Impact of Experience Sampling Methods on Tap Pattern based Emotion Recognition

Surjya Ghosh
IIT Kharagpur
WB-721302, India
surjya.ghosh@iitkgp.ac.in

Vatsalya Chauhan
IIT Kharagpur
WB-721302, India
vatsalyachauhan1@gmail.com

Niloy Ganguly
IIT Kharagpur
WB-721302, India
niloy@cse.iitkgp.ernet.in

Bivas Mitra
IIT Kharagpur
WB-721302, India
bivas@cse.iitkgp.ernet.in

Pradipta De
SUNY Korea
Incheon, Korea
pradipta.de@sunykorea.ac.kr

UbiComp/ISWC '15 Adjunct, September 7-11, 2015, Osaka, Japan

ACM 978-1-4503-3575-1/15/09$15.00
http://dx.doi.org/10.1145/2800835.2804396

Abstract
Smartphone based emotion recognition uses predictive modeling to recognize user's mental states. In predictive modeling, determining ground truth plays a crucial role in labeling and training the model. Experience Sampling Method (ESM) is widely used in behavioral science to gather user responses about mental states. Smartphones equipped with sensors provide new avenues to design Experience Sampling Methods. Sensors provide multiple contexts that can be used to trigger collection of user responses. However, subsampling of sensor data can bias the inference drawn from trigger based ESM. We investigate whether continuous sensor data simplify the design of ESM. We use the typing pattern of users on smartphone as the context that can trigger response collection. We compare the context based and time based ESM designs to determine the impact of ESM strategies on emotion modeling. The results indicate how different ESM designs compare against each other.

Author Keywords
Emotion Detection, User Interaction, Data Collection, Tap Sensing, Experience Sampling

ACM Classification Keywords
H.1.2 [User/Machine Systems]: Miscellaneous;J.4 [Social and Behavioral Sciences]

Introduction

Experience Sampling Methods (ESM), or Ecological Momentary Assessment (EMA), techniques are widely used in psychology and behavioral science to collect momentary affect of people in daily life [5]. Although the merits of field studies in designing ESM are accepted, it is often challenging to conduct field studies. The ubiquity of smartphones in everyday life provides a non-intrusive way to design ESM. Electronic ESM, unlike pen-and-paper or interview based methods, rely on predictive modeling to determine user's emotion states.

In order to generate the predictive models, an important step is to collect ground truth data and correlate it to the monitored parameters. This is the learning or training phase that can significantly impact the accuracy of the model. In emotion recognition, ground truth is the state of the mind as reported by the user. Typically user's perceived emotion is collected by explicitly asking the user to respond about her state of mind. In smartphone based studies a pop-up survey serves the purpose. But the challenge is to choose the appropriate moment to trigger the pop-up that truly captures the momentary affect.

Smartphone sensors provide a variety of contexts, such as location, activity, social environment that can be used to trigger a user survey. However, many of these sensors are not continuously sampled. Lathia et al. showed that subsampling of sensor data to determine survey trigger points can bias the ESM based behavioral inference [6]. They posited that the use of multiple sensors may mitigate the problem since that may collectively increase the sample density. Alternatively, there are works which used controlled environments to generate the ground truth. In EmotionSense, Emotional Prosody Speech and Transcripts library is used to label data for training and classification [10]. Lu et al. used targeted interviews as a means to label stress effects on users [9]. Even simpler alternative is to use time based trigger, i.e. ask periodically, for user response. However, a clear preference for a ESM design is not obvious.

In this paper, we pose the question that given a continuous sample of the sensor data, does an obvious ESM approach exist? Typing activity, as well as, mouse usage has been shown to be good indicators of emotion states of users [1], [11]. We use tap behavior, or smartphone keyboard usage pattern, as the feature to detect emotion. We design three ESM methods for ground truth collection, and analyze the impact on emotion model generation. We designed an Android based system that can track user's typing activity and use it as a cue to trigger survey responses about emotion state. The time triggered ESM approach uses different periodicity. We use the survey results to train and test the model, evaluate the effectiveness of the generated models, and identify the relationship between timeliness of ESM engagement and accuracy of user responses.

