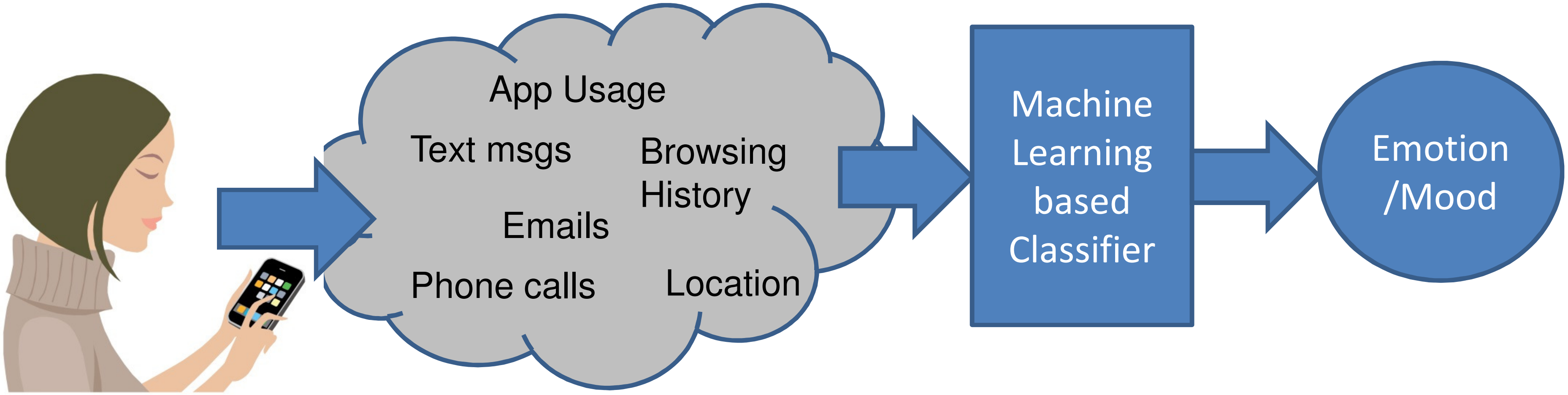
Towards Designing an Intelligent Experience Sampling Method for Emotion Detection

Surjya Ghosh, Niloy Ganguly, Bivas Mitra
Indian Institute of Technology, Kharagpur

Pradipta De
Georgia Southern University

Smartphone-based Emotion Detection



- MoodScope: detects multiple mood states
- Lee et al. (CCNC 2012): Uses different sensors to collect context, and a modified Twitter app to gather touch behavior
- MouStress: detects stress behavior from mouse usage patterns

Assumption: It is possible to collect the ground truth (or emotion labels) reliably

Collecting Emotion Labels

- Experience Sampling Methods
  - [*Time-based*] Periodically ask the user to record the emotion
  - [*Event-based*] Detect a context (or event) to trigger a questionnaire to record emotion
- What if the requests are too frequent or misses important events
  - User may respond falsely
  - User may not respond at all
  - Quality of classification may drop

Can we design an intelligent ESM, which reduces survey fatigue and collects emotion labels timely ?

