Towards Designing an Intelligent Experience Sampling Method for Emotion Detection

Surjya Ghosh*, Niloy Ganguly*, Bivas Mitra*, Pradipta De†

*Department of Computer Science and Engineering, Indian Institute of Technology Kharagpur, INDIA 721302
†Department of Computer Sciences, Georgia Southern University, USA
Email: surjya.ghosh@iitkgp.ac.in, {bivas,niloy}@cse.iitkgp.ernet.in, pde@georgiasouthern.edu

Abstract—Experience Sampling Method (ESM) is widely used in idiographic approaches to collect within-person patterns. Planning a suitable survey schedule while designing an ESM based experiment is challenging as it must balance between survey fatigue of users, and the timeliness and accuracy of the responses provided by users. Even with the proliferation of ESM experiments, survey scheduling typically remains confined to use of fixed schedules, where periodic probes are sent to user, or event-based schedules, where depending on the number of events, large number of probes may interrupt user frequently. We propose a novel survey scheduling scheme, *Low-Interference High-Fidelity (LIHF)* ESM schedule, which is designed to reduce interference while retaining fidelity of user response. We integrated *LIHF* into an ESM application, called *TapSense*, that is used to infer user's emotion from typing characteristics on smartphone keypad. Conducting a 2-week field study involving 9 users, using proposed metrics we show that using *LIHF* there is 26% reduction in survey fatigue, 50% improvement in triggering survey probes in timely manner, and 8% improvement in predicting emotion states based on typing patterns compared to typical ESM scheduling techniques.

I. INTRODUCTION

Experience Sampling Methods (ESM) have emerged as a powerful tool for modern idiographic approaches in psychology and behavioral research [1]. By carefully probing a user at opportune time to elicit a survey response, idiographic methods can identify patterns of personalized behavior in diverse situations and experiences. Early designs of ESM used a pager to randomly signal a user for response [2]. With the proliferation of ubiquitous mobile devices, like smartphones, smart watches and other wearable computing devices, more intelligent and less intrusive survey scheduling can be designed. In literature attempts have been made to compare such ESM policies [3]. Several open source software platforms, like ESP [4], MyExperience [5], PsychLog [6], Personal Analytics Companion [7], are available on different mobile computing platforms to cater to ESM experiments.

One of the important steps in designing an ESM experiment is to plan the survey schedule. The simplest approach is to use a fixed schedule where the user is asked periodically to fill out a survey. This is referred to as the *Time-based* ESM [8]. Alternatively, context information can be used to trigger a survey, called an *Event-based* ESM [8]. *Event-based* ESM is easily supported by sensing user's present context transparently by monitoring the on-device sensors. Although *Event-based* ESM can trigger timely survey request, with a

Name of Work	ESM Schedule
StudentLife [12]	Periodic
Gao et al. [13]	Periodic
Kim et al. [14]	Periodic
MoodScope [10]	Periodic
Pielot et al. [11]	Periodic
MyExperience [5]	Context-driven
EmotionSense [15]	Context-driven

TABLE I: The table shows different ESM survey scheduling strategy used by works on emotion or mood recognition.

potential to elicit high fidelity response, it suffers from the fact that the number of events, and therefore, the number of survey requests can be very high. Large number of survey requests can easily deter users from responding honestly due to survey fatigue. Hence an approach that can balance the survey fatigue by limiting survey requests, while being timely to ensure fidelity of response, is ideal in an ESM experiment.

In this paper, we propose a novel survey scheduling strategy for ESM experiments, called *Low-Interference-High-Fidelity (LIHF)* schedule, that carefully reduces number of survey probes while retaining the accuracy of the responses. Our case study focuses on determining a person's daily emotion states, based on the circumplex emotion model [9]. Recent literature shows that smartphone activities may provide a strong signature of the emotional state of the user [10], [11]. Along this line, assuming smartphone typing behavior as the effect of four emotion states - Happy, Sad, Relaxed and Stressed, we conduct an ESM experiment using *LIHF* where user is asked to record one of the four states after some typing activity. Although we treat the typing activity as an event that can trigger a survey response, we reduce the survey count by spacing out the probes if an application is used very frequently.

