Ensemble Approach for Sensor-Based Human Activity Recognition

Sunidhi Brajesh
New York University
Brooklyn, New York, U.S.A
sb7253@nyu.edu

Indraneel Ray
New York University
Brooklyn, New York, U.S.A
ir944@nyu.edu

ABSTRACT

This paper discusses in detail our (Team:AISA) ensemble based approach to detect Human Activity for the Sussex-Huawei Locomotion-Transportation (SHL) recognition challenge. The SHL recognition challenge is an open competition wherein the participants are tasked with recognizing 8 different types of activities based on smartphone data collected from multiple positions - Hand, Hips, Torso, Bag. On the magnitude of sensor data, time and frequency domain features were calculated to achieve position independence. To make the model robust, we trained it with a random shuffle of the training and validation data provided. To find the optimal hyper-parameters, we parallely executed randomized search to choose the best performing model from about 200 models. We set aside 30% of this combined dataset for internal testing and the model predicted human activities with an F1-Score of 86% on this test dataset.

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing**; • **Computing methodologies** → **Ensemble methods**; *Cross-validation.*

KEYWORDS

SHL Recognition Challenge 2020; Activity Recognition; Random Forest; RandomizedSearchCV; Parallel Computation

ACM Reference Format:
Sunidhi Brajesh and Indraneel Ray. 2020. Ensemble Approach for Sensor-Based Human Activity Recognition. In *Adjunct Proceedings of the 2020 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2020 ACM International Symposium on Wearable Computers (UbiComp/ISWC '20 Adjunct), September 12–16, 2020, Virtual Event, Mexico.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3410530.3414352

UbiComp/ISWC '20 Adjunct, September 12–16, 2020, Virtual Event, Mexico

ACM ISBN 978-1-4503-8076-8/20/09…$15.00
https://doi.org/10.1145/3410530.3414352

1 INTRODUCTION

As smartphone and wearable device adoption becomes prevalent, the amount of sensor data collected from these devices has also increased exponentially; giving researchers the opportunity to extract meaningful insights. Few applications of these are, health monitoring and diagnosis [7], parking spot detection, traffic route monitoring [9], assistive technology, and elder-care, indoor localization and navigation, etc. Activity recognition is a core building block behind many of these[10]. It takes input in the form of raw sensor readings and predicts a user's motion activity.

The Sussex-Huawei Locomotion - Transportation (SHL) dataset [13] [5] with eight modes of locomotion, three users and four smartphone positions is one of the most comprehensive datasets for modeling activity recognition from mobile sensor data. The Sussex-Huawei Locomotion-Transportation (SHL) recognition challenge, 2020 is intended to recognize the eight modes of activities i.e. walk, run, still, subway, bike, car, bus, train, and subway from inertial sensor data by keeping a smartphone in different positions. The goal of this challenge is to recognize eight modes of locomotion and transportation (activities) independent of the position of the smartphone.

The key difference this year is the additional hand data provided in the training set along with last year's training data for the rest of the phone positions. Last year the participants were supposed to predict activity on the hand data provided in the test. This year the test data on which the predictions must be made is given from an unknown location and in a time shuffled manner to the participants.

We (Team:AISA) approach the challenge by first preprocessing the sensor data and then extracting a number of important features and finding the best hyper-parameters by using randomized grid search with parallel execution. We finally proceed to train a hyper-parameter tuned random forest classification model. Random forests [4][14] are computationally efficient to train and test, making them excellent candidates for real-world prediction tasks.

The paper is organized as follows: In Section 2, we describe the dataset and implement exploratory data analysis techniques on the dataset provided for this year's challenge. In section 3, we discuss about the data preprocessing and feature engineering in our pipeline. In section 4, we discuss the model used and techniques for improving the performance of the model. We conducted experiments before proposing the pipeline which is discussed in section 5. At last, we conclude by reporting the findings and scope of the future work.

2 SHL RECOGNITION CHALLENGE DATA

The SHL Recognition Challenge 2020 focussed on identifying 8 modes of transportation - Still, Walk, Run, Bike, Car, Bus, Train, Subway using inertial sensor data of a smartphone. The dataset was acquired from three users wearing four smartphones at positions - Hips, Bag, Hand and Torso simultaneously. It contained labeled data logged from 16 sensor modalities over 2812 hours.