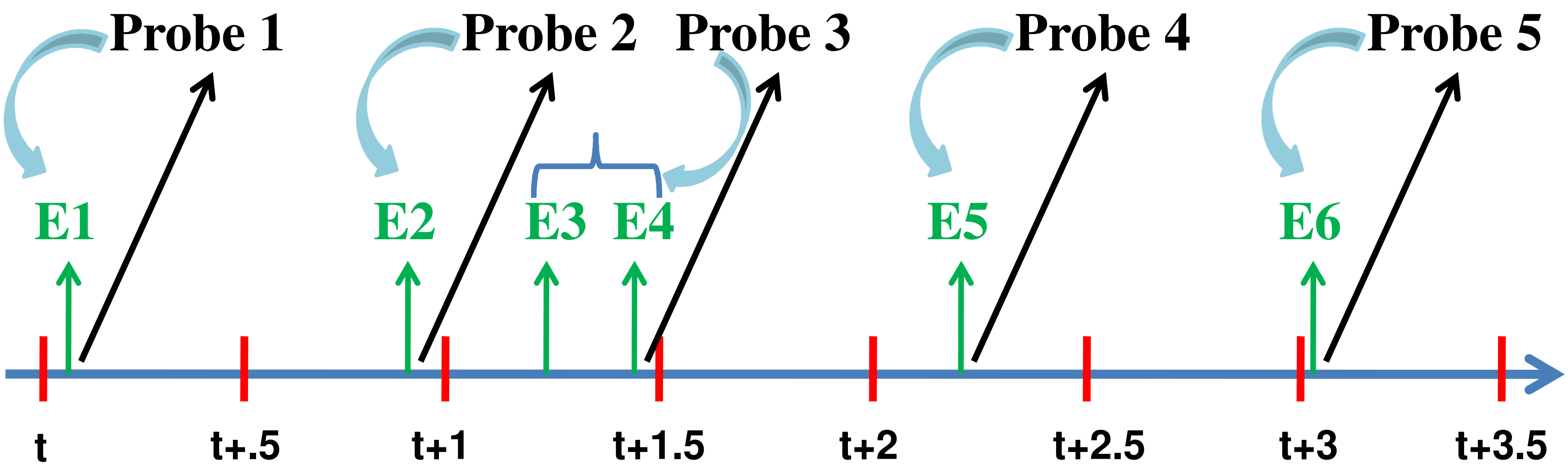
Outline

- LIHF ESM
- Case Study : TapSense
  - Scenario
  - Architecture
- DataSet
  - User segregation
  - ESM Trace Generation
- Evaluation
- Conclusion and Future Work

Limitations of Conventional ESM

ESM Schedule	Weakness
Time-based	-High elapsed time between label collection and occurrence of event -Possibility of missing out important event if the sampling interval is high
Event-based	- May issue too many probes if the app change occurs too frequently