Related Work

Experience Sampling Method (ESM) can be typically scheduled in three formats - event-driven, signal driven, or time-driven. In this section, we focus on the user feedback collection techniques used in the existing emotion or mood recognition systems.

Some work avoid the need for user feedback collection for labeling by using annotated library of the feature under observation. For example, in EmotionSense, Rachuri et al. use the Emotional Prosody Speech and Transcripts library to train the emotion classifier [10]. Thus EmotionSense circumvents the issue of in-situ collection of user

response. Similarly, in their work on assessing stress levels, Lu et al. collect the emotion labels corresponding to speech patterns by designing personal interviews that can elicit different emotions [9].

In the MyExperience system, Froehlich et al. implement in-situ feedback collection, and uses participant's reaction to events as the context for feedback collection [3]. However, use of time-driven probing is common among studies that require in-situ user feedback collection to label monitored features. For example, in StudentLife, Wang et al. periodically triggers EMA questionnaires to the users on their smartphones [12]. Gao et al. analyzed player emotion based on touch during gameplay [4]. They train the model by asking users to input their emotion after finishing each game level. Lee et al. designed a Twitter client app that allows user to record emotion by typing some text whenever she wishes [7]. When there are multiple observed features, event based user feedback collection is harder to design. In MoodScope, which is designed to capture user's mood state based on multiple features, the user feedback to train the model is collected by asking the user periodically [8].

In a naturalistic behavioral scenario, where the focus is not on extracting a specific type of emotion, sampling experience at right point of time helps to improve data quality [5]. Although it raises the implementation cost and poses burden on respondents, an well-designed ESM may be able to replace time-based sampling techniques. In summary, there is a lack of research comparing alternative design choices to collect user response. Event based or signal based sampling techniques, that are well established in behavioral science, can be useful in designing accurate emotion models. This work focuses on evaluating the impact of different emotion sampling techniques on emotion recognition.

Methodology

In this section, we describe the Android application, called TapSense, used for data collection and the ESM designs used to collect user response.

TapSense System Design

TapSense collects typing patterns on a smartphone, as well as, provides a survey question interface. When TapSense is launched, we replace the default keyboard by a custom virtual keyboard that allows us to trap the tap events. We record only the timestamp of each tap event, and the foreground application name. The survey questionnaire asks the user to choose one of the emotion states, Happy, Sad, Excited, Normal, that reflects her current state of the mind. The questionnaire is triggered automatically based on the ESM design described later. The survey responses and tap event logs are sent to a server periodically. User responses are used to label the tap events for training a SVM based model. Figure 1 shows the architecture of TapSense.

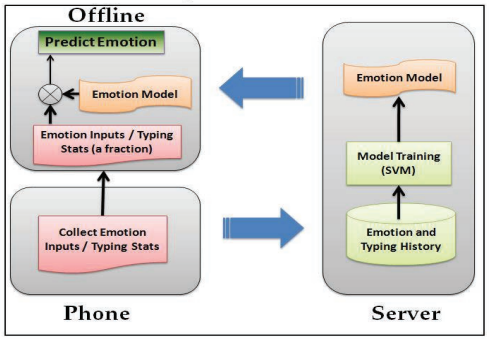


Figure 1: Tap based emotion recognition system (TapSense) architecture

The key feature used in the SVM modeler is the time interval between two tap events, called Inter-Tap Distance (ITD). We assume that different emotion states will affect the typing behavior of the user. For example, when excited or happy, a user may type faster, as compared to when sad or neutral. We also experiment with other features, like mean of all ITD values between two user surveys, and the application types.

ESM Design
We designed the ESM techniques to be either synchronous or asynchronous with the tap events. Synchronous design uses tap behavior or other contexts, while asynchronous design is oblivious of the tap behavior. In synchronous design, application switching is used as an event to trigger a user survey, while in signal based design an idle period during typing denotes a signal. Asynchronous design uses time based triggers that are periodic with different periodicity.

We describe below the ESM designs.