The Dataset was divided into 3 parts - *SHL-Training*, *SHL-Validation* and *SHL-Test* dataset. The *SHL-Training* dataset contained labeled data over 59 days, *SHL-Validation* data over 6 days and *SHL-Test* data over 40 days. *SHL-Training* data was the largest, containing raw sensor data at four positions from User 1. The *SHL-Validation* contained data for User 2 and User 3 at four positions while, *SHL-Test* contained data for User 2 and User 3 at a position, which was kept unknown.

The raw sensor data was sampled at frequency of 100Hz and included data from following sensors: acceleration (x, y and z), linear acceleration (x, y and z), magnetic field (x, y and z), gravity (x, y and z), gyroscope (x, y and z), Orientation (x, y, z and w) and Pressure. The data was segmented with a non-overlapping sliding window of 5 seconds and labels were provided per sample. The frames in *SHL-Training* and *SHL-Validation* data were consecutive while, in *SHL-Test*, the frames were shuffled.

3 PREPROCESSING AND FEATURES

Data Preprocessing

We observed that the SHL-Training and SHL-Validation data consisted of missing values in several samples. In these cases, for each sample of 5 seconds, we replaced the missing values with mean of the values in the frame.

Feature Engineering

Since our approach was to classify activities using classical machine learning, features had to be calculated for respective window of data. To overcome the impact of position and orientation, we calculate the magnitude for each of the sensor.

$$magnitude = \sqrt{x^2 + y^2 + z^2}$$

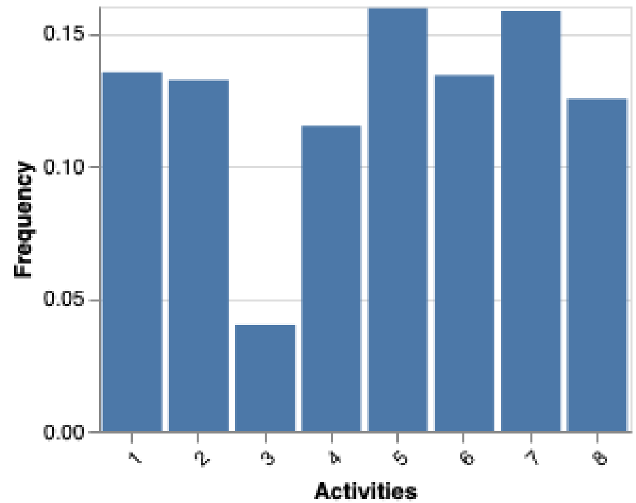

Figure 1: Combined distribution of activities across SHL-Training and SHL-Validation dataset:Still-1, Walk-2, Run-3, Bike-4, Car-5, Bus-6, Train-7, Subway-8

Based on the assumption that user is not likely to change activity during five second window, we selected window size as five-second for feature extraction. Also, *SHL-Test* data contained samples with window-size as five seconds, limiting the maximum size of window to five-second.

Statistical features were extracted [1] [11] from acceleration, linear acceleration, magnetic field, gyroscope and pressure magnitude for every five-second window. The statistical features can be categorized into time-domain and frequency-domain as described below:

- `Time-Domain`: For each window of 5 second, we calculate the mean, median, maximum, minimum, standard deviation, variance, interquartile range on the magnitude value of the sensor. Mean, median and mode measured the central tendency of data while standard deviation, variance and interquartile range described the spread of the data for every window.
- `Frequency-Domain`: Each window of 5 second data was transformed into frequency domain using fast fourier transform (FFT) to analyze spectral features. We extracted mean frequency, peak frequency, kurtosis, skewness, energy and entropy for each segment from respective sensors. Kurtosis and skewness quantify the peak and asymmetry in probability density function of the signal respectively. The Energy Spectral Density describes how the energy of a signal or a time series is distributed with frequency [6]. Entropy was calculated to discriminate between similar energy features.