LIHF Experience Sampling Method

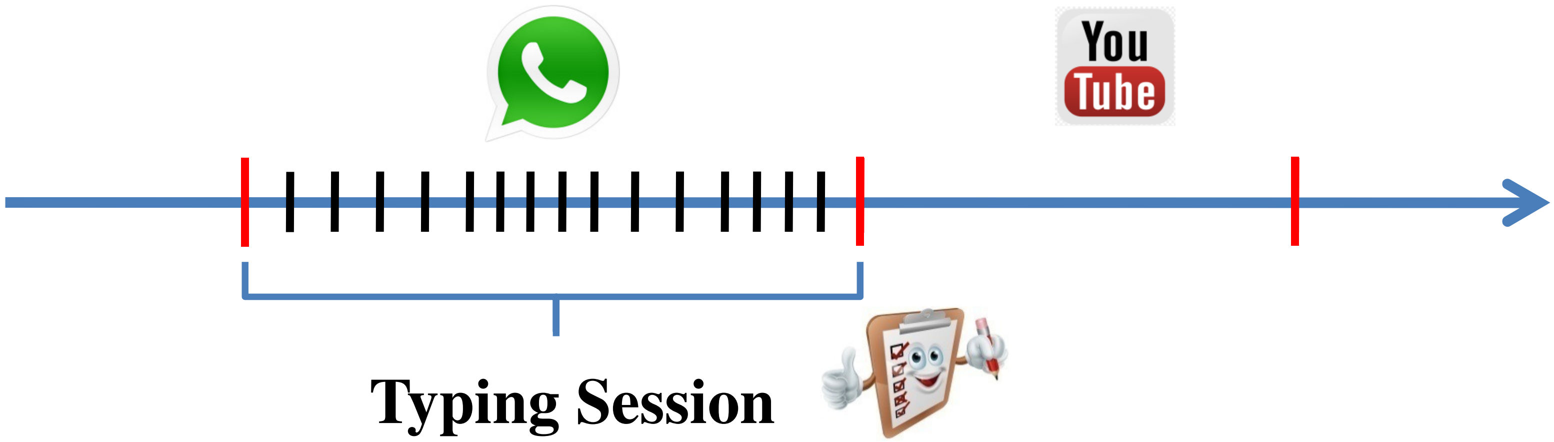


- *Low Interference High Fidelity (LIHF)* ESM
- Probe will be issued only if
  - An event has occurred *and*
  - A minimum time (say 30 mins) has elapsed since last probe

Case Study: TapSense App

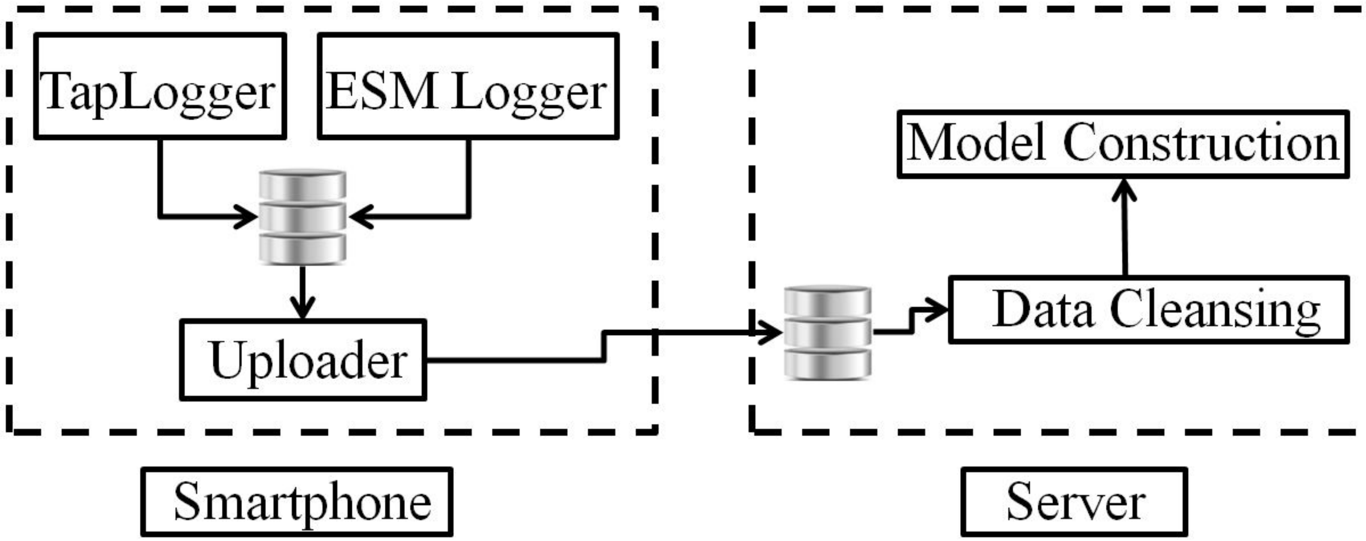
- An app that tracks the typing pattern of a user
  - Typing based Emotion detection system
- Design an ESM, which
  - reduces user engagement
  - collects emotion labels timely
  - yet produces reasonable emotion classification

Example Scenario



- Inter-Tap Distance (ITD)
  - Elapsed time between entering two character is *ITD*
- *Mean Session ITD*
  - Compute mean of all *ITD*s in a session, which is known as *Mean Session ITD*
  - Representation of typing speed

