HUMAN ACTIVITY RECOGNITION

Ram Sabarish Obla Amar Bapu
Institute of Signal Processing and Systems Theory
University of Stuttgart
Stuttgart, 70569
st169693@stud.uni-stuttgart.de

Swetha Lakshmana Murthy
Institute of Signal Processing and Systems Theory
University of Stuttgart
Stuttgart, 70569
st169481@stud.uni-stuttgart.de

February 16, 2021

ABSTRACT

Physical human activities are necessary to maintain a balanced momentum in an individual's life[1]. The unprecedented changes in the work-life balance have caused an imbalance.[2]. Hence, a system to monitor the Human Activity Recognition(HAR) is ventured in this paper. The Human Activities and Postural Transitions Dataset(HAPT) is a publicly available dataset which is used for this work[3]. The embedded inertial sensors[1] from the smartphone aids in collecting and recording the activity data. The dataset consists of experimental raw data collected from 30 subjects in the age group of 19-48. The 3-axial linear acceleration and angular velocity at a constant rate of 50Hz were captured. There are 12 classes recorded and labeled, viz three static activities (sitting, standing, lying), three dynamic activities (walking, walking downstairs, walking upstairs), and six postural activities (stand-to-sit, sit-to-stand, sit-to-lie, lie-to-sit, stand-to-lie, lie-to-stand). A 12 class Long Short-Term Memory(LSTM) neural network-based human activity classifier is implemented here. Further experimental insights helped in obtaining an overall test accuracy of 79.06%.
Key words: *Human Activity Recognition, HAPT dataset, Time-series data, Twelve Activities, LSTM*

1 Introduction

The transitional activities along with postural and fall detection play a significant role in human activity recognition among the elderly age group[2]. A mobile healthcare system helps in the easy retrieval of important data in times of emergencies. This dataset helps us in catering to the above needs as well as the other listed activities. The actions performed by the user based on the data collected from the embedded sensors in the phone is the main focus of this paper.

Subsequent sections of the paper help us in understanding the workflow. Section 2 introduces us to the dataset and its features. Section 3 helps us in understanding and configuring the input pipeline. Following which the LSTM model architecture used to classify the 12 classes is explained in Section 4. Evaluation and results of the compiled model are recorded in Section 5. The concluding aspects of the topic are seen in Section 6.

2 Data and Features

The dataset used here is HAPT[3]. The total number of subjects is 30 who performed two experiments to record the linear acceleration and angular velocity. The train, test, and validation are split in the ratio of 70:20:10. User1 to 21 accounts to the train data, user 22 to 27 the test data, and the remaining user 28 to 30 is used for validation.

3 Input Pipeline Creation [Swetha]

The raw data from all participants are used as the input data. Both the experiments performed per user is used. The input to the model consists of six channels in total, tri-axial inputs from each of the inertial sensors.

3.1 Visualizing and understanding the data[Ram Sabarish and Swetha]

An example plot for the training data for the user01 and exp01 for all the activities performed is as shown in Figure 1 and test data for user18 and exp37 is as shown in Figure 2. Two activities, walking, and sitting are also shown in Figures 3 and 4 respectively. In all the visualized graphs here, the x-axis denotes the time in seconds, y-axis denote the corresponding Accelerometer and Gyroscope values.

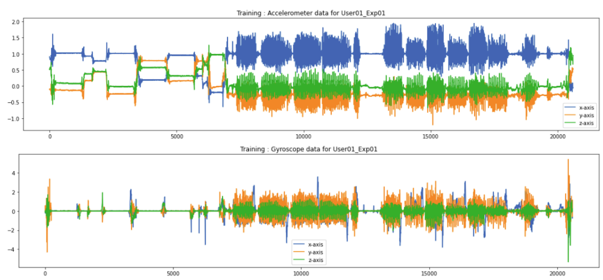


Figure 1: 3-axis Accelerometer-Gyroscope Train sensor values for User01 Exp01 for all 12 activities

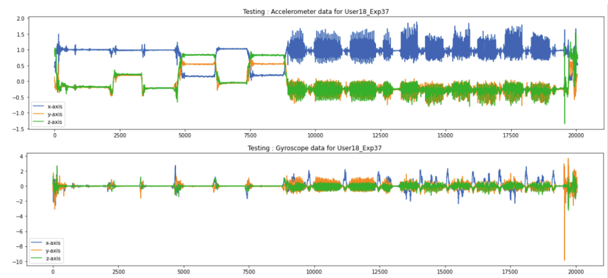


Figure 2: 3-axis Accelerometer-Gyroscope Test sensor values for User18 Exp37 for all 12 activities

