HUMAN ACTIVITY RECOGNITION

Peter Bauer and Kai Pan

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

1. Introduction

Human activity recognition with body-worn sensors is an active research field that aims to classify sensor data of wearables. In this project we used tri-axial sensor data from an accelerometer and a gyroscope to classify 12 different classes of activities.

2. Data Pipeline

- Sequence-to-label approach with window size 250 and window shift 75.
- Hard coded label assignment.
- Z-score normalizationa and low-pass filtering is applied.
- Unlabeled data is removed before TFRecords creation.
- Dataset gets resampled into two groups (transition and normal activities).

3. Model & Training

- Models are trained for 1500 steps and checkpoints are stored for best validation accuracy to prevent from overfitting.
- LSTM and GRU based models, Temporal Convolutional Network (TCN).
- Flexible architecture to quickly change hyperparameters.
- Adam optimizer with learning rate decay.
- Weighted losses for transition classes (optional).
- Bayesian hyperparameter optimization:
  - number of recurrent layers (LSTM,GRU)
  - size of recurrent layers
  - number of dense units
  - number of dense layers
  - dropout for dense and rnn layers
- Ensemble learning for better generalization.

4. Evaluation

Best results of each model after hyperparameter optimization using *Weights & Biases*.

Model	Acc [%]	Balanced-Acc [%]
LSTM	96.10 (92.78)	90.26 (84.72)
GRU	96.87 (95.20)	92.24 (89.21)
TCN	96.15 (95.17)	92.45 (85.78)
Ensemble	96.46	93.04

Table 1: *Best result in 5 runs of each model, and ensemble result. Values inside brackets denote the average value of the five runs.*

- Metrics
  - Total 12 classes confusion matrix.
  - The class i against all: recall, precision, F1-score.
  - 6 classes confusion matrix for static activities and transition activities.

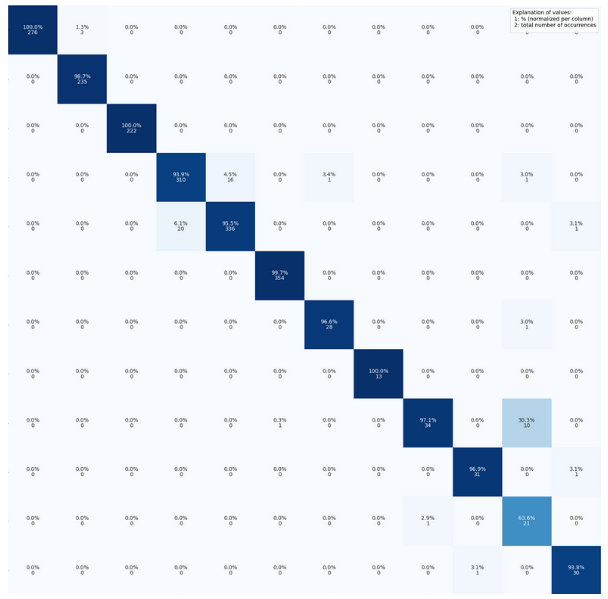

Figure 1: *Total 12 class confusion matrix of our best model (GRU-based)*

- Visual appearance shows a high correspondence between the predictions of our model and the ground truth labels.

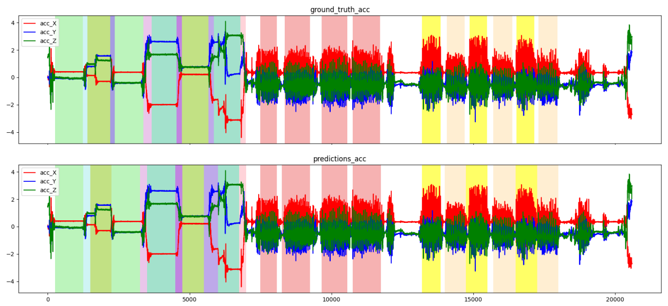

Figure 2: *Visualization of the accelerometer data. Ground truth labels (top), predictions (bottom).*

5. Ablation Study

To increase the performance of our proposed classifiers even more, an ablation study on the loss-weights (for transition activities), kernel initializers, and the cutoff frequency for data preprocessing was performed. Furthermore, we investigated in the relationship between window size and classification performance.

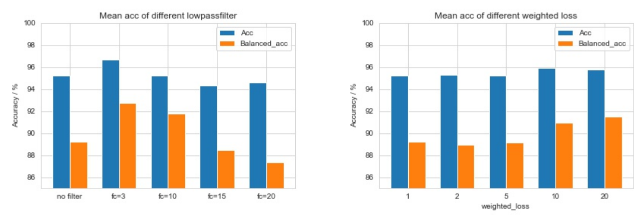

Figure 3: *Ablations of low-pass filter and weighted loss.*

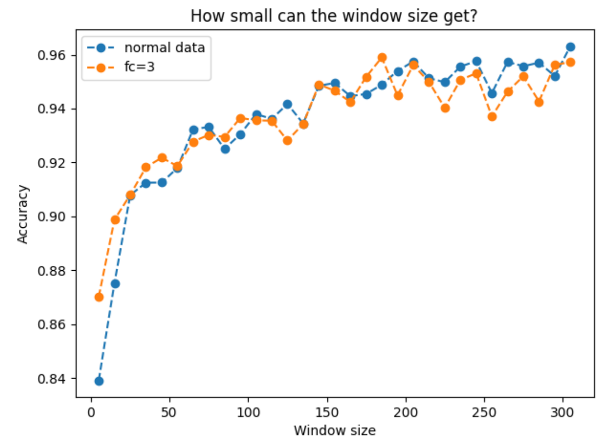

Figure 4: *Influence of the window size.*

6. Conclusion

In total we implemented three different models, all capable of achieving accuracy scores above 90% (five run average). While LSTM-based models and TCN are able to achieve competitive results, the GRU-based architectures were found to be more reliable in terms of reproducibility.

Deep Learning Lab 2022, February 8, Stuttgart, Germany