Several works have looked at emotion or mood determination from different cues [10], [11]. Each of them rely on an ESM module with a specific scheduling strategy, as shown in Table I. *Time-based* or *Event-based* are the most common survey scheduling used. We compare the *LIHF* strategy with these common approaches in terms of survey fatigue and timely label collection with defined metrics. Our experiments are conducted in-situ as an user goes about her daily activities. We used the *LIHF* approach to gather the user responses once, and then use the data to generate response traces as if it were collected using *Time-based* or *Event-based* ESM approach. The accuracy of emotion prediction is considered as the performance metric indicating the effectiveness of the ESM

strategies. In a two weeks data collection experiment involving 9 users, *LIHF* shows better accuracy against the *Time-based* and *Event-based* ESM, in determining emotion states for each user.

The rest of the paper is organized as follows. In (§II), we perform a preliminary study to understand the limitations of conventional ESM approach and considering the insights from this study we propose the novel ESM in (§III). We discuss in detail the methodology of implementing the proposed ESM for typing based emotion detection in (§IV), which is followed by analysis and generation of the dataset used in the experiments (§V). Then we propose the metrics and compare the proposed ESM with conventional ESM approaches in (§VI). We conclude in (§VII) along with pointers to future directions.

II. MOTIVATIONAL STUDY

The objective of the preliminary experiment is to study the characteristics of *Event-based* ESM in collecting user responses. Since we focus on emotion detection based on smartphone typing activities, we collect the typing behavior of the user. We consider switching from a typing based application as the event to schedule a ESM survey questionnaire, which provides the ground truth about the emotion experienced by the user during the preceding *typing session*.
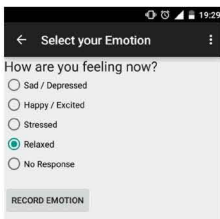

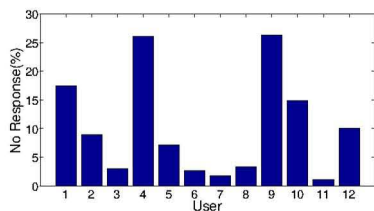Fig. 1: Survey popup to collect emotion labels from user

Fig. 2: Distribution of *No Response* in *Event-based* ESM

A. Apparatus

We implement an Android virtual keyboard to trace the typing events. The survey questionnaire requests the user to input an emotion state, as shown in Figure 1. The user may also choose to ignore the request by selecting *No Response*. We identify individual *typing session* as a sequence of typing activity within an application without leaving the app. Once the user switches to a different app, it triggers an event to collect the survey response, shown to the user as a pop-up.

B. Participants

We installed the ESM application on smartphones of 12 university students (aged between 18-24, 10 male, 2 female) and recorded their emotion labels and typing patterns for 2 weeks.

C. Insight from Study

A high proportion of *No Response* labels were recorded by users, as shown in Figure 2. Post experiment survey revealed

that for 11 out of 12 users the *Event-based* scheduling triggered large number of survey requests, demanding high user engagement and leading to annoyance. This insight confirmed that user engagement in a naive *Event-based* ESM design can drive away participants, and may even distort user responses due to survey fatigue.

III. LOW-INTERFERENCE-HIGH-FIDELITY ESM

In this section, we present the design of an ESM schedule that can reduce the survey fatigue by limiting the number of survey probes, while retaining the benefit of collecting emotion labels from users as close to the event as possible. Although each *typing session* may evoke different emotions, we assume that unless the user stays in an app long enough, it is worthwhile to delay the survey probe till a number of *sessions* has accumulated. This may compromise the timeliness of emotion label collection, but reduces the number of probes compared to vanilla *Event-based* approach. The proposed ESM schedule, called *Low-Interference High-Fidelity (LIHF)*, triggers a survey request only if (a) event of interest has occurred and (b) a minimum time has elapsed since the last probe.

