U</td>
<td>Y</td>
<td>U</td>
<td>U</td>
</tr>
</tbody>
</table>

Y: yes, N: no, U: undetermined

Data Distribution of Each Variable in our Bayesian networks

Distribution of all participants’ data for each variable was depicted in Figure 3.

It appeared that variable "Call" and application category variables such as "Economic Information" and "Shopping" had a small proportion of collected data. The variable "Health" had an imbalanced data distribution such that "LL" was exceedingly frequent. The variables "Total Time" and "Total Frequency" had large amount of value "L"; the variable "Weather" contained extremely few "Rain".

Figure 3: Application data distribution.
Predictive accuracy in different time windows

We tested, in depressed patients and healthy subjects, time windows of 30, 60, 120, and 180 minutes regarding the predictive accuracy of the emotional state. For all participants, a time window of 30 minutes had the best accuracy of 54%. In general, prediction accuracy of depressed patients was better than the healthy users in each time window. In depressed patients, the 180-minute time window achieved the best accuracy; on the contrary, in healthy users, a graded relationship was present that longer time window had lower predictive accuracy. Generally, the Bayesian network performed significantly better than the benchmark random guessing method. We further analyzed the proportion of time using smartphone in each time window, overall, depressed patients spent significantly more time (19.3-20.8%) than healthy users (11.2-11.8%), which contributed to more training data acquisition.

Figure 4. Predictive accuracy of the emotional state in different time windows in depressed
Choice of training data selection

We compared the average performance when we used all available data, the data of the first two weeks and the data of the recent two weeks as the training data. Figure 5 showed the results. Generally, the accuracy of the model trained with the prior 2-week data was better than that with the first 2-week data or with all available data. It appeared that models trained with temporally proximal data had better predictive accuracy.

Figure 5: The average performance comparison between two different training data.
Variable analysis

Figure 6a and Figure 6b depicted the prediction performance when each time we removed a specific variable from the Bayesian network. For example, in Figure 6a, removing “sleep” significantly improved the predictive accuracy for patients P2 and P3, but not for other patients. In Figure 6b, the “total time” and “weather” were crucial predictors for subject Q4, such that removing them led to a decrease of model performance. “Work day” was important for subject Q2, but was counter-productive for subject Q3.

We then compared the results with the perceived attributes shown in Table 2 and found mixed patterns of correlation. For example, in subject Q4, “total time” of application use and “weather” were important in model building, and he did acknowledge that weather and application use affected his emotion. On the contrary,
subject P2 thought that sleep affected her emotion, but it was counterproductive to include sleep when modeling her emotion in the Bayesian network.

In summary, the influence of a given variable on predictive performance varies considerably among subjects; this is consistent with the well-known phenomenon of inter-subject variability regarding emotion detection[21]. Meanwhile, the correlation of model-identified and perceived attributes was mixed.

Figure 6: The prediction performance with removal of one specific variable from the Bayesian network in 30-min time slot for depressed patients and normal participants.

(a) Depressed patients

(b) Healthy participants
DISCUSSION

In this exploratory study, we developed a smartphone system capable of detecting momentary emotion based on contextual information and smartphone use patterns, both in depressed patients and in healthy subjects. Detection accuracy is superior in patients than in healthy ones, and their optimal time window for data acquisition is different. Our system needs to be retrained with more recent data to achieve a better performance. Prominent inter-subject variability is noted, such that each individual has her unique set of emotional attributes.

Limitations

The generalizability of our findings is limited by the small sample size and relatively short duration of experiment. Our system works only on Android phones, which lead to selection bias. Absence or imbalanced distribution of data for certain
predictors may result in biased modeling. Some information potentially associated with emotion, such as physical activities, semantic contents in text messages or emails and pitches of speech[22], was not included in our system.

Comparison with previous works

The strength of our study is that our system is able to infer momentary emotion, but not emotional status reported prior to study initiation[8, 9], which is more valuable for detecting states and changes of emotion to provide prompt intervention. Saeb et al. used correlation analysis and fail to identify consistent associations of GPS sensor data and momentary affect[23]. Burns and colleagues applied machine-learning methods to infer emotion state based on 36 phone sensor values, predictive accuracy was low[24]. On the contrary, in bipolar patients, Grunerbl et al. was able recognize their mood state and state changes with a smartphone-based detection system[7]. LiKamwa and associates analyzed phone usage to infer momentary emotion, predictive accuracy reached 61% and was improved by using personalized data to train the classifier[25]. To our knowledge, this is the first study examining an emotion-aware system both in depressed patients and healthy subjects. We found that depressed patients spent significantly more time with smartphones than healthy individuals, which may influence the performance of detection algorithm, as well as choice of time frames for training data. Our predictive
accuracy is moderate, but the Bayesian network provides unique advantages: contribution of each variable regarding emotion detection is illustrated, which may be used to improve individual's awareness regarding the association of momentary emotion and contextual as well as behavioral factors. Moreover, our system can be refined periodically by removing unproductive variables and including clinically meaningful factors.

Implications and future directions

There is evidence that patients with depressive disorder had deficits of emotion processing even if they are recovered, which may contribute to recurrence of depression[26]. Accordingly, Kramer et al provided feedback to momentary positive affect reported by depressed patients, which led to an improvement of depressive symptoms[27]. An emotion-aware mobile system has the potential to bypass the requirement of subjective reporting and directly provide intervention. Research with larger sample size, longer experimental period and more diverse clinical as well as general population is needed to confirm our preliminary findings. In our report, system-identified contributors of momentary emotion may not be compatible with subjective reporting; possible explanations may include an unawareness or misperception of one's emotional determinants. Alternatively, the detection algorithm may need to be refined according to subjects' feedback and clinical judgment. Once a
consistent set of determinants are identified, specific behavioral interventions can be built accordingly.

The findings here suggest that, both in depressed patients and in healthy subjects, it is possible to infer momentary emotion with individualized detection algorithm, while identifying personal attributes to improve emotional awareness and design intervention.
References


10. Saeb S, Zhang M. Mobile Phone Sensor Correlates of Depressive Symptom Severity in Daily-Life Behavior: An Exploratory Study. 2015;17(7):e175. PMID: 26180009
11. Hung GC, Yang PC, Chang CC, Chiang JH, Chen YY. Predicting Negative Emotions Based on Mobile Phone Usage Patterns: An Exploratory Study. JMIR research protocols 2016 Aug 10;5(3):e160. PMID: 27511748


25. LiKamWa R LY, Lane ND, Zhong L. Can your smartphone infer your mood? . PhoneSense 2011: International Workshop on Sensing Applications on Mobile Phones; 2011 Nov. 1-4 of Conference; Seattle, WA.

26. Levens SM, Gotlib IH. Updating emotional content in recovered depressed individuals: Evaluating deficits in emotion processing following a depressive episode.

PMID: 25889375