Smartphone Sensor-Based Human Activitiy Recognition (HAR) using Deep Learning Models

Prabhat Kumar*and S Suresh
Department of Computer Science, Institute of Science,
Banaras Hindu University, Varanasi – 221 005, India.
Email: prabhat.kumar13@bhu.ac.in

Abstract

The deep learning models have been widely accepted for recognizing human activities in recent decades. These models have significantly beaten the classical machine learning techniques in terms of their strong feature extraction capabilities. This is often an expensive and time-consuming process to retrain the model while any changes are made in the HAR (Human Activity Recognition) community. In real life, the activities recognition is a challenging issue because human beings are often performed activities not only simple but also complex and heterogeneity type. We have deep learning models i.e. Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long short-term memory (LSTM) for recognizing the heterogeneity type activities. The activities recognition performance of the purposed model was measured using a publically available KU-HAR dataset. We achieved an F1-score of 98.03%, 98.22%, and 97.96% for CNNs, RNNs, and LSTM, respectively.

Introduction

Human Activity Recognition (HAR) has become an emerging research area in the last decades due to the advancement of wireless sensor technology. The typical application areas of HAR have widely covered the various domains including the activities analysis in the smart home, surveillance application, wellness management, elders caring, gesture recognition, abnormal activities detection, health caring, body temperature and indoor condition monitoring in quarantine due to COVID-19, physical exercise recognition in the gym, patients caring, and more. The sensor-based data gained the attention of researchers because this is low cost, easy implementation, location independence, and injurious free radiation. The accelerometer and gyroscope sensors are mostly used for activity recognition, easily available with digital gadgets such as smartphones and smartwatches.

Contribution

We designed the novel Deep-HAR models i.e. CNNs, RNNs, and LSTM for recognizing human activities. The CNNs model learns and extracts the effective features from raw sensor data using current and temporal activity dependencies. LSTMs are a special kind of RNN, capable of learning long-term dependencies which make RNN smart at remembering things that have happened in the past and finding patterns across time to make its next guesses make sense. The proposed model needs a little bit of the preprocessing the experimental datasets, which makes undoubtedly accepted for implementation of real-time activities recognition system. These experimental datasets are suffering from the class imbalanced problem that becomes defective issues for various classifiers. Hence, our proposed model has robustness against the class imbalanced problem. The experimental outcomes are the recognition of human activities. To accomplish the research work, the experimental dataset is KU-HAR (heterogeneity activities). For evaluating the recognition performance, we have used the F1-score because the activity sample distribution is imbalanced.

KU-HAR Dataset

For heterogeneity-type activities, we have used KU-HAR as the experimental dataset. The heterogeneity activities are those activities that contained unique properties among the group of activities. In other words, the activity classes are different from each other in terms of the associated actions, although some of them are similar, such as walking forward, backward, and in circles. There were a total of 90 subjects, performed the scripted eighteen activities in the experimental environment. The annotated eighteen activities in the KU-HAR dataset, listed as Walk-Circle, Walk-Backward, Table-Tennis, Push-Up, Run, Jump, Stair-Down, Stair-Up, Walk, Sit-Up, Pick, Lay-Stand, Talk-Sit, Lay, Talk-Stand, Stand, Sit, and Stand-Sit. These activities were sampled at 100 Hz on triaxial two smartphone-based accelerometers and gyroscope sensors.

Sensors Deployment in Human Body

In the figure, we illustrate the variation in activity patterns due to the deployment of sensors at different body parts. Here, the small circle filled with blue and red colors shows the known and unknown body labels. The research challenge is that we can caption the missing activity pattern at unknown certain body parts by using known activity patterns at different body parts.

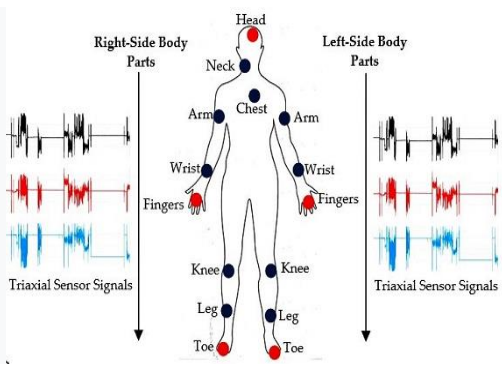

Figure: Sensor deployment points

Activities Data Sample Distribution

The KU-HAR dataset has gained the heterogeneity characteristic that means each class label is distinguished from the remaining classes. The walk circle activity has the lowest samples whereas the stand-sit contained the highest number of samples. However, the remaining walk-backward, table-tennis, push-up, run, jump, stair-down, stair-up, walk, sit-up, pick, lay-stand, talk-sit, lay, talk-stand, stand and sit activities have the equivalent number of samples distribution. Here, the class imbalanced problem has made less influence on the KU-HAR dataset.

