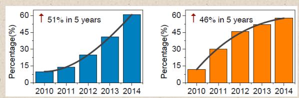
Effectiveness of Deep Neural Network Model in Typing-based Emotion Detection on Smartphones

Surjya Ghosh*, Niloy Ganguly*, Bivas Mitra* & Pradipta De#

*Department of Computer Science & Engineering, Indian Institute of Technology, Kharagpur, India
#Department of Computer Sciences, Georgia Southern University, USA

Introduction

- **Background**
  - Large number of emotion related apps due to the advances in Affective Computing (e.g. Emotion-aware music player, personality-aware OSN service)
  - Ubiquity of smartphone makes it suitable for emotion inference



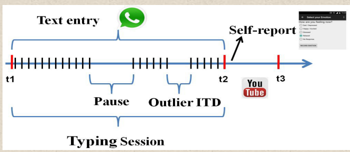
Growth pattern of emotion related app for Android and iOS [1]

- **Motivation**
  - Typing activity on smartphone carries emotion signature [2, 4]
  - Monitoring typing is unobtrusive, low-resource overhead
  - Deep Neural Network (DNN) is very efficient in similar tasks like facial expression detection
- **Problem Statement**
  - Can DNN be effectively used to infer multiple emotions based on typing on smartphone ?
  - What are the resource implications of deploying such a DNN based model ?

Methodology

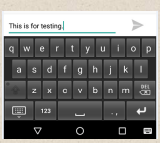
- **Typing-based Emotion Detection Scenario**

- Identify typing session
  - time spent on a single app uninterruptedly
  - extract typing features
- Collect emotion self-reports
  - four emotions based on Circumplex model
  - *happy, sad, stressed, relaxed*
- Construct model to detect emotion
  - combining typing features and self-reports

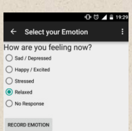


Schematic of typing-based emotion detection

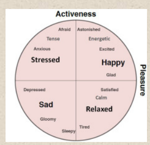
- **Data Collection**



Application keyboard



Emotion self-report collection

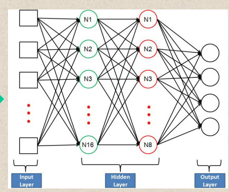


Circumplex model [3]

- **Model Construction**

Feature name	Feature description
MSI (Mean session ITD)	Avg. of all ITDs in the session
RMSI (Refined MSI)	Avg. of non-outlier ITDs in the session
Itd^{i}_{per} , $i \in \{25, 50, 75, 90\}$	i^{th} percentile value of ITDs in the session
Mean_word_time	Average time to complete a word
Std_word_time	Std dev. of word completion times
Session_dur	Duration of the session
Pause_time	Sum of ITDs greater than 30 secs.
No_pauses	No of ITDs greater than 30 secs.
Abs_session_dur	Session_dur - Pause_time
Dur_per_char	Session_dur / No. of chars in a session
Dur_per_word	Session_dur / No. of words in a session
Backsp_{per}	% of backspace in the session
$\mathrm{Splchar}_{per}$	% of non-alphanumerics in the session

Set of typing features



Personalized DNN model

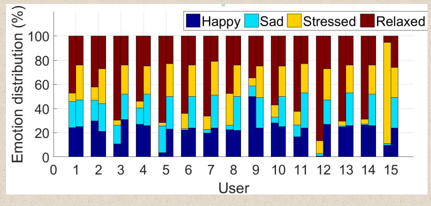
- 16-size feature as input
- Softmax activation with cross-entropy loss
- Dropout is used for regularization

Field Study and Dataset

- Android based application used as experiment apparatus to trace user's typing and collect emotion self-report
- 15 students (12 male, 3 female, aged between 24-33 years)
- 3-week in-the-wild study
- Installed the app in the smartphone of the volunteers for collecting typing details and emotion self-reports

- Total typing sessions: 8301
- Average number of typing sessions per user: 553

- Distribution of emotion samples is found to be skewed as users often reported *relaxed* state
- Sample imbalance is overcome using SMOTE



Distribution of emotion self-reports for every user. For every user, first bar shows the distribution in original data and the second one shows the distribution after applying SMOTE. All emotions are almost equally distributed after applying SMOTE.

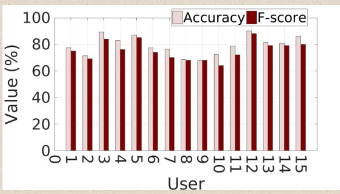
Experiment Setup

- Used batch size of 8 and dropout rate of 0.2 (based on grid search) for superior performance
- Performed 10-fold cross-validation and measure emotion classification performance
  a) Accuracy b) F-score

- Deployed the trained model on the device to measure resource overhead
  a) CPU utilization b) Memory consumption c) Inference time

Evaluation: Emotion Classification

- Average accuracy: 80% (std dev 7.1%) [Min: 69%, Max: 90%]
- Average F-score: 75% (std dev 7%)



User-wise Accuracy, F-score

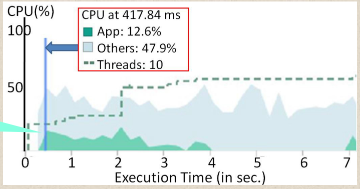
- All emotions except *sad* are identified with an F-score greater than 70%
- *Stressed* emotion is identified with highest F-score (92%)

Emotion-wise F-score

Evaluation: Resource Overhead

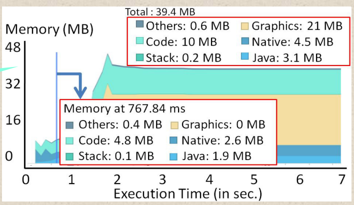
- **CPU Utilization**

Peak CPU Utilization is less than 15% across different smartphones



- **Memory Consumption**

Cumulative memory consumption is less than 40 MB



- **Inference time**

Average inference time ~3 msec.

Parameter	Mean	Std dev
Inference time (in msec.)	3.2	9.6

Conclusion

- Propose a personalized Deep Neural Network model, which can determine four emotion states *(happy, sad, stressed, relaxed)* based on typing features on smartphone
- It returns an average accuracy of 80%, (std dev. 7%)
- Inferring emotion on smartphone using DNN model is not resource-intensive (peak CPU utilization: 15%, cumulative memory consumption: 40 MB, inference time: 3.2 ms)

References

[1] Boyuan Sun, Qiang Ma, Shanfeng Zhang, Kebin Liu, and Yunhao Liu. 2015. iSelf: Towards cold-start emotion labeling using transfer learning with smartphones. In IEEE Infocom.

[2] Ghosh, S., Ganguly, N., Mitra, B., & De, P. 2017. TapSense: combining self-report patterns and typing characteristics for smartphone based emotion detection. In ACM MobileHCI.

[3] James A Russell. 1980. A circumplex model of affect. Journal of Personality and Social Psychology 39, 6(1980), 1161–1178.

[4] Surjya Ghosh, Niloy Ganguly, Bivas Mitra, and Pradipta De. 2017. Evaluating effectiveness of smartphone typing as an indicator of user emotion. In IEEE ACII.

Email: surjya.ghosh@iitkgp.ac.in

Research Group: CNeRG, IIT KGP (http://www.cnergres.iitkgp.ac.in/)

Acknowledgement: We would like to thank the "**Mobicom 2018**" organizers for the student scholarship.