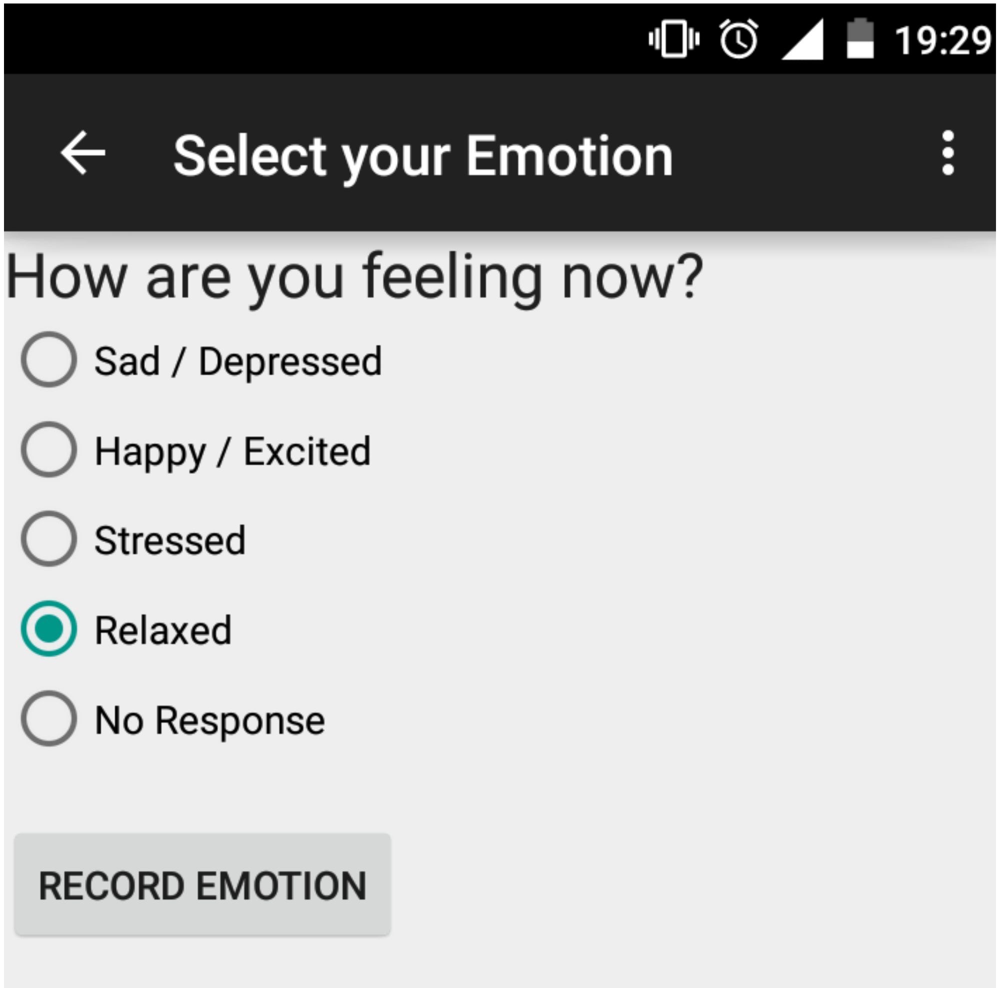
System Architecture



TapSense System Architecture

- Taplogger
  - Tap Data collection
- ESMLogger
  - Implements LIHF ESM
- Model Construction
  - Personalized, decision-tree based

Survey Collection Interface



Higher “No Response” may indicate that the user is not engaging ➔ the user was probed at an inopportune time.

