Assignment 3: Evaluating Performance of Machine Learning Algorithms on Human Activity Recognition

RISHIRAJ ADHIKARY, Indian Institute of Technology, Gandhinagar, India

Motion sensor data is instrumental in classifying human activity. This assignment analysed the raw time-series data and temporal and frequency domain features curated over time series in classifying six human activities. While the classical machine learning approach performs better with curated features, sequence models perform better only when both accelerometer and gyroscope data are available. I observed that removing gyroscope data from the list of curated features reduced classification accuracy by 2%. I conclude that activity recognition can be performed using accelerometer data alone but requires hand-crafted feature selection.

Additional Key Words and Phrases: activity recognition, DTW, KNN, SVM, LSTM

ACM Reference Format:
Rishiraj Adhikary. 2021. Assignment 3: Evaluating Performance of Machine Learning Algorithms on Human Activity Recognition. In . ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Human activity recognition is a classic problem of identifying events performed by humans. In this assignment, I will use a well known Human Activity Recognition (HAR) [1] dataset and perform time series classification to classify six activities described in the next section. The HAR dataset is a well-curated feature on the time series data. But, to maintain the rigour of the analysis in this assignment, I have used both the raw time series data and the curated features as an input to three machine learning models. The following section describes the dataset, followed by the approach and result.

2 DATASET

The Human Activity Dataset (HAR) [1] comprises of 3-axial linear accelerometer and gyroscope data on human activity, viz. walking, walking upstairs, walking downstairs, sitting, standing and lying. Researchers collected the data using a Samsung smartphone at a constant rate of 50Hz. The experiments were carried out with 30 volunteers within an age bracket of 19-48 years. As mentioned in the UCI Machine Learning repository[1], “The sensor signals (accelerometer and gyroscope) were pre-processed by applying noise filters and then sampled in fixed-width sliding windows of 2.56 sec and 50% overlap (128 readings/window).”

The data repository contains two kinds of data. First is the raw values of the accelerometer and gyroscope. Second, the authors of HAR used the accelerometer and gyroscope data to create temporal and frequency domain features. The featured data is a 561 length vector each for each activity. There are 1722 walking activities, 1544 walking upstairs activities, 1406 walking downstairs activities, 1777 sitting activities, 1906 standing activities and 1944

[1] https://archive.ics.uci.edu/ml/datasets/human+activity+recognition+using+smartphones

Ubicomp Class '21, Semester 1, 2021, Gandhinagar, India

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM. . . $15.00
https://doi.org/10.1145/1122445.1122456

lying activities. There are a total of 10299 activities.

Although the authors provided the training and testing data split, I chose to combine the entire data and create a balanced class dataset of 3000 activities where each activity appears 500 times.

3 APPROACH

This assignment aims to evaluate the performance of different machine learning models (ML) on the HAR dataset. Multi-sensor sensing is often challenging in a realistic sensing environment due to lack of time synchronisation, more sensing space and higher energy needs. Motivated by the challenges of multi-sensor sensing, I also evaluated the performance of ML models on a trimmed version of the dataset. I used two versions of the dataset for the machine learning model that follows. The different datasets are explained below.

(1) **(D1) Only Accelerometer Data:** In all the machine learning models analysed, I initially used only the 3-axis accelerometer raw time series data and abstained from using the gyroscope data.
(2) **(D2) Featurized data:** I also evaluated all the models (except sequence model) on the full data. Full data refers to the 561-feature vector with time and frequency domain variables created from the raw data of the accelerometer and gyroscope.

.

Furthermore, for D2, I tried to answer two questions.

- (**Q1:**) Which class predictions remain accurate even after removing Gyroscope data?
- (**Q2:**) What will be the effect on accuracy if frequency domain information is removed?

I chose to answer the above question only for one of the machine learning classifiers. The machine learning approach for classification follows:

Baseline: To obtain the baseline performance, I used the most frequently occurring class to generate predictions. **K Nearest Neighbor (KNN) with Dynamic Time Warping (DTW), SVM and LSTM**: K Nearest Neighbors is a classification algorithm that takes an unlabeled observation and compares it to a population of labelled observations. By changing the distance metric of KNN to DTW, we can apply KNN to time-series data. The objective was to obtain the prediction accuracy for both D1 and D2 datasets. To answer Q1 and Q2 above, I used a support vector machine (SVM). Finally, I also fed the raw accelerometer data into a one layer Long Short term Memory (LSTM) recurrent neural network (RNN). For each of the models, I used 33% of the data as a test set. I chose the best choice of hyperparameter for KNN with DTW (number of neighbours) and the best selection of hyperparameter for SVM (C, gamma and kernel) using a grid search cross-validation approach with four-folds.

4 EVALUATION AND RESULT

I observed that with D1 (Raw accelerometer) data, SVM with RBF kernel had an accuracy of 88.2% and surpassed KNN with DTW by 1%. With D2 (curated features) data, KNN with DTW had an accuracy of 98.5%. With 98.1%, SVM with RBF kernel had the second-best performance. Table 1 shows the overall result. Next, I evaluated the question, "Which class predictions remain accurate even after removing Gyroscope data?". On removing all the features retrieved from gyroscope data, the overall accuracy dropped by around 2%. It shows that accelerometer data alone can be used for activity recognition without a significant reduction in accuracy. On the second question, "What will be the effect on accuracy if frequency domain information is removed", the observation was similar, i.e the overall accuracy dropped only by 2%. It is worth mentioning that the lowest F1 score of 0.91 was rcorded for the standing activity among all the activities.

No.	Algorithm	Data	Test Accuracy
0	Baseline (Most Frequent)	Raw Accelerometer and Curated Feautures	14.5%
1	DTW with KNN	Raw Accelerometer Data	87.2%
2	DTW with KNN	Curated Features	98.5%
3	SVM (RBF)	Raw Acceleromter Data	88.2%
4	SVM (RBF)	Curated Features	98.1%
5	SVM (Linear)	Curated Features Without Gyroscope Data	96.9%
6	SVM (Linear)	Curated Features Without Frequency Domain Information	96.9%
7	LSTM (Single Layer)	Raw Accelerometer Data and Gyroscope Data	88.6%
8	LSTM (Single Layer)	Raw Accelerometer Data	18.2%
9	LSTM (Single Layer)	Raw Gyroscope Data	34.0%

Table 1. SVM with RBF kernel performs the best with all the features. SVM (RBF) maintains its best performance even if only curated features on accelerometer data is considered.

REFERENCES

[1] Davide Anguita, Alessandro Ghio, Luca Oneto, Xavier Parra, Jorge Luis Reyes-Ortiz, et al. 2013. A public domain dataset for human activity recognition using smartphones.. In *Esann*, Vol. 3. 3.