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
- **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

- **Data Collection**
  - Application keyboard
  - Emotion self-report collection
  - Circumplex model [3]

- **Model Construction**
  - Set of typing features
  - Personalized DNN model
  - 16-size feature as input
  - Softmax activation with cross-entropy loss
  - Dropout is used for regularization

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

- **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.

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

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, Qing Ma, Shanfeng Zhang, Kehua Liu, and Yunhao Liu. 2015. iSelf: Towards cold-start emotion labeling using transfer learning with smartphones. In IEEE Infocom.

Acknowledgement: We would like to thank the “Mobicon 2018” organizers for the student scholarship.