DataSet

- Study duration – 2 Weeks
- Number of users – 15
  - University students
  - 12 males, 3 females, aged between (24 – 33) years
- Data collected
  - 1291 survey requests corresponding to 2156 typing sessions
  - Only one user marked 2% of labels as “No Response”
    - Sharp contrast to Event-based Sampling where large number of users marked “No Response”

User Identification

- Computed mean session ITD from every typing session
- Performed ANOVA test
- For 9 users, the test reveals $p < .05$

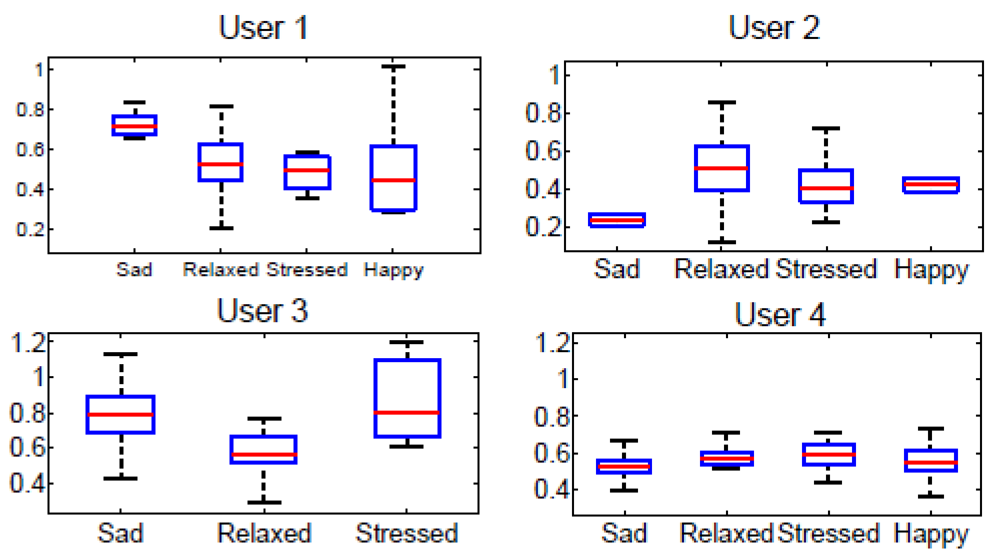
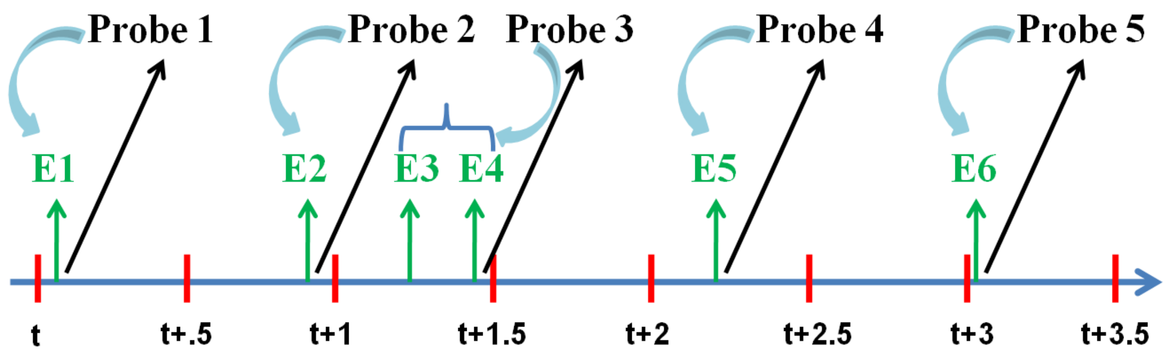
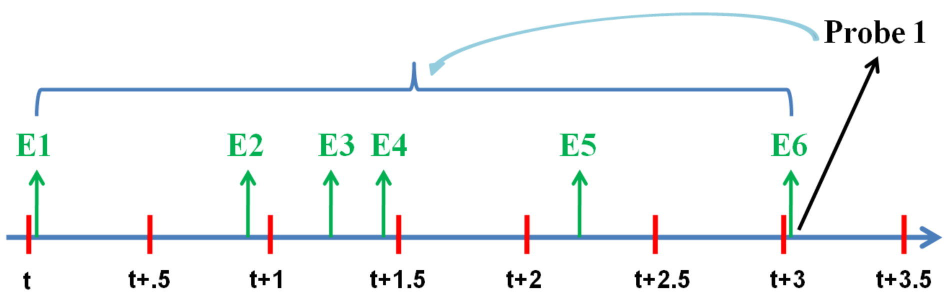


Fig. 4: Distribution of ITD for different users.Emotion states and *Mean session ITDs* in seconds are plotted along X and Y-axis respectively.