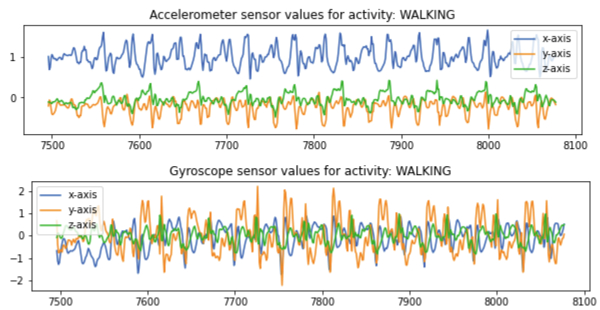


Figure 3: WALKING activity for User01 Exp01

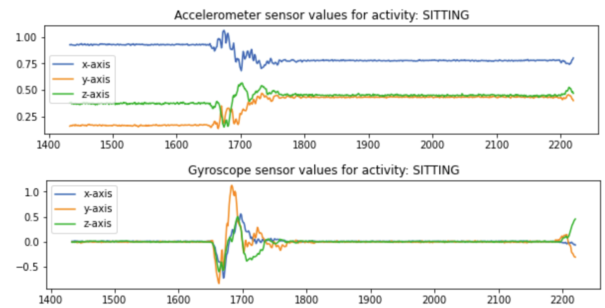


Figure 4: SITTING activity User01 Exp01

3.2 Data Preprocessing[Swetha]

The noisy rows are removed after which the six-channel data is normalized using the Z-Score normalization. The raw score x is converted into a standard score using,

$$z = \frac{x - \mu}{\sigma} \tag{1}$$

μ is the mean of the population , sigma is the standard deviation of the population

The dataset predominantly consists of unlabeled data. These are marked with a label 0. Discarding of the unlabeled data is avoided to prevent the random initialization of weights in the hidden layers. One hot encoding method is used to label all the other performed activities. A window size of 250 is used here to ensure that all activities are not lost during their transitions. Also, overlapping technique of 50% between consecutive windows is employed here.

4 Model Architecture [Ram Sabarish]

LSTM networks are a type of recurrent neural network(RNN) intelligent of learning sequence prediction problems that depend on a particular order. The two technical problems overcome by LSTMs are vanishing gradients and exploding gradients[4]. These two parameters depend on how well a model is trained. The output from the previous step is fed as input to the current step in an RNN. This can be summarized as a feed-forward neural network that has internal memory. This thus helps us in recognizing the upcoming sequence in the given data. LSTM comprises 3 gates namely forget, input, and output gate. The memory and gating mechanism is a preferred choice for dealing with time-series data. Hence the following architecture is used for the HAPT dataset.

The model architecture used in this work comprises of a 256 unit LSTM layer followed by a dropout layer with a dropout rate of 0.2. This is followed by another 128 unit LSTM layer. Finally a 12 unit Dense layer is added to predict the classes. The simple architecture is as shown in Figure 5,

4.1 Hyper-parameter Tuning using HParams [Ram Sabarish]

The above model has been obtained using hyperparameter tuning. This is visualized in tensorboard as shown in Figure 6,

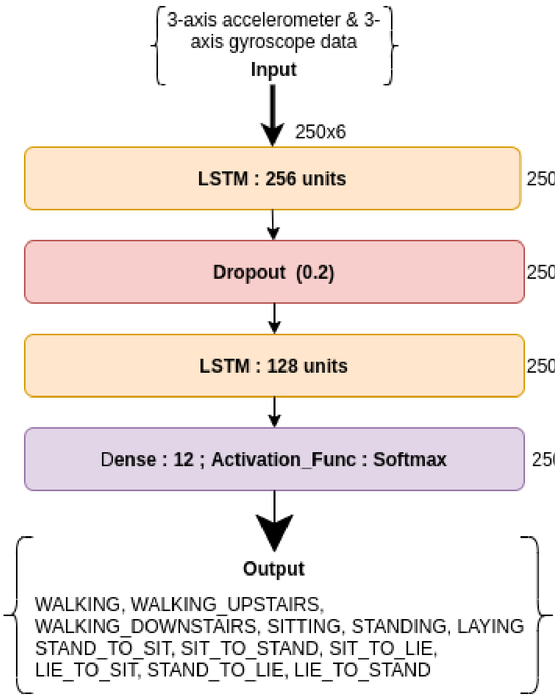
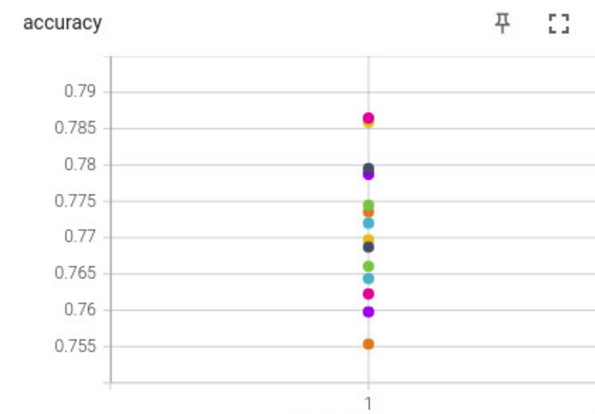


Figure 5: Model Architecture



HP_LSTM_NEURONS	128, 256
HP_EPOCHS	5, 10
HP_OPTIMIZER	RMSProp, Adam
HP_DROPOUT	0.1, 0.2

Figure 6: Tensorbaord Visualization

5 Evaluation and Metrics [Ram Sabarish and Swetha]

The observed overall test accuracy for all the 12 activities is **79.06%**. The training and validation accuracy and loss is as shown in Figure 7. Th number of epochs is 15.