Algorithm 1: *LIHF* Experience Sampling Method

Input: *EventLog*, *ESMLog*, Events of Interest (E)
Output: ESM Probe
while *true* **do**
    $N \leftarrow$ Set of events in *EventLog* table
    $P \leftarrow N \cap E$
    if $(P! = \emptyset)$ **then**
        t1 $\leftarrow$ Timestamp of last ESM probe
        t2 $\leftarrow$ Current timestamp
        if *time difference* $(t2, t1) > W$ **then**
            F $\leftarrow$ Current foreground application
            if $F = LockScreen$ **then**
                Do not fire ESM probe
            else
                1. Fire ESM probe
                2. Update ESM probe timestamp in *ESMLog*
                table
    Sleep T seconds;

The *LIHF* scheduling, as shown in Algorithm 1, monitors and records a collection of pre-defined events, N, in the *EventLog*. Whenever an event belonging to E (say an application switch after sufficient typing) is detected in *EventLog*, the algorithm checks whether a survey should be triggered. The *ESMLog* maintains timestamp of each survey request. A new survey is scheduled only after time, W since the last probe. All the events between two successive surveys are labeled with the input provided by the user. However, it may so happen that once a probe is about to be issued, the screen is locked which may delay the label collection. Such labels that suffer unplanned delay can be filtered out in a post-processing step.

The frequency of the probe is regulated by configuration parameters T and W. By introducing W, we ensure that there is a minimum gap between the two successive ESM probes,

and limit the survey frequency. Choice of W also dictates how close to the event the label is collected, and should be carefully selected to prevent recall bias. To prevent continuous polling for events of interest, the parameter T is chosen to delay the event collection.

IV. METHODOLOGY

Apart from verifying survey fatigue and timeliness of label collection, we also verify whether emotion labels collected using *LIHF* ESM policy improves the emotion detection quality. For this purpose, we design a typing based emotion detection system and integrate the *LIHF* ESM schedule into it for emotion ground truth collection.
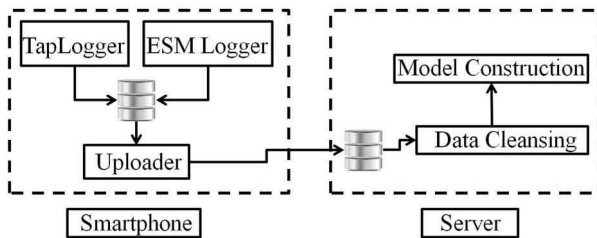

Fig. 3: *TapSense* Architecture

A. Apparatus

We design and implement an Android application, *TapSense*, as shown in Figure 3. It consists of the following key components.

- *TapLogger* module tracks user's typing activity. It is implemented by installing an instrumented QWERTY keyboard using Android IME facility [16]. We collect the timestamp of each tap event, as well as, record the associated application name and any non alphanumeric character typed in the *EventLog*.
- *ESMLogger* implements the *LIHF* ESM strategy. Instead of considering every switch from a typing based application as an event, we consider switching application after typing a pre-defined amount of text, L, as an event. Periodically after T seconds, it is checked if there is an application switch. If there is a switch, and the last ESM probe was triggered W minutes ago, then a new survey is scheduled. We set the values of these configuration parameters as, $L = 80$ characters, $W = 30$ minutes and $T = 15$ seconds. Both *TapLogger* and *ESMLogger* modules run on the smartphone and the *Uploader* module sends the collected data to the server via WiFi.
- *Data Cleansing* steps are performed at the server end to remove the emotion labels marked as *No Response*. We also remove those *sessions* and associated emotion labels for which the elapsed time between typing completion and label collection is more than 3 hours.
- *Model Construction* is done at the server end. We build a simple, single-feature personalized emotion detection model using decision tree (C4.5) in Weka. For every *session*, we compute the time interval between consecutive tap events,

defined as Inter-Tap Distance (ITD). We calculate the *mean session ITD* (a signature of typing speed) for every *session* by computing average of all ITDs present in a *session* and use it as the only feature for model construction.