- **Signal-based Sampling(SB)**: A pause in typing is measured by an idle time threshold of 2 minutes. If a user has paused typing, then the user is asked for a response. This approach can be effective in capturing the momentary mental state, but leads to high user engagement.
- **Event-based Sampling(EB)**: When a user switches an application, we ask for a response. This approach reduces the user engagement. However, if an application, like Instant Messaging, is used over a long period where user went through variations in emotion, then this approach will fail to capture those fine grained variations in emotion.
- **Time-based Sampling(TB)**: We ask the user at regular intervals of time to record her mental state. In our experiments we have used two time intervals: 3 hours and 30 minutes, which are named as TB1 and TB2 respectively. We choose two different intervals to verify whether increasing emotion sample count without any attention to the observed event can lead to improvement in emotion modeling.

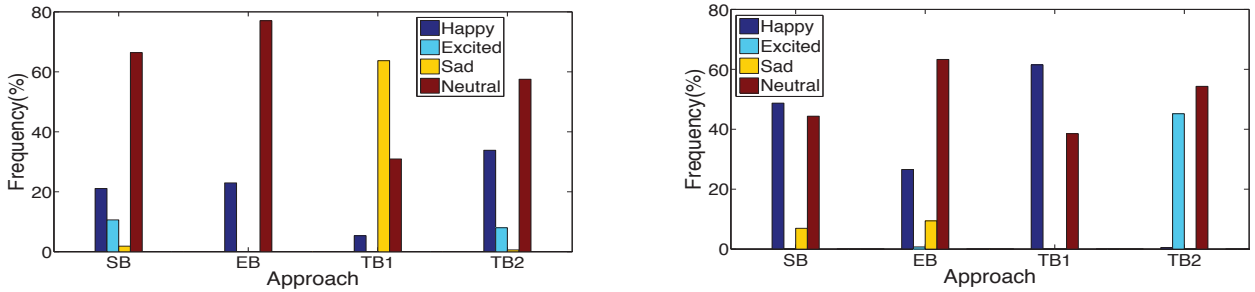
Experimental Setup and DataSets

In this section, we present the experimental setup and the data collected from user studies.

Participants
Our requirement for the target participants was that user regularly uses typing based applications. We chose 10 students, between the age of 19 and 24, and surveyed them about their usage of typing based applications, like WhatsApp, Facebook, SMS. We selected 2 participants who spend on average more than 120 minutes in Whatsapp, send on average 5 emails or SMSes and spend on average more than 60 minutes in other IM apps. We installed TapSense on a Sony-Xperia and a Samsung GT-I9082 Android based phones, and collected the tap data and user feedback for 16 days. We changed the ESM design for user response collection after 4 days.

Overview of Data
In this section, we present an overview of the collected data. For user-1 and user-2, we show the distribution of the emotion labels, and the average values of Inter-Tap Distance (ITD).

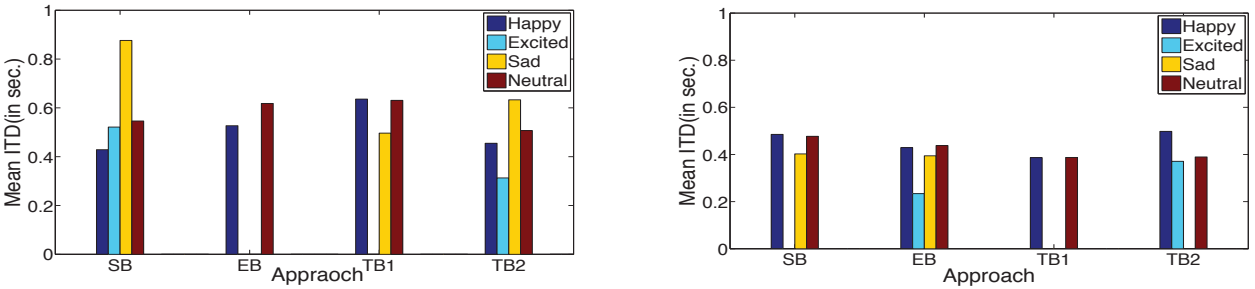
Figure 2a, and 2b show the emotion label distribution for user-1 and user-2 respectively. For both users, proportion of neutral emotion reported is high in all ESM designs as expected. But for other emotion states, there is an impact of the ESM approach chosen. When an ESM approach samples more, then more labels can be captured. Hence



(a) Distribution of emotion labels from user-1 responses

(b) Distribution of emotion labels from user-2 responses

Figure 2: Distribution of emotion labels based on the responses collected from user-1 and user-2 using different ESM techniques.