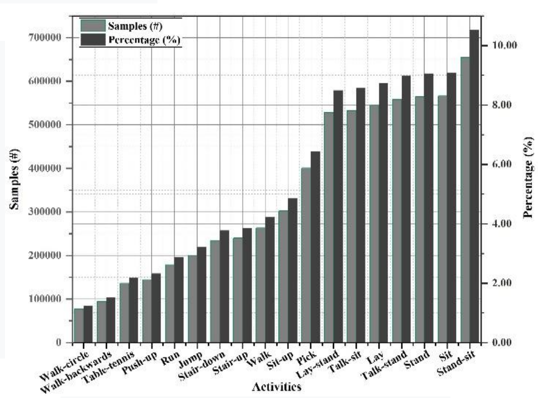

Figure: KU-HAR Dataset Sample

Classical Block Diagram of HAR

The first phase of the HAR is to continuously acquire the data while subjects performing the scripted activities using embedded sensors. Here, we need to apply the data preprocessing for removing the anomalies and outliers. The second phase is segmentation, responsible for slicing the time-series sensor data into equal size of window length. The third phase is feature extraction, based on the time, frequency, and time-frequency domain. However, we should careful while the segmentation and feature extraction phase, because the classification performance is directly influenced by segment length and quality of feature, extracted dataset. In the model training and testing phase, we prefer either machine or deep learning techniques as per our need. Finally, the classification phase predicts the activity class labels on the input streaming of sensor data. However, the researchers have started to prefer deep learning over issues related to machine learning techniques such as essential sensor data preprocessing, lack of unique procedures for feature extraction, and hungry for the huge amount of unlabelled dataset.

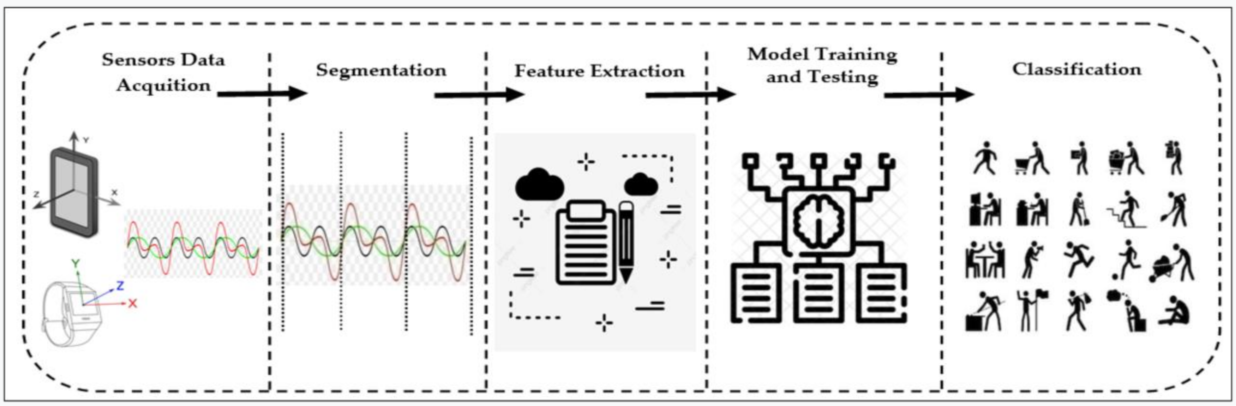

Figure: Human Activities Recognition Process

Deep Learning Models

The convolutional layer is the first layer of a convolutional network. While convolutional layers can be followed by additional convolutional layers or pooling layers, the fully-connected layer is the final layer. With each layer, the CNN increases in its complexity, identifying greater portions of the input data. The convolutional layer is the core building block of CNN, and it is where the majority of computation occurs. Pooling layers, also known as downsampling, conducts dimensionality reduction, reducing the number of parameters in the input. Similar to the convolutional layer, the pooling operation sweeps a filter across the entire input, but the difference is that this filter does not have any weights. However, in the fully-connected layer, each node in the output layer connects directly to a node in the previous layer.