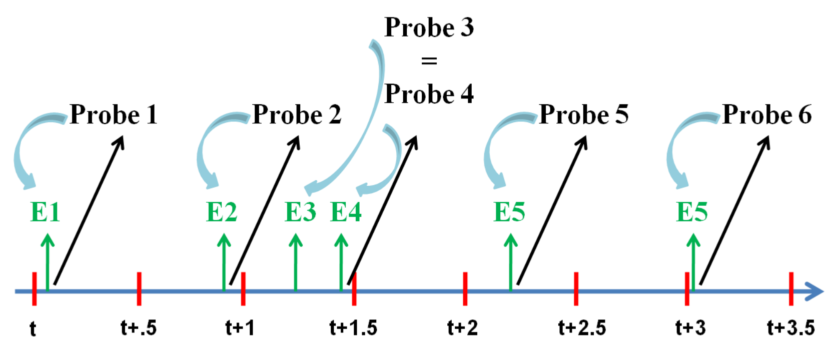
ESM Trace Generation



LIHF ESM Probes



Equivalent Time-based ESM Probes



Equivalent Event-based ESM Probes

Evaluation

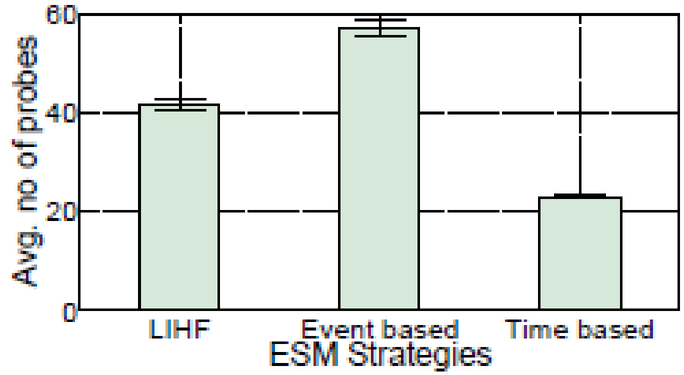
- Evaluation Metrics
  - User Engagement
    - Compares intrusiveness in terms of number of probes issued
  - Timeliness of Labels
    - Measures how close to the event, the probes is issued
    - Elapsed time between typing and label collection
  - Classification Accuracy
    - Measures performance of emotion classification

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

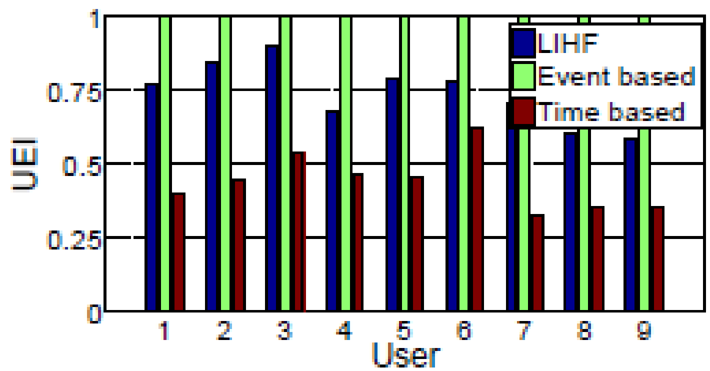
Evaluation Metrics

ESM Type	# of probes	Avg. elapsed time	UEI	RoL
Event-based	n_e	d_e	n_e / max(n_e, n_t, n_h)	d_e / max(d_e, d_t, d_h)
Time-based	n_t	d_t	n_t / max(n_e, n_t, n_h)	d_t / max(d_e, d_t, d_h)
LIHF	n_h	d_h	n_h / max(n_e, n_t, n_h)	d_h / max(d_e, d_t, d_h)

How intrusive is the LIHF ESM approach?