B. Field Study

A group of 15 university students (aged between 18-24, 12 male, 3 female) were recruited for the study. We installed the application on their smartphones and asked them to use the app for recording emotion for two weeks. We informed the participants that once they change the application after typing, they may receive a pop-up, where they have to record the current emotion state. It was also advised to skip emotion recording by selecting *No Response* if the pop-up appears at an inopportune time.

V. DATASET

We recorded a total of 1291 survey requests corresponding to 2156 *typing sessions* using *LIHF* ESM. Only one user marked 2% of the labels as *No Response*, showing a considerable improvement compared to that of *Event-based* ESM.

A. Feasibility of Typing Speed for Emotion Detection

In order to verify whether *LIHF* ESM can improve the quality of emotion prediction, we first test the feasibility of using typing speed for emotion detection. In [17], Lee et al. have used various typing features including typing speed for emotion detection for Twitter client. We verify the same in our dataset using typing speed as the only cue for emotion detection. We compute *mean session ITD*s of each *session* for individual emotion states. In Figure 4, we have shown *mean session ITD*s of different *sessions* using boxplot for 4 representative users.
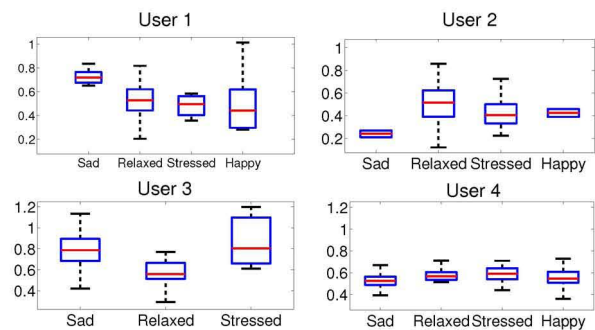

Fig. 4: Distribution of ITD for different users. Emotion states and *Mean session ITD*s in seconds are plotted along X and Y-axis respectively.

We observe that for user 1 and 2, *mean session ITD*s for Sad is distinctly different from other emotion states, whereas for user 3, same can be observed for Relaxed state; however no such emotion state discrimination can be observed for user 4. We saw that out of 15 users, 9 users have variation in *mean session ITD*s across the emotion states. In order to validate the same statistically, we group the *mean session ITD*s of similar emotion category for every user and perform a between-group

one-way ANOVA test. The test result for each of the 9 users validates the hypothesis that *mean session ITD* corresponding to at least one emotion state is different from that of the other emotion states. This is confirmed at confidence level 95% yielding *p*-value < 0.05.

R - H	S - H	T - H	S - R	T - R	S - T
4	8	5	8	3	7

TABLE II: Tukey HSD test overview to identify the number of users having significantly different *mean session ITD* for a given emotion pair. A user can have multiple distinguishable emotion pairs. R,H,S,T denote emotion states *Relaxed,Happy,Sad,Stressed* respectively.

In order to verify the difference in *mean session ITD* for any emotion pair, we perform post hoc comparisons using the Tukey HSD [18] for these 9 users. Table II summarizes Tukey HSD test result. Every cell indicates the number of users having distinctly different typing speed for the given emotion pair. A user can have multiple distinguishable emotion pair. Depending upon the distribution of emotion samples, better emotion classification accuracy is expected for users with higher number of distinguishable emotion pairs. Finally, we conduct the data cleansing step on these filtered users, resulting in 650 probes (corresponding to 888 *sessions*) for analysis.
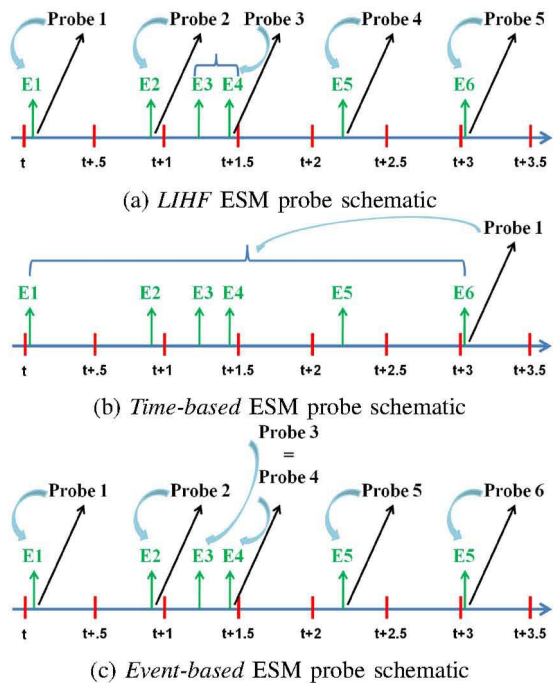