(a) Mean ITD of user-1 based on the typing pattern

(b) Mean ITD of user-2 based on the typing pattern

Figure 3: Mean ITD values of user-1 and user-2 for different ESM techniques. User-1 exhibits significant variation in typing speed across emotion states, while variation in typing speed for user-2 is much lower.

Signal-based triggering (SB) captures more states both for user-1 and user-2. However, there are instances where we did not receive any emotion label for some states (e.g. user-1 in Excited state for EB and TB1). It is possible that excited state is a transient state and can be captured when the user is probed at the moment itself.

Figure 3a and 3b show the average value of ITD for the two users across different emotion states over different ESM designs. Between two emotion labels, $e1$ and $e2$, all the tap events are tagged to belong to the label $e2$. We compute the mean ITD based on the ITD of tap events with same emotion label.

User-1 shows high variation in tapping pattern across emotion states compared to user-2. The standard deviation of mean ITD across emotion states for user-1 as 0.1954 sec, while for user-2 it is 0.0456 sec. This indicates that choice of the parameter to monitor can affect the emotion model significantly.

Results

In this section, we analyze the influence of ESM designs on emotion modeling. We designed a SVM based classifier that uses Inter-Tap Distance (ITD) as the single feature to build an emotion model. We also tested with additional features to understand the relationship between ESM techniques and increasing features on emotion modeling. We train the classifier with 80% of the samples and test with remaining 20% of the samples based on the dataset collected from the two users.

How does ESM techniques influence accuracy of emotion detection?

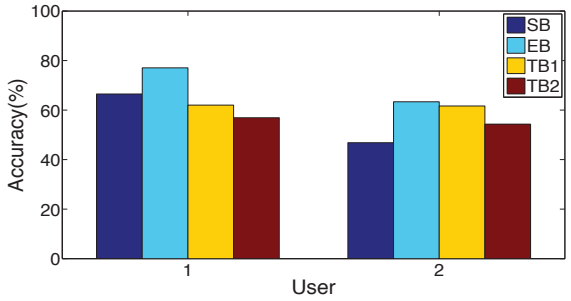


Figure 4: Emotion state classification accuracy of different classifiers using a single feature (ITD). The result is shown for data collected using different ESM designs.

We measured the accuracy of identifying an emotion based on data collected using each ESM approach described in ESM Design earlier. Figure 4 shows the prediction accuracy for each ESM technique. We observe that for both user-1 and user-2 Event-based Sampling (EB) works better than the other techniques. For user-2, the performance corresponding to each approach does not vary significantly. Overall, depending on the ESM techniques there are variations in accuracy for both the users.

Note that for user-1 the variance in ITD across different emotion states is higher than that of user-2. Since we are only using ITD as the classification feature, therefore, the accuracy is lower for user-2 irrespective of the ESM technique. Better performance of EB for user-1 can be attributed to lack of data, but even signal-based (SB) technique performs better than the Time based (TB) ones. This indicates that collecting user's emotion labels at the appropriate moment provides the true picture of the user's mental state if there is high variance in the selected feature. In time based approach, the mental state recorded by the user does not seem to correlate as closely as the other approaches.

For user-2, we had observed from Figure 3b that the variation in ITD is low. Hence, irrespective of event or time based approaches, the accuracy values are not significantly different. Infact, for user-2 TB performs better than SB because low variance in ITD makes it harder to classify with higher volume of sample points. Note that the number of sample points will be the lowest for TB1, then TB2 due to higher frequency of sampling, and the highest for SB. We can infer that if the selected feature does not have significant variation then time sampling may suffice as opposed to using event or signal based sampling.

How different are the ESM approaches?
We investigate the extent of difference across the ESM approaches. If we assume that the data collected from the users by the ESM approaches are similar, then training a model using data from one approach should perform well when tested on data collected using a different approach. We cross-validate the approaches for each user by training the classifier using one sampling approach (Train App in Table 1), and testing on the samples obtained from a different approach. In all the cases, we use 80% as training sample, 20% as test sample. The classifier accuracy is shown in Table 1.