(a) Average number of probes per day



(b) *UEI* Comparison

Fig. 8: Intrusiveness comparison across ESM strategies

In case of LIHF ESM, there is an average improvement of 26% in UEI with respect to Event-based ESM

Are labels collected close to an event?

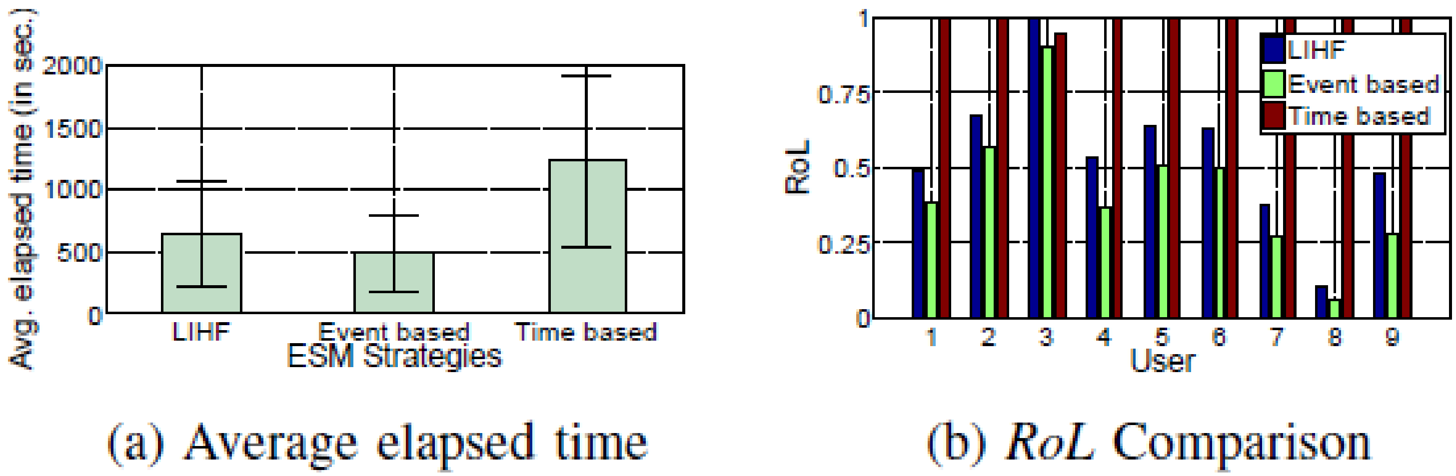
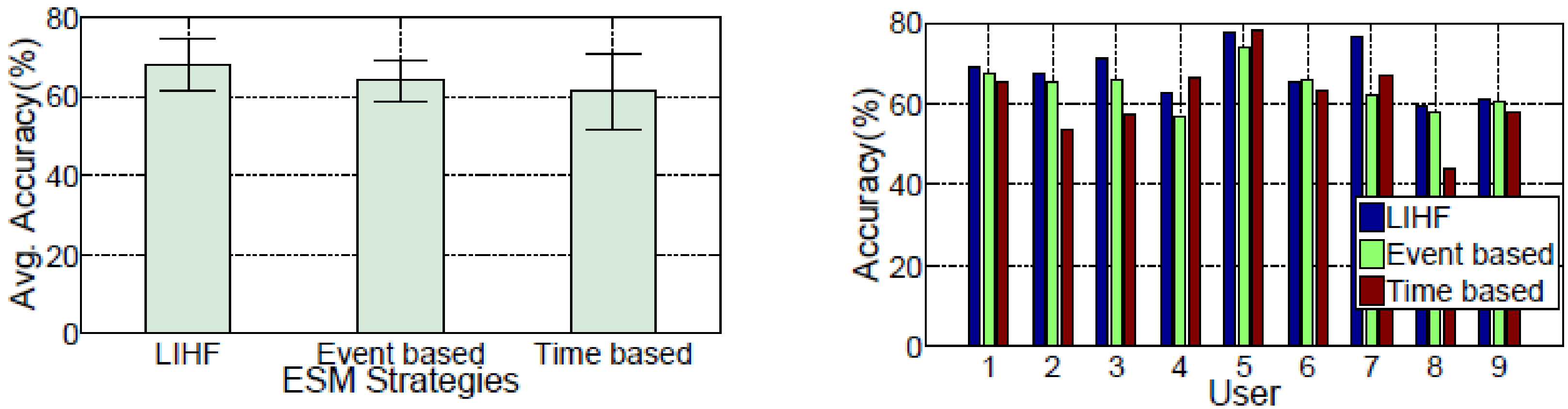


