

Long Short Term Memory network for Recognition of Daily Human Activity

Gaurav Arora¹, Anvaya Ahlawat², Mandeep Payal³

{garora039@gmail.com¹, anvayaahlawat12@gmail.com²,
mandeepayal99@gmail.com³}

Maharaja Surajmal Institute of Technology^{1,2,3}

Abstract: Previously Human Activity Recognition has been solved by using the engineered features, but the main problem with this approach is that it requires specific domain knowledge. Adopting classical machine learning models has been effective but these methods completely ignore the time signal obtained from the sensors. But due to recent advancements, deep learning techniques like Long Short Term Model and recurrent neural networks have been effectively used to provide good results and classify human activities correctly as compared to machine learning models. So in this paper, we propose a LSTM framework to classify these activities and also compare its performance with classical machine learning models.

Keywords: deep learning, LSTM, neural networks, machine learning, human activity recognition, convolution

1 Introduction

Mobile phones have become an important part of our life. With continuous advancements in the technology, it has become possible to accommodate more and more sensors in the phones which in turn can be used to track our daily activities. The readings attained from these sensors depends on the movement of the smartphones. This process of classifying various human activities from the data collected by these sensors (mainly accelerometer and gyroscope) and using them to provide useful feedback is known as Human Activity Recognition. It is not a simple problem as there is a large number of observations available and there is no straight way of relating the readings recorded by the sensors. So one way is to directly apply deep learning architecture like RNN and LSTM on the raw sensor data and the second way is to generate handcrafted features from the recorded sensor signals and then apply classical machine learning models on them, but the latter approach is not always possible as it requires good domain knowledge.

Recognition of various human activities finds its applications in many diverse fields. It can be used in daily life monitoring which includes tracking the total number of steps taken, the number of hours slept, calories burnt. It is also used to track the activities of elderly people so as to provide them with better care and assisting them. It is also being used intensively in the medical field[1,2,3]. For example, patients with heart disease are required to follow a particular diet and to exercise regularly, so it can help the patients track their activities and provide valuable feedback. Similarly, it can be used to detect abnormal activities in patients with dementia, Alzheimer [2] and other mental illness. It is also used in sports [4] to track the fitness of the players and to suggest them a better workout plan based on their various daily activities.

2 Related work

With the advancements in sensing technology and the Internet of Things, it has become very easy to track various human activities like sitting, walking, standing without any use of costly hardware.

2.1 Dynamic Sliding Window Approach

The main aim of Human activity Recognition is to identify the actions of each person from a set of recorded observations. To perform this a fixed-length window sliding approach is opted, where the two parameters: the window and the shift have to be fixed. The authors in [5] proposed a different approach using dynamic windows based on events. Their approach adjusts dynamically the window size and the shift at every step.

2.2 Convolutional networks and Extreme Learning Machine classifier

Due to no requirement of domain knowledge in feature extraction, the authors in [6] proposed a hybrid deep framework based on convolution operations, LSTM recurrent units, and ELM classifier.

3 Methodology

3.1 Dataset

The dataset used in this paper is collected from the UCI Machine Learning repository [7]. The data was collected with the help of an experiment conducted on thirty volunteers of the age group 19-48 years. The persons were required to perform six activities walking, walking downstairs, walking upstairs, standing, sitting and laying. A smartphone was attached to their waists which contained gyroscope and accelerometer to record three-axial angular velocity and linear acceleration. Each

reading was divided into a window of 2.56 seconds with 50% overlapping. Some feature engineering was performed like taking Fourier transforms, max, mean, entropy. We finally get a feature vector of 561 features. Then the obtained dataset was divided into two sets 70% training and 30% test randomly.