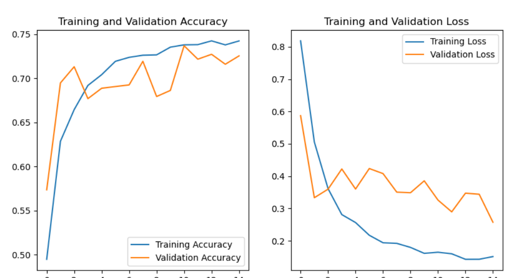


Figure 7: Training and Validation Accuracy and Loss

5.1 Observation

- Activities such as walking upstairs, sitting, lying produce very high f1-score(greater than 0.9).
- The other activities such as walking downstairs, standing also produce satisfactory results.
- Fewer transitional activities such as stand-to-lie and lie-to-stand produce lower results compared to the above listed activities. This can be seen in the Confusion Matrix as shown in Figure 8. Also the f1-score, precision and recall as tabulated as in Table 1.

Table 1: Label-wise precision, recall and f1-score for the 12 Activities

Activity	precision	recall	f1-score
WALKING	0.95	0.53	0.68
WALKING_UPSTAIRS	0.83	0.97	0.90
WALKING_DOWNSTAIRS	0.82	0.92	0.87
SITTING	0.92	0.93	0.93
STANDING	0.56	0.98	0.71
LAYING	0.95	0.98	0.96
STAND_TO_SIT	0.54	0.57	0.56
SIT_TO_STAND	0.58	0.76	0.66
SIT_TO_LIE	0.73	0.57	0.64
LIE_TO_SIT	0.48	0.79	0.60
STAND_TO_LIE	0.30	0.70	0.42
LIE_TO_STAND	0.50	0.40	0.44

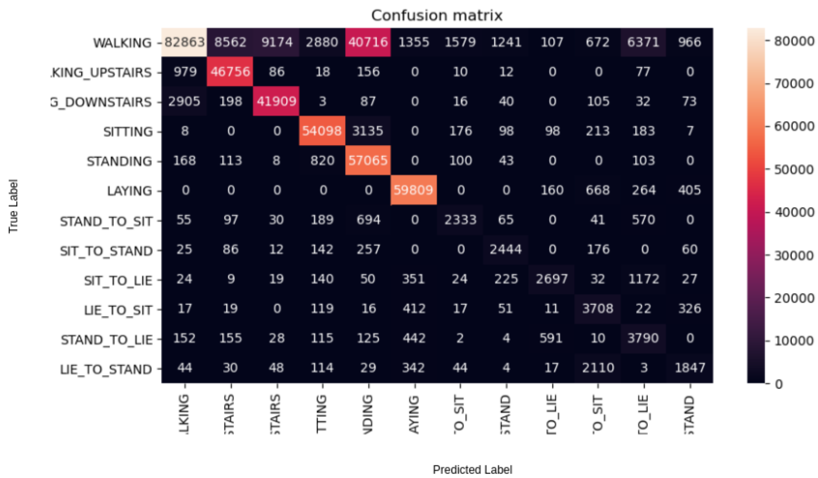


Figure 8: Confusion Matrix for the 12 Activities

6 Conclusion and Future Scope

The deep learning model developed here has helped us in obtaining accurate results for activities such as walking upstairs, sitting, lying . Fewer activities are having a lower F1-score. This is an outcome of the transitions during the change in the activities. These results can be improvised by adding a combination of both Convolutional and RNN. A futuristic scope can be laid upon these activities by improvising the data pre-processing techniques[1]. Also, the entire scope of the project can make headway in the field of HAR for obtaining user results instantaneously, enhancing the health monitoring domain.

References

[1] D. Anguita et al. "A Public Domain Dataset for Human Activity Recognition using Smartphones". In: *ESANN*. 2013.

[2] Girija Chetty, Matthew White, and Farnaz Akther. "Smart Phone Based Data Mining for Human Activity Recognition". In: *Procedia Computer Science* 46 (2015). Proceedings of the International Conference on Information and Communication Technologies, ICICT 2014, 3-5 December 2014 at Bolgatty Palace Island Resort, Kochi, India, pp. 1181–1187. ISSN: 1877-0509. DOI: `https://doi.org/10.1016/j.procs.2015.01.031`. URL: `https://www.sciencedirect.com/science/article/pii/S1877050915000320`.

[3] HAPT dataset. *Smartphone-Based Recognition of Human Activities and Postural Transitions Data Set from University of California, Irving. The dataset can be accessed here*. `http://archive.ics.uci.edu/ml/datasets/Smartphone-Based+Recognition+of+Human+Activities+and+Postural+Transitions#`. [Online; accessed 19-July-2008]. 2008.

[4] Yu Guan and Thomas Ploetz. "Ensembles of Deep LSTM Learners for Activity Recognition using Wearables". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 1 (Mar. 2017). DOI: `10.1145/3090076`.