Overall 65 features were calculated for model training. For each window of five-second, corresponding label was determined using mode of the labels provided in *SHL-Training* and *SHL-Validation* dataset.

Data Preparation

The *SHL-Training* and *SHL-Validation* consisted of frames consecutive in time at all four positions. The *SHL-Test* data contained shuffled frames for an unknown position. To develop a model that generalizes across all four positions and three users, features extracted for both *SHL-Training* and *SHL-Validation* were combined together in single training dataset. The frames in this combined dataset were further shuffled and split into *Training* dataset, 70% and *Test* dataset 30%.

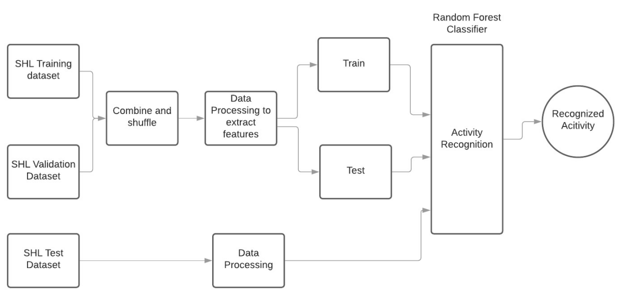

Figure 2: Structure of Activity Recognition System

4 METHODOLOGY

The goal of this year's challenge was to develop a user-independent pipeline to recognize eight modes of locomotion. A Decision Tree is a non-parametric model implemented based on divide and conquer strategy. It is mainly used as base classifier in many activity recognitions researches [11] as it is easy to build and interpret. Since, the location was unknown in *SHL-Test* data, we needed to develop model that generalizes across all locations given in *SHL-Training* and *SHL-Validation* dataset. To improve accuracy and prevent overfitting, we opted for ensemble of decision trees, Random Forest. In Random Forest classifier, multiple decision trees were built from different sample drawn with replacement from the *Training* dataset. To reduce co-relation between trees, at each split in a tree, random sample of m features were selected from full set. The output from each decision tree was combined by averaging the probabilistic prediction instead of letting each classifier vote for a single class. This led to improved accuracy in prediction on *Test* dataset. We used scikit-learn's Random Forest. Since, we were using mode to extract label for each frame in *Training* dataset, the predicted activity was same for each observation in a five-second frame. We achieved F1-Score of 86% on *Test* dataset. Figure 2 illustrates the structure of our activity recognition system.

Hyperparameter Tuning

In most cases Random Forest works reasonably well with the default values of the hyperparameters specified in scikit-learn. Out of SVM, KNN, XGBoost and Random Forest, we found Random Forest to have the best combination of accuracy and training time, making it suitable for our analysis. Tuning the parameters helped us achieve an increase in accuracy in comparison to the default values.

Probst et al. provide some guidance on how to tune parameters effectively [8]. The number of trees should be set high: the higher the number of trees, the better the results in terms of performance and precision of variable importances. However, the improvement obtained by adding trees diminishes as more and more trees are added. Apart from this, we also tuned maximum tree depth to achieve a balance between model flexibility and avoiding overfitting.

Since there are many other parameters that need to be assessed, it would make sense to automate the optimal hyper-parameter search among a combination of many such parameters. Sklearn provides options for both exhaustive grid search and randomized grid search. Bengstra et al. [3] have shown that random experiments are more efficient than grid experiments for hyper-parameter optimization in the case of several learning algorithms on several data sets. It is also computationally more efficient to perform randomized hyper-parameter search. To make the model more robust, we chose a 10 fold cross validation with random shuffle on the training data while doing randomized search with 20 iterations; giving us the best among 200 models.

We finally arrived at the following optimal hyper-parameter values after performing randomized search:

- `Criterion`: 'Gini'
- `max_depth`: 46
- `n_estimators`: 125

The oob_score with default parameter values was 80.64% and after parameter tuning, the oob_score improved to 85.12%. The mean cross validation accuracy for this parameter with 10 fold cross validation was found to be 84.96%. The F1-score on *Test* dataset was 86% and accuracy was 85.57%. Parameter hyper-tuning resulted in reduced model size without any degradation in performance. The final trained model size is 278.3 MB.

