User Activity Context Recognition From Smartphone Data Using Deep Neural Networks

Shashank Manjunath[1], Bethany K. Bracken[1], Stan German[1], Camille Monnier[1], Mike Farry[1]

[1]*Charles River Analytics Inc., Cambridge, MA, United States*

Introduction: Currently, 92% of US adults carry mobile phones, and 77% carry smartphones with advanced sensors[1]. Smartphone applications are already being used to identify disease and promote healthy behaviors, but these applications require active participation by the user, such as through taking and uploading pictures to the application. Sensors already integrated into modern smartphones can be used to passively and continuously monitor users without requiring explicit user interaction. Therefore, there is a timely opportunity to develop mobile-based technologies that allow for early detection of disease, and consequently better treatment of the patient. Recent advances in machine learning[2] allow for continuous monitoring of weak signals that indicate physiological state. The goal of this research is to enable continuous, real-time, and accurate assessment of individual's physical state and activity through data captured passively and unobtrusively by cellphone sensors.

Materials and Methods: To perform classification of user state and activity, we propose a modified deep neural network architecture based on DeepSense[3]. DeepSense provides a sensor fusion framework, taking in accelerometer, gyroscope, and audio data from a smartphone.

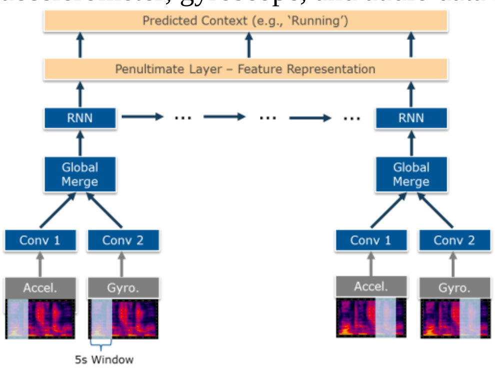

Figure 1: DeepSense network architecture as applied to accelerometer and gyroscope smartphone data. Audio data is fed through the same architecture after conversion to Mel-frequency cepstrum coefficients (MFCC). Deepsense first performs convolution operations on each sensor type, before fusing the outputs of these operations together in a merge layer. A recurrent neural network is then used to analyze temporal correlations before final predictions are made.

We further modified the architecture to enable prediction of multiple co-occurring activities, e.g., the user is both *sitting* and *eating*. To achieve this multi-label classification, we modified the training loss function using a simplifying assumption that activities are independent. We used the UCSD Extrasensory dataset to evaluate performance. This dataset contains over 300,000 minutes of accelerometer, gyroscope, audio data, and sparse annotations for 51 different activities [4]. In our experiments we train on 36 individual users and test on the remaining users. One major issue of this dataset in the multi-label context is extreme class imbalance since a user spends many more samples not performing a particular activity as compared to the number of samples where they do perform the given activity. To mitigate this class imbalance, the loss function is weighted by the ratio of positive to negative occurrences of each activity in the training data[5]. We use balanced accuracy (BA) to evaluate model performance on a particular activity defined as the mean of the true positive rate and the false positive rate.

Results and Discussion: On the ExtraSensory dataset, our DeepSense architecture was able to achieve an average BA of 72.72% over the 51 contexts labelled by users in the dataset, demonstrating that the modified DeepSense architecture can achieve improved classification of user activity. The algorithm was additionally able to achieve >80% BA on 18 activities.

Conclusions: We demonstrated that deep neural network architectures can be applied to user physical activity recognition based on passively collected data. Additionally, we proposed a novel method of predicting multiple user activities using a single sample of user data. Future work includes integrating meta-sensors such as battery level and WiFi status, into the inference pipeline.

Acknowledgements: This material is based upon work supported by United States Air Force and DARPA under Contract No. FA8750-18-C-0056 entitled Health and Injury Prediction and Prevention Over Complex Reasoning and Analytic Techniques Integrated on a Cellphone App (HIPPOCRATIC App). The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government

References: [1] Smith, A. *Pew Research Center*, 2017, vol. 12. [2] Bengio, Y. *arXiv:1206.5538*. 2012. [3] Yao S. *arXiv:1611.01942*. 2016. [4] Vaizman, Y. *IEEE Pervasive Computing*, vol. 16, 62-64. [5] Vaizman, Y. *Proceedings of the ACM on IMWUT*, vol. 1.