(a) *LIHF* ESM probe schematic

(b) *Time-based* ESM probe schematic

(c) *Event-based* ESM probe schematic

Fig. 5: Schematic to display how *Event-based* and *Time-based* traces are generated from *LIHF* ESM.

B. ESM Trace Generation

In order to compare the different ESMs one must collect emotion labels under identical conditions. In a human centric in-situ experiment it is impossible to control all parameters. Hence we generate trace for *Time-based* ESM and *Event-based* ESM from the data collected using *LIHF* ESM.

In Figure 5, a schematic is given to depict how *Event-based* and *Time-based* traces are generated from *LIHF* ESM. E_i denotes the application switching event after sufficient typing. In case of *LIHF* ESM, there are 6 such events, however only 5 probes were issued (Figure 5a). No probe is issued after $E3$ because it occurs within time-window ($W = 30$ minutes) since last probe (Probe 2). In order to generate the equivalent *Time-based* trace, probes are considered at 3 hour interval. As a result, there will be only one probe Probe 1 and all events $E1$ to $E6$ will be tagged with it (Figure 5b). But in case of conversion to *Event-based* ESM, all events are treated separately, as a result there will be in total 6 probes (Figure 5c).

Next, we define the transformation procedure formally. We represent the corresponding application, emotion label and timestamp of label collection for a *session* as triplet $e_i = < A_i, l_i, t_i >$, where A_i, l_i represent to the applications and emotion label associated with a *session* i respectively, and t_i represents the corresponding timestamp of label collection.

1) Generation of Equivalent Time-based Trace: We consider the sampling interval of *Time-based* ESM as 3 hours. We parse every association of *LIHF* ESM e_j and find all e_i such that $(t_j - t_i) <= 3$ and produce the equivalent association $e_i' = < (A_i \cup A_j), l_j, t_j >$ for *Time-based* ESM. In summary, all *sessions* that are traced within 3 hours are considered as part of a single *session* and labeled with the latest emotion state. This results in total 366 probes (corresponding to 888 *sessions*) for *Time-based ESM*.

2) Generation of Equivalent Event-based Trace: We consider changing application after typing as an event. We scan every association of *LIHF* ESM e_i and produce equivalent associations e_i' of *Event-based* as follows. $\forall a_i \in A_i$, group a_i sequentially such that $a_i = a_{i+1} = ... = a_k$ and $a_k \neq a_{k+1}$ and e_i' is defined as $e_i' = < \{a_i, a_{i+1}, ..., a_k\}, l_i, t_i >$. Evidently, as long as the application is same, it is considered part of a single *session* and whenever the application is changed, it is considered as a different *session*. This results in total 888 *sessions* or probes for *Event-based* ESM.
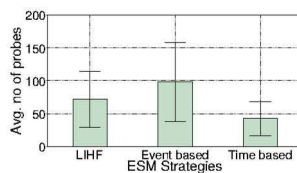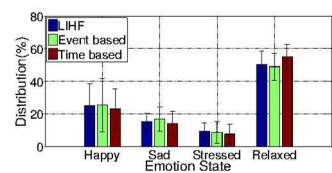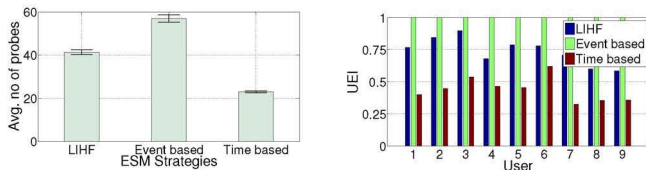

Fig. 6: Average number of probes per user

Fig. 7: Emotion distribution

In Figure 6, we display the average number of probes fired per user for every ESM. In order to compare the ESM strategies, it is also required to verify that the emotion labels are present in equivalent proportion for each ESM approach. We verify the same in Figure 7.