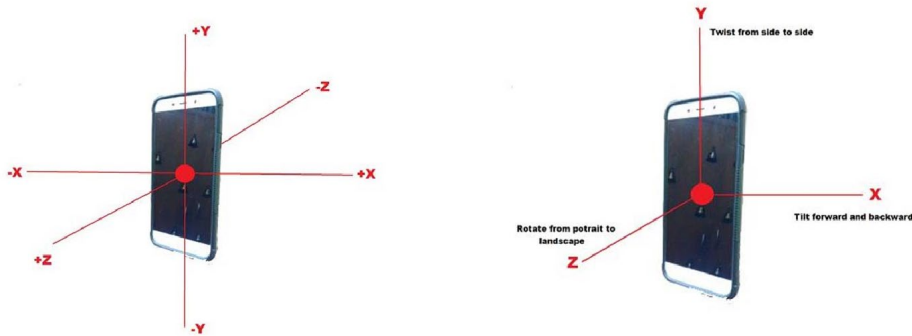


Fig 1: Accelerometer and Gyroscope

3.2 Exploratory Data Analysis and Data Preprocessing

One of the common problem that affects the performance of the machine learning models is the existence of imbalanced dataset which leads to overfitting. From the Fig 2 it can be observed that our data is not imbalanced and each class has sufficient amount of data points for each activity.

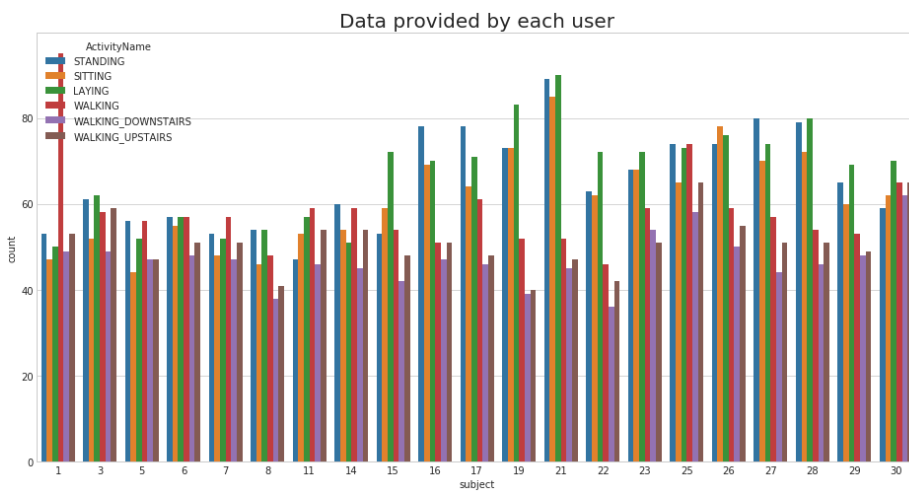


Fig 2: Count of the data provided by each user for each activity

We did a comparisons of the histograms with x axis as tBodyAccMagmean (Fig 3) of the stationary activities and the moving activities and find out that both the group of activities are completely different from each other.

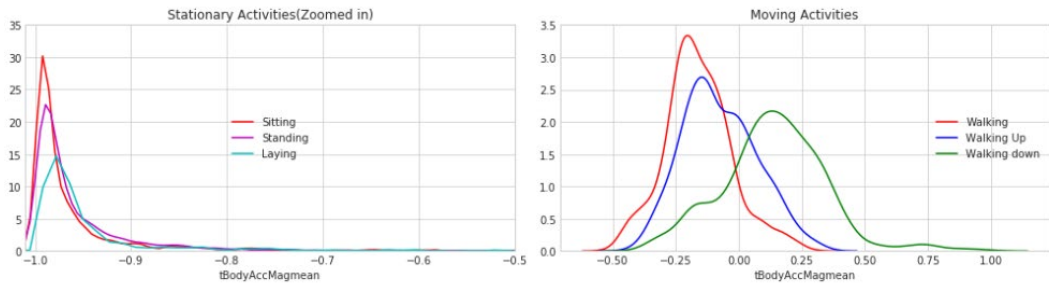


Fig 3: Histogram of Body acceleration mean

We then performed t-SNE on the dataset with different values of perplexities. It is a popular method of exploring high-dimensional data. It is a very useful dimensionality reduction technique. It creates two-dimensional maps from high dimensional data. From the Fig 4 we can see that the activities laying, walking, walking_downstairs, walking_upstairs are pretty much separated and can easily be classified but in case of standing and sitting activities, the clusters overlap and will be difficult to separate.