Computation Methods

We trained the random forest model parallelly using all logical cores to reduce the training time. We had the following hardware and software packages available for our testing.

- 2.3 GHz CPU i9 processor with 16 virtual cores and 16 GB RAM.
- Language: Python 3.7.5
- Library: Scikit learn 0.22.1

In scikit-learn implementation, n_jobs=-1 was passed as parameter during model training which increased the CPU utilization to 90%. The training time for the model was 5 minutes and 3 seconds and parameter hyper-tuning time was 20hr 30mins. It took 1 second to predict labels for *SHL-Test* dataset.

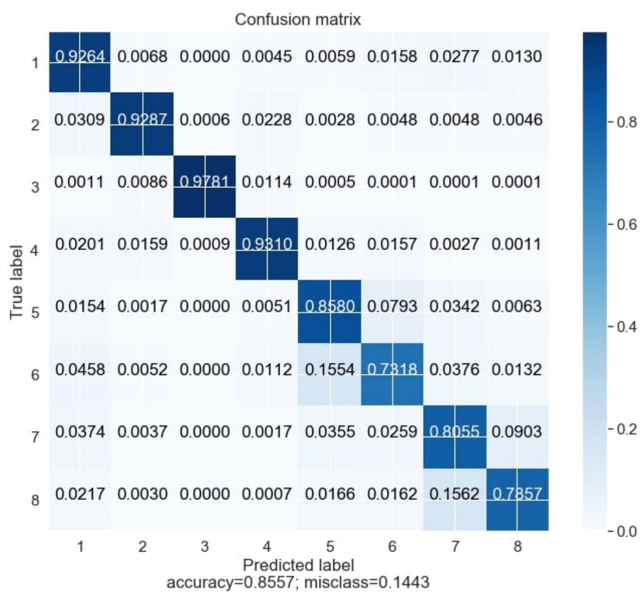

Figure 3: Normalized confusion matrix on test data : Still-1, Walk-2, Run-3, Bike-4, Car-5, Bus-6, Train-7, Subway-8

5 EXPERIMENTS

For the prediction, we utilized data from all positions of both *SHL-Training* and *SHL-Validation* dataset. We calculated the magnitude of the sensor streams and extracted stastical features from it. We then combined the features of *SHL-Training* and *SHL-Validation* dataset according to the position. We split this data in 70:30 ratio for training and testing. Initially, we evaluated our model with default parameters on the training dataset using four fold cross validation. The idea was to train the model on three locations and validate on the fourth location. The order of frames were preserved in the process. The performance was not satisfactory as the oob_score was found to be 74%.

Another approach was to build an ensemble of four random forest models for each body position. Each model was fed with the statistical features which we had calculated. The final position was determined by taking a vote from each model. This approach also did not yield satisfactory results and the maximum accuracy score we achieved was 71%. The model failed to classify subway, train and bus probably because of the similarities in the nature of these activities.

To remove dependency of position and order of activities, after combining *SHL-Training* and *SHL-Validation* dataset, we shuffled the rows. This data was then split into 70:30 ratio. The model was evaluated with 10-fold cross validation and observed an increase in oob_score to 80% with same parameters. Parameter tuning was done for this model using RandomizedSearchCV with 10-fold cross validation and 10 iterations. The model parameters that maximized the F1-score on held-out data were returned as best the parameters. The final model was trained using the best parameters returned. The oob_score on the training dataset was then 85.12% and F1-score of this model on Test dataset was 86%. Figure 3 provides confusion matrix for each activity for this model.

6 CONCLUSION

In this paper, we have presented ensemble approach for activity recognition using scikit-learn's Random Forest. The model was evaluated using shuffled 10-fold cross validation. The final model was trained with parameters obtained by implementing RandomizedSearchCV with ShuffleSplit 10-fold cross validation in 10 iterations. The F1-Score for the final model on *Test* dataset is: Still-88%, Walk-94%, Run-99%, Bike-94%, Car-84%, Bus-77%, Train-80%, Subway-81%.