Table 1: Classification Accuracy in Cross-validation

User#	Train App	SB	EB	TB1
1	SB	66.50	77.05	31.0
1	EB	66.50	77.05	31.0
1	TB1	4.31	7.57	62.0
2	SB	46.82	27.41	60.87
2	EB	44.40	63.35	38.33
2	TB1	48.69	26.57	61.66

For user-1, SB and EB behave identically. This indicates that sampling emotion on application switch (EB) and sampling when there is a pause in typing (SB) carry similar information. The observation can be explained by the smartphone usage behavior of the users. Figure 5 shows that user-1 frequently switches among applications with app usage duration peaking at 2 mins. Since the idle threshold interval is also 2 mins, therefore, EB and SB collect similar emotion information from the user. Similar usage pattern has also been reported by Ferreira et al. [2]. On the other hand, user-2 uses an application for longer durations, which reflects in lower classification accuracy in the cross-validation between EB and SB for user-2.

Cross-validation between TB1 and other approaches lead to poor accuracy. TB1 labels the samples at regular intervals, unlike SB and EB. Therefore, TB1 fails to gather momentary changes in emotion which is captured by SB and EB.

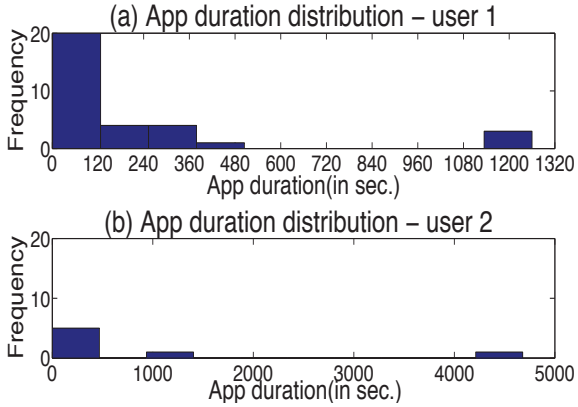
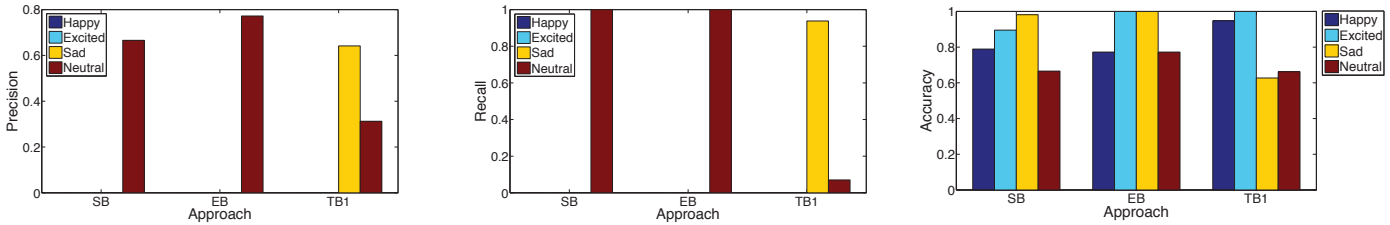


Figure 5: Application usage pattern of user-1 and user-2. User-1 frequently switches between applications showing short usage durations per application, while user-2 tends to use an application for a longer duration.

What is the role of sampling approaches on detecting individual emotion states?
Since the volume of data for each emotion state is different, we investigate how the ESM approaches performed in detecting each state individually. We calculate the precision, recall and accuracy metrics for each emotion state, as shown in Figure 6 and Figure 7.

We observe that for user-1 in SB and EB, the classifier could identify only neutral states, whereas in approach TB1 both sad and neutral states are identified. Incidentally, in Figure 2 we notice a high volume of the



(a) Precision **(b)** Recall **(c)** Accuracy

Figure 6: Different evaluation metrics for user-1: precision, recall, accuracy values for different emotion states.

corresponding emotion states, which helps to increase the precision. Similarly, we observe that the recall rate of neutral state is very high in SB and EB and it is low in TB1 which concurs with Figure 2. Similar results are observed for user-2.

We also observe that sad and excited states are detected more accurately than normal and happy states. This is due to fewer sample points for sad and excited states (refer Figure 2).

How does ESM approaches depend on additional features?