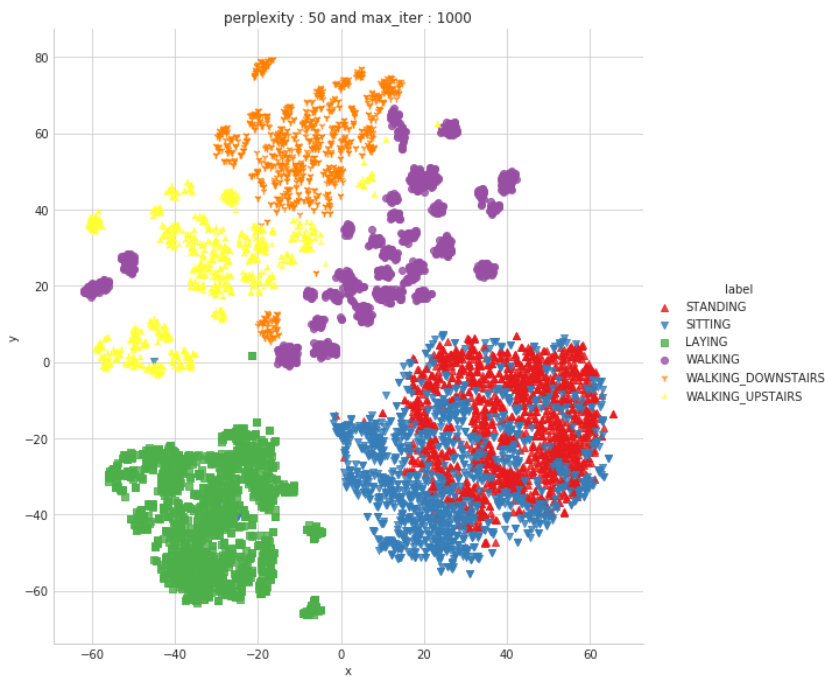


Fig 4: T-distributed Stochastic Neighbor Embedding (t-SNE)

4. Results and Discussion

Our model contains 2 LSTM layer and 1 dropout layer with dropout percentage 50% to reduce the overfitting of the model as seen in Fig 5. We limit our model to only 2 LSTM layers because if we increase the number of these layers then the number of

total trainable parameters will become more than the total data points in the dataset and will lead to overfitting. We are using sigmoid activation function and RMSprop

| Layer (type) | Output Shape | Param # |
|-------------------------|--------------|---------|
| lstm_3 (LSTM) | (None, 32) | 5376 |
| dropout_3 (Dropout) | (None, 32) | 0 |
| dense_3 (Dense) | (None, 6) | 198 |
| Total params: 5,574 | | |
| Trainable params: 5,574 | | |
| Non-trainable params: 0 | | |

as the optimizer and calculated loss as categorical cross-entropy.

Fig 5: Long Short Term Memory based model architecture

After compiling and fitting of the model, we now tested our model on the dataset. With a simple 2 layer LSTM network, we get an accuracy of 92.14% and a loss of 0.30. In the below confusion matrix(Fig 6), we can see that our model is doing a good job in classifying laying, walking, walking downstairs and walking upstairs activities, but it still has a problem differentiating between sitting and standing activities. This can be further improved.

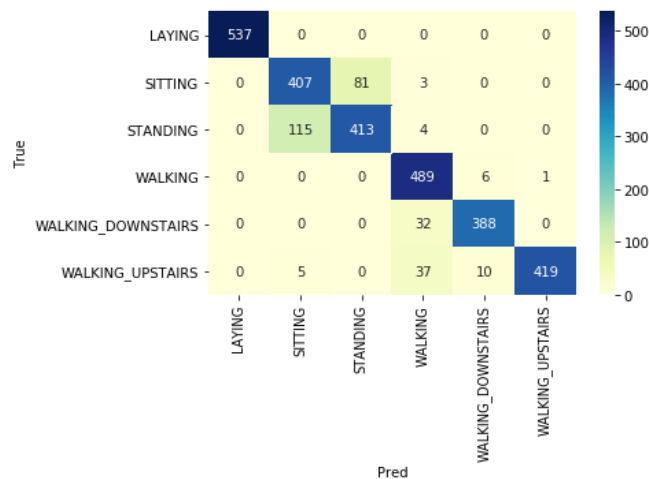


Fig 6: Confusion matrix

The Table 1 shows the comparison of accuracy of different machine learning models and the LSTM model. The LSTM model performs better than most of the machine learning models without the requirement of any prior domain knowledge and no need of feature engineering. This showcase the advantages of using a deep framework based on the Long Short Term Memory recurrent layers, for activity recognition. This structure is used to study multichannel time series data. It captures the significant features of sensor signals using the LSTM cells and various convolution operations.

