HUMAN ACTIVITY RECOGNITION

Universität Stuttgart

Banupriya Valluvan and Hanjie Li

Institute of Signal Processing and System Theory, University of Stuttgart, Germany

1. Introduction

Human Activity Recognition (HAR) is a topic in ubiquitous computing that is currently being researched. The basic purpose of a HAR system is to properly analyze human actions by watching and understanding ongoing occurrences. There are multiple applications based on the HAR systems, for example in healthcare monitioring and human-robot interaction (HRI).

2. Dataset and Features

- The HAPT dataset contains data from the tri-axial accelerometer and gyroscope of a smartphone, both captured at a frequency of 50 Hz. The dataset has to be split further into train (70%), validation (10%), and test (20%) datasets based on the subjects.
- There are 12 activities labelled in the dataset: 3 dynamic activities including walking, walking downstairs and walking upstairs, 3 static activities including standing, sitting and lying, 6 transition activities including stand-to-sit, sit-to-stand, sit-to-lie, lie-to-sit, stand-to-lie and lie-to-stand.

3. Data Preprocessing

- The first step of the data preprocessing is sorting out the unlabelled data.
- To reduce the offset and variances in the input data, 6 input channels should be normalized separately. Furthermore, each experiments was performed in different time length. - To make the input data in fixed-length sequences, sliding window technique is deployed. The train dataset is splitted into a series of windows of 250 samples with 50% overlap. The validation and test dataset have the windows of 250 samples without overlap.
- To load the data more efficiently, the TFRecord format is used. The preprocessed input data is written into 3 TFRecords for training, validation and test. The orginal data can be extracted, when the datasets are loaded.

4. Model Architecture

The model is trained with LSTM and CNN-LSTM.

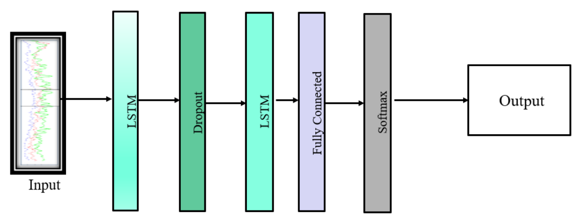

Figure 1: Pure-LSTM Model Architecture

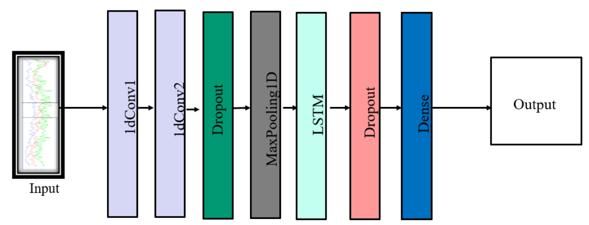

Figure 2: CNN+LSTM Model Architecture

-Dropout layers are used to prevent over-fitting. Here, batch size is 64. The model is trained for a total of 3000 steps. The Adam optimizer is used and Categorical cross-entropy computes the loss.

5. Evaluation and Metrics

To evaluate the performance of the model, the confusion matrix is used as the metric. The results of using different models are compared. The achieved test accuracy is around **91.2%**with pure LSTM model and around **94.2%** with CNN-LSTM model.

Compared to the pure LSTM model, CNN-LSTM shows better performance on the classification of transition activities. The confusion matrix of the CNN-LSTM model is shown in figure 1.

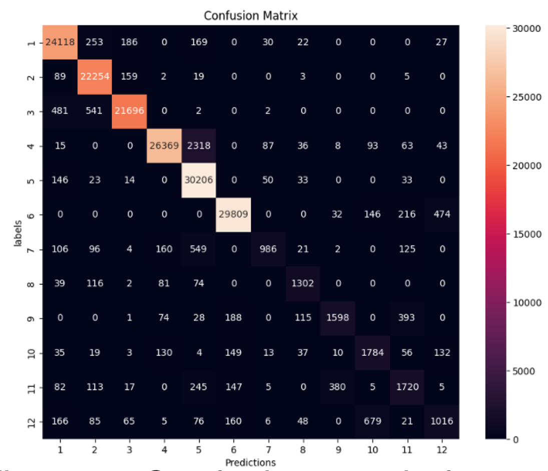

Figure 3: Confusion matrix for evaluation of cnn-lstm model

To visualize the results of the model, a sequence from the test dataset was selected. In figure 2, the accelerometer and gyroscope data in x,y,z axis is shown in first two subplots, the background color represents the corresponding activities in ground truth. The third subplot represents the predicted activities with the background color. The colorbar is shown in the last subplot.

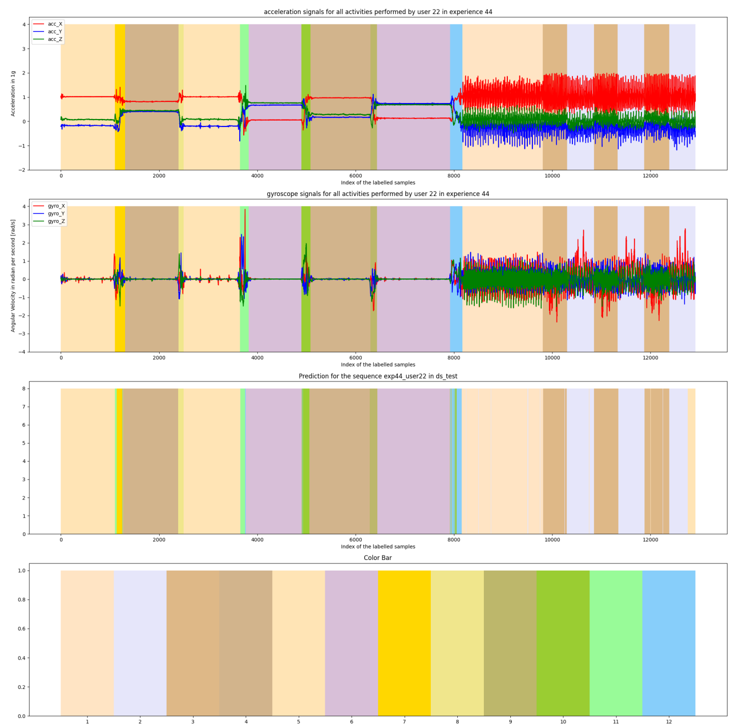

Figure 4: Visualisation of a sequence from test dataset

6. Conclusion

According to our measurements and other previous results, the final conclusion of this study is that, the usage of a CNN-LSTM layer on HAPT dataset with appropriate model parameters can achieve a better performance compared to pure LSTM.

Deep Learning Lab 2022, February 8, Stuttgart, Germany