Application Category: We focus on the category of applications based on typing activity. We classify the typing-intensive applications into three categories; (Cat-1) Instant Messaging (IM), (Cat-2) Texting, (Cat-3) Others. IM applications, viz. Whatsapp, GTalk, Facebook chat, are category-I, while email, SMS are category-II, and the rest are category-III.

If there is a variation in typing patterns across application categories, use of application category as a feature can improve accuracy. But as shown in Figure 8, the results are similar to using only ITD. Since category-I applications (IM) constitutes more than 80% samples for both the users, therefore, use of application category as a feature does not significantly alter the accuracy values.

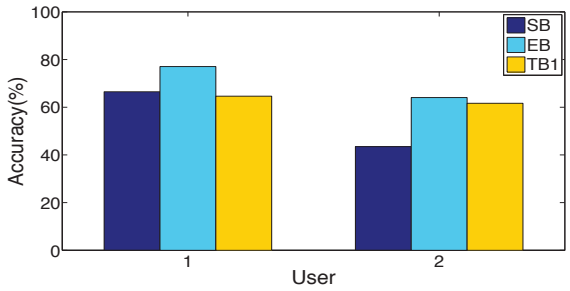
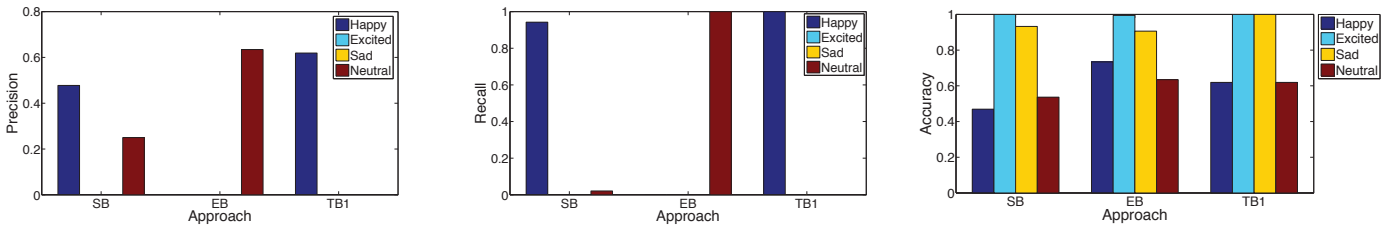


Figure 8: Overall emotion state classification accuracy of different classifiers built using multiple features (ITD, Application category). The results are shown for each ESM design.

Mean ITD: We include the mean ITD for each emotion state as a new feature in our model. This feature becomes useful to distinguish two emotional states where the corresponding ITDs exhibit significant overlap in values but mean ITDs are fairly distant. Accuracy results in Figure 10 highlight the fact that users whose typing



(a) Precision **(b)** Recall **(c)** Accuracy

Figure 7: Different evaluation metrics for user-2: precision, recall, accuracy values for different emotion states.

speed vary across emotional states (user-1) exhibit significant improvement across all the approaches. However users whose typing speed does not vary much across emotion states (user-2), we do not observe much variation in accuracy.

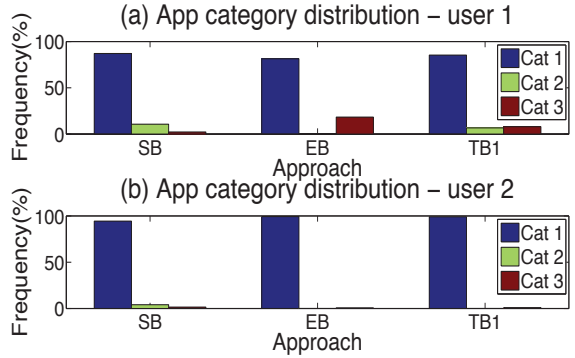


Figure 9: Distribution of different categories of typing applications used by user-1 and user-2 during data collection by different ESM designs. Both the users heavily use Cat-1, i.e. Instant Messaging, applications.

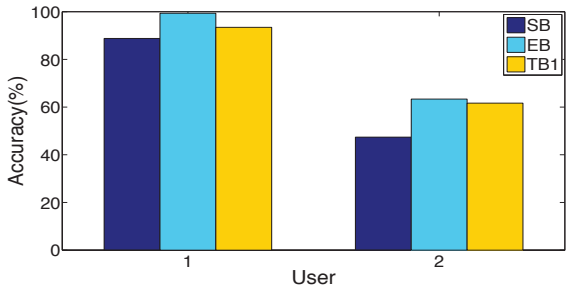


Figure 10: Emotion state classification accuracy of classifiers built using features, ITD, and Mean ITD. The results are shown for data collected using different ESM designs.