Table 1: Comparison table

| Model | Accuracy | Error |
|--------------------------------|-----------------|--------------|
| Logistic Regression | 96.27% | 3.733% |
| Linear SVC | 96.61% | 3.393% |
| rbf SVM classifier | 96.27% | 3.733% |
| Decision tree | 86.43% | 13.57% |
| Random Forest | 91.31% | 8.687% |
| Gradient Boosted Decision Tree | 91.31% | 8.687% |
| LSTM | 92.14% | 7.86% |

5. Conclusion

In this paper we explored the Long Short Term Memory based deep network to solve the Human activity recognition problem and compared its performance with the classical machine learning approaches. LSTM does a good job in classifying activities . It even gives better accuracy than ensemble models like decision trees and random forests. It has wider application in medical research and study of human framework.

6. Future work

More advanced deep learning models could be applied like a model which uses both CNN and LSTM. The right amount of hyperparameter tuning of the batch size, epochs and the number of layers can give better results. Bidirectional LSTMs and stacked LSTMs could also be used to make the performance better. The most important part of this problem was to deal with time-series data and to choose the model which best fits the data and give satisfactory results. A large amount of null and missing values can degrade the performance of your model, so dealing with them is very important. The LSTMs can easily overfit, so choosing the right value of

hyperparameters – batch size, epochs and dropout percentage play a major role in the performance of the model. Apply this approach on a more diverse and bigger dataset so that a more deep LSTM network could be used to give the results.

References

- [1] S. Sendra, L. Parra, J. Lloret, and J. Tomás, “Smart system for children’s chronic illness monitoring,” *Information Fusion*, vol. 40, pp. 76–86, 2018.
- [2] P. C. Roy, S. Giroux, B. Bouchard, A. Bouzouane, C. PhuaAndrei, and J. B. TolstikovL. Chen, C. Nugent, J. Biswas, and J. Hoey, “A possibilistic approach for activity recognition in smart homes for cognitive assistance to Alzheimer’s patients,” in *Activity Recognition in Pervasive Intelligent Environments, Atlantis Ambient and Pervasive Intelligence*, pp. 33–58, Atlantis Press, 2011.
- [3] C. Xu, J. He, X. Zhang, C. Wang, and S. Duan, “Detection of freezing of gait using template-matching-based approaches,” *Journal of Sensors*, vol. 2017, Article ID 1260734, 8 pages, 2017.
- [4] J. Margarito, R. Helaoui, A. M. Bianchi, F. Sartor, and A. Bonomi, "User independent recognition of sports activities from a single wrist-worn accelerometer: a template-matching-based approach," *IEEE transactions on bio-medical engineering*, vol. 63, no. 4, pp. 1–796, 2015.
- [5] Ortiz Laguna J., Olaya A.G., Borrajo D. (2011) A Dynamic Sliding Window Approach for Activity Recognition. In: Konstan J.A., Conejo R., Marzo J.L., Oliver N. (eds) *User Modeling, Adaption and Personalization. UMAP 2011. Lecture Notes in Computer Science*, vol 6787. Springer, Berlin, Heidelberg.
- [6] Jian Sun, Yongling Fu, Shengguang Li, Jie He, Cheng Xu, and Lin Tan, “Sequential Human Activity Recognition Based on Deep Convolutional Network and Extreme Learning Machine Using Wearable Sensors,” *Journal of Sensors*, vol. 2018, Article ID 8580959, 10 pages, 2018.
- [7] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra and Jorge L. Reyes-Ortiz. A Public Domain Dataset for Human Activity Recognition Using Smartphones. 21th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN 2013. Bruges, Belgium 24-26 April 2013.