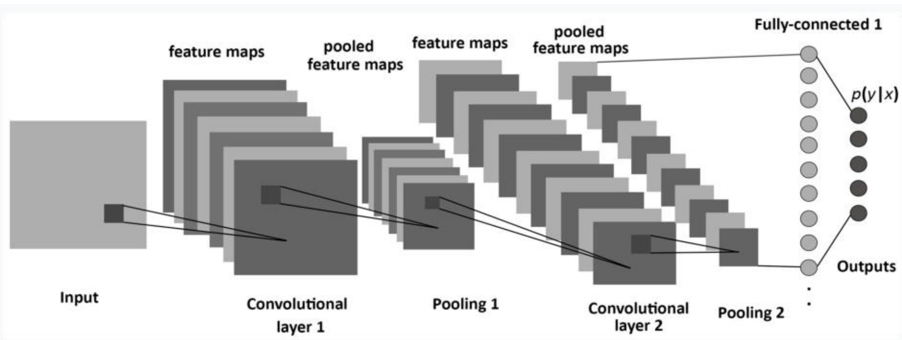

Figure: CNN Architecture

Recurrent Neural Networks suffer from short-term memory. If a sequence is long enough, they'll have a hard time carrying information from earlier time steps to later ones. An LSTM has a similar control flow as a recurrent neural network. It processes data passing on information as it propagates forward. The core concept of LSTM's are the cell state, and its various gates. The cell state act as a transport highway that transfers relative information down the sequence chain. The differences are the operations within the LSTM's cells. These operations are used to allow the LSTM to keep or forget information.

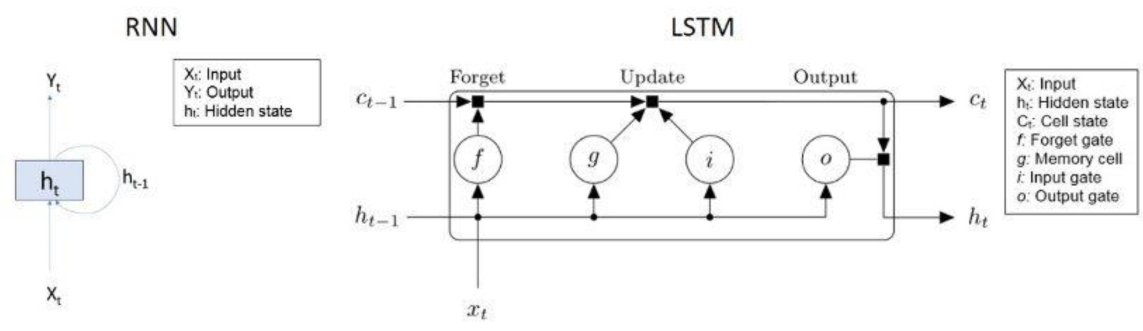

Figure: RNN and LSTM Architecture

Deep Learning Models Configuration

Deep Learning Model	Hyper Parameters	Values
CNN Architecture	**Input Shape**	(4935019 X 7), (1233754 X 4)
	Kernel Size	3
	Filters	512
	Padding	Same
RNN Architecture	**Units**	30
LSTM Architecture	**Units**	30
Model Compilation	**Activation Function**	Softmax
	Loss	Sparse Categorical Crossentropy
	Learning Rate	0.0001
	Optimizer	Adam
	Batch Size	1024
	Epochs	20

Figure: Configuration Table

Confusion Matrix

A confusion matrix is a table that is used to define the performance of a classification algorithm. A confusion matrix visualizes and summarizes the performance of a classification algorithm.