Fig. 9: Comparing Recency of Label (RoL) Collection across ESM Strategies

In case of LIHF ESM, average elapsed time is reduced by 50% with respect to Time-based ESM

Does ESM schedule influence emotion classification?



(a) Average classification accuracy (b) Classification accuracy for individual users

Fig. 10: Comparing accuracy for different ESM approaches

LIHF ESM performs best in recognizing the emotion states

Trade off between study duration and emotion classification

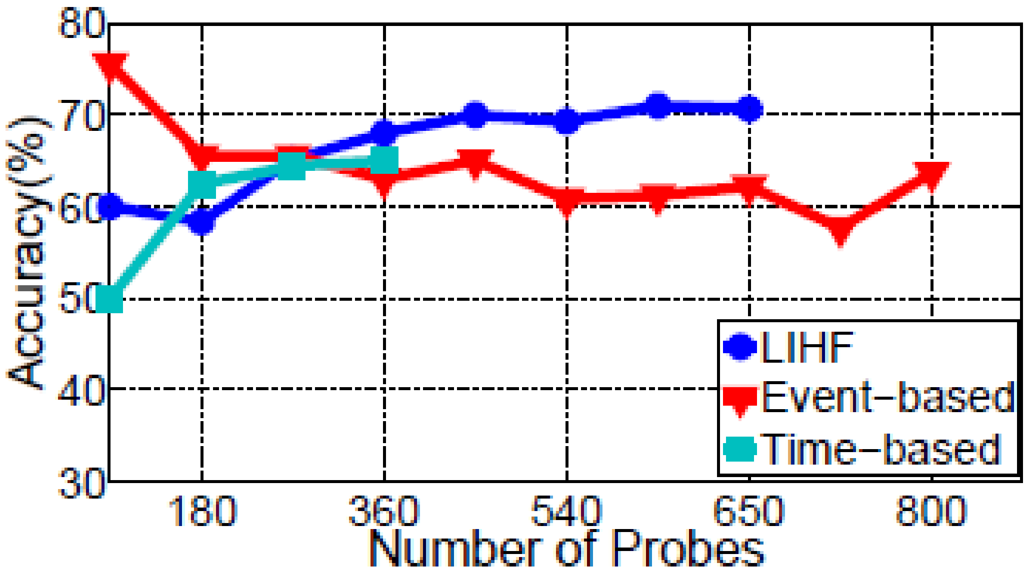


Fig. 11: Accuracy comparison with number of probes

LIHF ESM outperforms others once sufficient labels are collected

Conclusion

- Proposed a new ESM techniques which trades of between Time-based and Event-based ESM
- Validated the ESM using a Typing-based emotion detection system, which indicates using proposed ESM there is
  - 26 % reduction in survey fatigue
  - 50% improvement in timely label collection
  - 8% improvement in emotion classification accuracy


THANK YOU