VI. RESULTS

We compare three ESM approaches in terms of (a) intrusiveness or survey fatigue, (b) timeliness of emotion label collection and (c) accuracy in classifying emotion states.

(a) Average number of probes per day

(b) *UEI* Comparison

Fig. 8: Intrusiveness comparison across ESM strategies

(a) Average elapsed time

(b) *RoL* Comparison

Fig. 9: Comparing Recency of Label (RoL) Collection across ESM Strategies

A. Evaluation Metrics

We define the following metrics for comparison.

1) User Engagement Index: The intrusiveness of an ESM approach is measured using the User Engagement Index (*UEI*), defined as follows. Let n_h^i, n_t^i and n_e^i denote the total number of ESM probes fired in *LIHF*, *Time-based* and *Event-based* ESM respectively for a user 'i'. Then UEI_i for user 'i' is,

$$UEI_i = \frac{N_i}{max(n_t^i, n_e^i, n_h^i)}$$

where N_i is n_t^i or n_e^i or n_h^i corresponding to the ESM approach.

2) Recency of Label: The timeliness of label collection is measured using Recency of Label (*RoL*). *RoL* is the elapsed time between an event, i.e. completion of typing followed by application switch, and the survey request to collect emotion label. *RoL* is computed as follows. Let d_h^i, d_t^i and d_e^i denote the average elapsed time between typing and labeling for *LIHF*, *Time-based* and *Event-based* ESM respectively for a user 'i'. Then RoL_i of user 'i' is,
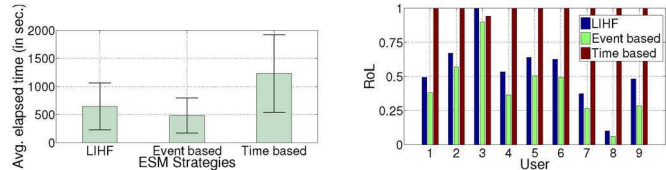
$$RoL_i = \frac{D_i}{max(d_t^i, d_e^i, d_h^i)}$$

where D_i is d_t^i or d_e^i or d_h^i corresponding to the ESM approach.

3) Classification Performance: The influence of ESM approach on emotion classification is evaluated using 10-fold cross validation based on decision tree based classification model generated using the data corresponding to each ESM type. We partition the *sessions* present in the dataset of every user into 10 sets containing roughly equal number of *sessions*. We use one set as test data and train the model using *sessions* from remaining sets. This is repeated for 10 times by leaving out a different set every time and the classification accuracy is computed as an average of each iteration.

Accuracy of classification is computed as follows.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

where TP, FP, FN, TN means true positive, false positive, false negative, and true negative respectively for emotion detection. *Accuracy* denotes the number of emotion states inferred correctly based on typing characteristics. The accuracy indirectly depends on the fidelity of the labels collected using ESM survey.

B. How intrusive is the LIHF ESM approach?

The number of survey requests issued by each ESM schedule is compared in Figure 8a. We observe that *Time-based* ESM issues minimum number of probes, *Event-based* ESM issues maximum number of probes while *LIHF* ESM lies in between. We also compare the intrusiveness for each user using the *User Engagement Index* metric in Figure 8b. For all users, *UEI* for *LIHF* ESM is lower than that of *Event-based* ESM. Across all users, there is an average improvement of 26% in *UEI*. *Time-based* ESM is the best in terms of *UEI*, but is less accurate, as shown later. As *LIHF* ESM schedule enforces a minimum elapsed time between two successive probes, it generates less number of probes and reduces intrusiveness compared to *Event-based* ESM approach.