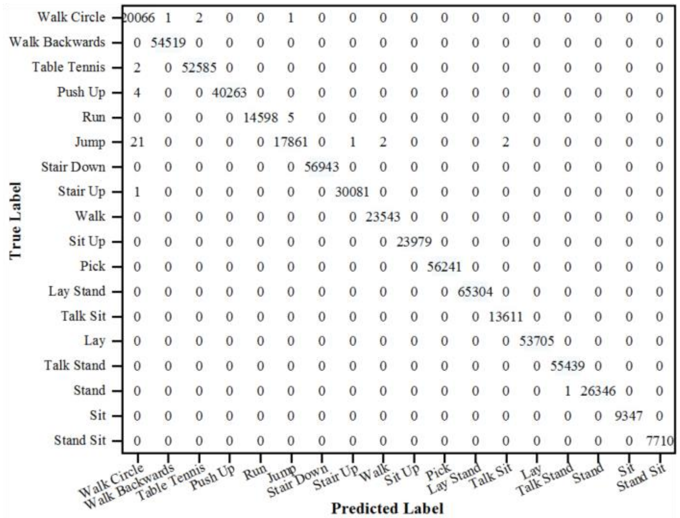

Figure: Confusion Matrix

Accuracy and Loss Score While Training and Validating the Models

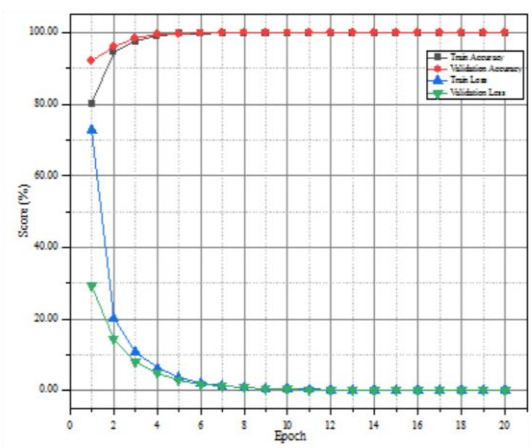

Figure: CNN Model

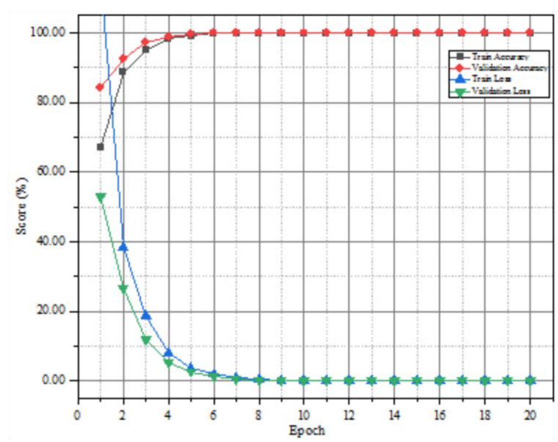

Figure: RNN Model

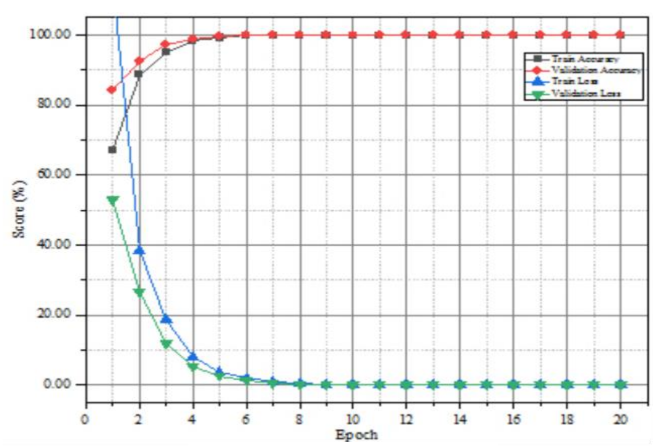

Figure: LSTM Model

Conclusion and Future Work

For evaluating the performance of the Deep Learning models, we have used the accuracy, precision, recall, and F1-Score. Due to the imbalance of activity classes, the F1-score has been used to evaluate the best performing model i.e. RNN Model, which scored 98.22%. Here, we have achieved our best to design and develop a successful way of ensembling the deep learning models. However, each model has certain merits and demerits. The wisely ensembling of the different models and assigning particular responsibilities according to their merits has been accomplished. Moreover, we are still working to improve the performance of the model. Also, the concept of a transfer learning approach for handling more complex activities.

References

- H., Homes, S., Mekruksavanich, S.: LSTM Networks Using Smartphone Data for Sensor-Based Human Activity Recognition in Smart Homes. 1–25 (2021).
- Minh Dang, L., Min, K., Wang, H., Jalil Piran, M., Hee Lee, C., Moon, H.: Sensor-based and vision-based human activity recognition: A comprehensive survey. Pattern Recognit. 108, 107561 (2020).
- Dataset: Sikder, N.,Nahid, A. Al: KU-HAR: An open dataset for heterogeneous human activityrecognition. Pattern Recognit. Lett. 146, 46–54 (2021)