Conclusion

In this work, we explored the impact of different Experience Sampling Methods (ESM) in collecting user feedback. We focus on emotion recognition based on user's typing pattern on smartphones. We designed an Android application, TapSense, that can collect tap events. Based on different ESM designs, TapSense can notify user to fill out a questionnaire about her mental state. We showed that different triggers for user data

collection can lead to variations in the emotion model used for prediction. The results raise the question about which design is suitable for applications like emotion recognition: (a) a simple ESM design, like periodic user feedback collection, coupled with a number of features for generating the model? or, (b) an ESM design that is adapted to the monitored feature, which may reduce the complexity of feature selection to build the model. We plan to focus on these questions in our ongoing work and validate them with larger datasets.

Acknowledgement

This research was supported by ITRA under the project titled "Post-Disaster Situation Analysis and Resource Management Using Delay-Tolerant Peer-to-Peer Wireless Networks" (DiSARM, ITRA/15(58)/MOBILE/DISARM/01), and by MSIP(Ministry of Science, ICT and Future Planning), Korea, under the "IT Consilience Creative Program" (NIPA-2013-H0203-13-1001) supervised by the NIPA (National IT Industry Promotion Agency).

References

[1] Epp, C., Lippold, M., and Mandryk, R. Identifying emotional states using keystroke dynamics. In *Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems* (2011).

[2] Ferreira, D., Goncalves, J., Kostakos, V., Barkhuus, L., and Dey, A. K. Contextual experience sampling of mobile application micro-usage. In *Proceedings of the 16th ACM MobileHCI* (2014).

[3] Froehlich, J., Chen, M. Y., Consolvo, S., Harrison, B., and Landay, J. A. Myexperience: a system for in situ tracing and capturing of user feedback on mobile phones. In *Proceedings of the 5th Mobisys*, ACM (2007).

[4] Gao, Y., Bianchi-Berthouze, N., and Meng, H. What does touch tell us about emotions in touchscreen-based gameplay? *ACM Trans. on Computer Human Interactions 19*, 4 (Dec. 2012).

[5] Hektner, J. M., Schmidt, J. A., and Csikszentmihalyi, M. *Experience sampling method: Measuring the quality of everyday life*. Sage, 2007.

[6] Lathia, N., Rachuri, K. K., Mascolo, C., and Rentfrow, P. J. Contextual dissonance: Design bias in sensor-based experience sampling methods. In *Proceedings of ACM international joint conference on Pervasive and ubiquitous computing* (2013).

[7] Lee, H., Choi, Y. S., Lee, S., and Park, I. Towards unobtrusive emotion recognition for affective social communication. In *Consumer Communications and Networking Conference (CCNC)* (2012).

[8] LiKamWa, R., Liu, Y., Lane, N. D., and Zhong, L. Moodscope: Building a mood sensor from smartphone usage patterns. In *Proceeding of the 11th ACM Mobisys*, ACM (2013).

[9] Lu, H., Frauendorfer, D., Rabbi, M., Mast, M. S., Chittaranjan, G. T., Campbell, A. T., Gatica-Perez, D., and Choudhury, T. Stresssense: Detecting stress in unconstrained acoustic environments using smartphones. In *Proceedings of ACM Conference on Ubiquitous Computing* (2012).

[10] Rachuri, K. K., Musolesi, M., Mascolo, C., Rentfrow, P. J., Longworth, C., and Aucinas, A. Emotionsense: A mobile phones based adaptive platform for experimental social psychology research. In *Proceedings of ACM International Conference on Ubiquitous Computing* (2010).

[11] Sun, D., Paredes, P., and Canny, J. Moustress: detecting stress from mouse motion. In *Proceedings of ACM SIGCHI Conference on Human Factors in Computing Systems* (2014).

[12] Wang, R., Chen, F., Chen, Z., Li, T., Harari, G., Tignor, S., Zhou, X., Ben-Zeev, D., and Campbell, A. T. Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (2014).