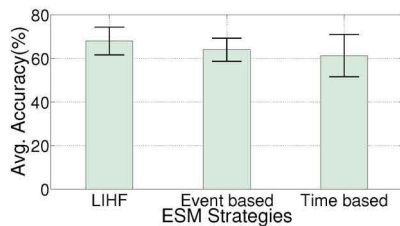### C. Are labels collected close to an event?

If an emotion label from user is collected as soon as an event has occurred, it reduces recall bias. We compare the average elapsed time in recording an emotion label for each ESM schedule, as shown in Figure 9a. *Event-based* ESM has the least average elapsed time, whereas *Time-based* ESM has the highest average elapsed time. *LIHF* ESM policy trades off between these. For individual users, the *Recency of Label (RoL)* shows the elapsed time before label collection, as shown in Figure 9b. We observe for every user (except U3), average elapsed time has reduced by about 50% compared to *Time-based* ESM, whereas it is slightly higher (10%) than *Event-based* ESM. Thus *LIHF* ESM is closer to *Event-based* approach in terms of timeliness of data collection, while reducing the number of survey requests.

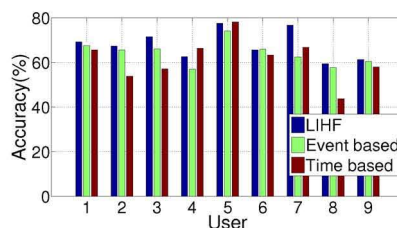### D. Does survey schedule influence emotion classification?

We build 3 different decision tree based classifier based on the data corresponding to each ESM schedule. We compute *mean session ITD* for every *session* and use it as a feature for classifying emotion. The average accuracy of emotion classification using 10-fold cross-validation is shown in Figure 10a. We also report the classification accuracy for each user in Figure 10b.

It is observed that *LIHF* ESM performs best in recognizing the emotion states. We have ensured that there is uniformity in distribution of emotion samples collected using different ESM approaches (Figure 7). Given all other factors to be invariant, except ESM survey schedule, accuracy of classification can be affected by the choice of ESM strategy. There can be different reasons, like survey fatigue, latency in collecting emotion labels because of which emotion classification is impacted.

(a) Average classification accuracy

(b) Classification accuracy for individual users

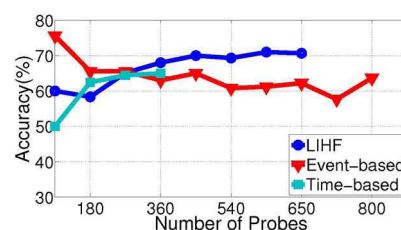Fig. 10: Comparing accuracy for different ESM approaches

Fig. 11: Accuracy comparison with number of probes

E. Trade off between study duration and emotion classification

Figure 11 depicts how the number of samples influence quality of emotion prediction. It also helps to understand the tradeoff between study duration and emotion classification.

For *Time-based* ESM, initially the accuracy is low and it improves once sufficient samples of each type of emotion state is collected. But based on Figure 8a, since *Time-based* ESM has least probing rate, collecting equal number of samples will take the longest duration for it. On the contrary for *Event-based* ESM the initial accuracy is very high but it drops gradually. Initial high accuracy can be attributed to the presence of only one or two emotion states. As a result, the classification is easy, but with more samples the classification performance degrades. Since *Event-based* ESM has highest probing rate, it can collect same number of *sessions* earliest, but these labels may not reveal user's actual emotion response because of survey fatigue, resulting in poor classification. However, in case of *LIHF* ESM schedule, the initial accuracy is not as high as that of *Event-based* ESM, but it surpasses *Event-based* ESM once it collects sufficient number of *sessions* (close to 300). Since the probing rate of *LIHF* ESM schedule is comparatively low, it will take more time to collect same number of samples, but the quality of the labels will be high resulting in better classification performance.