We observed that there were relatively few samples provided for activity-Run in Sussex-Huawei Locomotion - Transportation (SHL) dataset [13] [5] from Figure 1. To overcome the issue of imbalanced dataset, weighted random forest model can be evaluated on this dataset. Furthermore, the window size has been considered of a fixed length of 5 seconds. It is possible that adjusting the window size to be of a shorter duration 2-3 seconds might yield better results [2]. The recognition result for the testing dataset will be presented in the summary paper of the challenge [12].

REFERENCES

[1] Masud Ahmed, Anindya Das Antar, Tahera Hossain, Sozo Inoue, and Md. Atiqur Rahman Ahad. 2019. POIDEN: position and orientation independent deep ensemble network for the classification of locomotion and transportation modes. 674–679. https://doi.org/10.1145/3341162.3345570

[2] Anindya Das Antar, Masud Ahmed, Mohammad Shadman Ishrak, and Md Atiqur Rahman Ahad. 2018. A Comparative Approach to Classification of Locomotion and Transportation Modes Using Smartphone Sensor Data. *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (2018).

[3] James Bergstra and Yoshua Bengio. 2012. Random Search for Hyper-Parameter Optimization. *Journal of Machine Learning Research* 13, 10 (2012), 281–305. http://jmlr.org/papers/v13/bergstra12a.html

[4] L. Breiman. 2001. Random Forests. *Machine Learning* (2001). https://doi.org/10.1023/A:1010933404324

[5] H. Gjoreski, M. Ciliberto, L. Wang, F.J.O. Morales, S. Mekki, S. Valentin, and D. Roggen. 2018. The University of Sussex-Huawei Locomotion and Transportation Dataset for Multimodal Analytics With Mobile Devices. *IEEE Access* 6 (2018), 42592–42604. https://doi.org/10.1109/access.2018.2858933

[6] Vito Janko, Nina Resçiç, Miha Mlakar, Vid Drobnic, Matja Gams, Gasper Slapnicar, Martin Gjoreski, Jani Bizjak, Matej Marinko, and Mitja Lustrek. 2018. A New Frontier for Activity Recognition: The Sussex-Huawei Locomotion Challenge. *Proceedings of the 2018 ACM International Joint Conference and 2018 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (2018).

[7] Sumit Majumder and M. Jamal Deen. 2019. Smartphone Sensors for Health Monitoring and Diagnosis. *Sensors (Basel)* 19, 9 (May 2019), 2164. https://doi.org/10.3390/s19092164

[8] Philipp Probst, Marvin N. Wright, and Anne-Laure Boulesteix. 2019. Hyperparameters and Tuning Strategies for Random Forest. *ArXiv* abs/1804.03515 (2019).

[9] J. E. Siegel, D. C. Erb, and S. E. Sarma. 2018. A Survey of the Connected Vehicle Landscape—Architectures, Enabling Technologies, Applications, and Development Areas. *IEEE Transactions on Intelligent Transportation Systems* 19, 8 (2018), 2391–2406.

[10] X. Su, H. Tong, and P. Ji. 2014. Activity recognition with smartphone sensors. *Tsinghua Science and Technology* 19, 3 (2014), 235–249.

[11] X. Su, H. Tong, and P. Ji. 2014. Activity recognition with smartphone sensors. *Tsinghua Science and Technology* 19, 3 (2014), 235–249.

[12] L. Wang, H. Gjoreski, M. Ciliberto, P. Lago, K. Murao, T. Okita, and D. Roggen. 2020. Summary of the Sussex-Huawei locomotion-transportation recognition challenge 2020. *Proceedings of the 2020 ACM International Joint Conference and 2020 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers* (2020).

[13] L. Wang, H. Gjoreski, M. Ciliberto, S. Mekki, S. Valentin, and D. Roggen. 2019. Enabling Reproducible Research in Sensor-Based Transportation Mode Recognition With the Sussex-Huawei Dataset. *IEEE Access* 7 (2019), 10870–10891. https://doi.org/10.1109/access.2019.2890793

[14] L. Xu, W. Yang, Y. Cao, and Q. Li. 2017. Human activity recognition based on random forests. In *2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD)*. 548–553.