VII. CONCLUSION AND FUTURE WORK

Main objective of this study is to design an ESM schedule, which can reduce survey fatigue, collect emotion labels timely as well as maintain the fidelity of collected ground truth. Towards that goal, we proposed an ESM schedule, called *Low-Interference-High-Fidelity (LIHF)* and compared it's performance with conventional ESM schedules like *Time-based* and *Event-based* approaches. We integrated the *LIHF* ESM into a typing based emotion detection Android application *TapSense* and conducted 2 week long study, which revealed that the proposed ESM outperforms the conventional ESMs in all 3 aspects - intrusiveness, timeliness and accuracy of emotion classification.

Modeling the relationship analytically between data quality and survey frequency and timeliness can result in better personalized ESM schedules. Also, combining diverse context information, like activity, location while triggering a probe can help in reducing survey probes and be more successful in getting user attention leading to accurate user responses.

REFERENCES

[1] T. S. Conner, H. Tennen, W. Fleeson, and L. F. Barrett, "Experience sampling methods: A modern idiographic approach to personality research," *Social and personality psychology compass*, vol. 3, no. 3, pp. 292–313, 2009.

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

[3] A. Kapoor and E. Horvitz, "Experience sampling for building predictive user models: a comparative study," in *Proceedings of ACM SIGCHI*, 2008.

[4] L. F. Barrett and D. J. Barrett, "An introduction to computerized experience sampling in psychology," *Social Science Computer Review*, vol. 19, no. 2, pp. 175–185, 2001.

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

[6] A. Gaggioli, G. Pioggia, G. Tartarisco, G. Baldus, D. Corda, P. Cipresso, and G. Riva, "A mobile data collection platform for mental health research," *Personal and Ubiquitous Computing*, vol. 17, no. 2, pp. 241–251, 2013.

[7] "Personal analytics companion," https://www.pacoapp.com/.

[8] S. Consolvo and M. Walker, "Using the experience sampling method to evaluate ubicomp applications," *IEEE Pervasive Computing*, vol. 2, no. 2, pp. 24–31, 2003.

[9] J. A. Russell, "A circumplex model of affect," *Journal of Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, 1980.

[10] R. LiKamWa, Y. Liu, N. D. Lane, and L. Zhong, "Moodscope: Building a mood sensor from smartphone usage patterns," in *Proceedings of 11th ACM Mobisys*, 2013.

[11] M. Pielot, T. Dingler, J. S. Pedro, and N. Oliver, "When attention is not scarce-detecting boredom from mobile phone usage," in *Proceedings of ACM UbiComp*, 2015.

[12] R. Wang, F. Chen, Z. Chen, T. Li, G. Harari, S. Tignor, X. Zhou, D. Ben-Zeev, and A. T. Campbell, "Studentlife: assessing mental health, academic performance and behavioral trends of college students using smartphones," in *Proceedings of ACM UbiComp*, 2014.

[13] Y. Gao, N. Bianchi-Berthouze, and H. Meng, "What does touch tell us about emotions in touchscreen-based gameplay?" *ACM Trans. on Computer Human Interactions*, vol. 19, no. 4, Dec. 2012.

[14] H.-J. Kim and Y. S. Choi, "Exploring emotional preference for smartphone applications," in *IEEE CCNC*, 2012.

[15] K. K. Rachuri, M. Musolesi, C. Mascolo, P. J. Rentfrow, C. Longworth, and A. Aucinas, "Emotionsense: A mobile phones based adaptive platform for experimental social psychology research," in *Proceedings of 12th ACM UbiComp*, 2010.

[16] "Android ime," http://developer.android.com/guide/topics/text/creating-input-method.html.

[17] H. Lee, Y. S. Choi, S. Lee, and I. Park, "Towards unobtrusive emotion recognition for affective social communication," in *IEEE CCNC*, 2012.

[18] D. Lane and N. Salkin, "Tukey's honestly significant difference (hsd)," *N. Salkind der.), Encyclopedia of Research Design içinde, Thousand Oaks, CA: SAGE Publications*, vol. 1, pp. 1566–1571, 2010.