Where Are You? Human Activity Recognition with Smartphone Sensor Data

Gulustan Dogan
University of North Carolina Wilmington
Wilmington, NC, USA
dogang@uncw.edu

Iremnaz Cay
Istanbul Sabahattin Zaim University
Istanbul, Turkey
iremnaz.cay@std.izu.edu.tr

Sinem Sena Ertas
Istanbul Sabahattin Zaim University
Istanbul, Turkey
sinem.ertas@std.izu.edu.tr

Şeref Recep Keskin
Gazi University
Ankara, Turkey
seref.recep.keskin@gazi.edu.tr

Nouran Alotaibi
University of North Carolina Wilmington
Wilmington, NC, USA
na8103@uncw.edu

Elif Sahin
University of North Carolina Wilmington
Wilmington, NC, USA
es5510@uncw.edu

ABSTRACT

This paper describes our submission as Team-Petrichor to the competition that was organized by the SHL recognition challenge dataset authors. We compared multiple machine learning approach for classifying eight different activities (Still, Walk, Run, Bike, Car, Bus, Train, Subway). The first step was feature engineering, a wide set of statistical domain features were computed and their quality was evaluated. Finally, the appropriate machine learning model was chosen. The recognition result for the testing dataset will be presented in the summary paper of the SHL recognition challenge.

CCS CONCEPTS

• **Computing methodologies → Activity recognition and understanding**; **Supervised learning by classification**.

KEYWORDS

Activity Recognition, Locomotion Classification, Transportation Mode Prediction, Machine Learning

ACM Reference Format:
Gulustan Dogan, Iremnaz Cay, Sinem Sena Ertas, Şeref Recep Keskin, Nouran Alotaibi, and Elif Sahin. 2020. Where Are You? Human Activity Recognition with Smartphone Sensor Data. 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, 4 pages. https://doi.org/10.1145/3410530.3414354

1 INTRODUCTION

Human activity recognition aims to identify activities based on data collected by sensors [2]. It is one of the growing research areas that meet application demands, such as mobile computing, surveillance-based security, context-aware computing, and assistive living [1]. One of the difficulties facing this type of research is that each research group uses its own datasets and their own recognition tasks, making it hard to compare methodologies to advance research in this field [8, 11]. The literature shows that it is possible to use supervised machine learning and deep learning methods to recognize the human activity [3, 5]. The SHL recognition challenge makes comparing methodologies easier by providing unified datasets to competitors. The main goal of this SHL recognition challenge is to develop a methodology to recognize eight locomotion and transportation activities (Still, Walk, Run, Bike, Bus, Car, Train, Subway) from the inertial sensor data of a smartphone [8, 11].

The 2020 SHL recognition challenge is mainly focused on recognizing modes of transportation in a user-independent manner with an unknown phone position. The goal was to recognize the user transportation (activities) from data coming from the phone of the "test" user while the location of that phone on the "test" user is not specified. For the training data, the four phone positions for a "train" user is given. Also, a small "validation" dataset is provided, which includes data from the "test" user and all the four possible phone locations on the test user.

Our submission (team name: Petrichor) proposes a machine learning model to predict the transportation modes using feature engineering on the given raw sensor data as input. We did compare multiple machine learning models and used the one that has higher accuracy. We also had results with higher accuracy when using a voting classifier for three machine learning algorithms, which are Random Forest, Gradient Boosting and Gaussian Naive Bayesian.

2 DATASET AND TASK

The source of dataset for this SHL recognition challenge is the Sussex-Huawei Locomotion Dataset [4, 10]. This dataset was recorded by three participants in 8 transportation and locomotion activities which are: Still, Walk, Run, Bike, Car, Bus, Train, and Subway. Each participant carried four smartphones at four body positions independently, and the position of the phone is was unknown to them. This dataset contained train, validate, and test subsections. The train, validation and test data was generated by segmenting the whole data with a non-overlap sliding window of 5 seconds.

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/10.1145/3410530.3414354

The raw data for this dataset is from acceleration, gravity, rate of turn, linear acceleration, magnetic field, orientation of the device in quaternions, atmospheric pressure, and activity classes. The sampling rate of all the sensors is 100 Hz. The goal (task) is to recognize 8 modes of locomotion and transportation (activities) from the inertial sensor data of a smartphone. This is done by recognizing the user activity of data coming from the phone of the "test" user, that is the combination of user 2 and 3.

The number of days used for training data was 59, for the validation data 6 days and for test data 40. The frames for the train data are consecutive in time (196072 lines x 500 columns). Training data is provided for "train" user (user 1), of all the 4 phone positions (bag, hips, torso, hand), one of them will be identical to the "test" user, and the "validation" is provided. This includes data from the "test" user and all the 4 possible phone locations on the test user. The validation data is extracted from the already released preview of the SHL dataset (28789 lines x 500 columns). The test data (57573 lines x 500 columns) contains the raw sensors data from the other two users (2 and 3), and one phone location (unknown to the participants), with the same files as the train dataset but no class label, this is the data on which Machine Learning predictions has been made. The frames have been shuffled, so that the real time performance of the classification can be challenged.

3 METHOD AND EXPERIMENTS

Feature extraction has been done to reduce high dimensionality of the data. Feature extraction is done by using various methods, such as taking the mean, standard deviation (STD) and mean absolute deviation (MAD). Finally, a model is created through machine learning techniques. After training the model, our model will be evaluated in the validation phase and an estimate is provided about the effectiveness of its performance. The accuracy of the model on testing data will be shared in summary publication of the SHL recognition challenge. Below, our approach to feature extraction and machine learning is discussed.

3.1 Data Processing

Each file contains 196072 lines x 500 columns, corresponding to 196072 frames each containing 500 samples (5 seconds at the sampling rate 100 Hz). Each file in the validation data contains 28789 lines x 500 columns, corresponding to 28789 frames each containing 500 samples (5 seconds at the sampling rate 100 Hz). To keep data's original frame size, all sensor files have 500 columns. Any missing or inappropriate values were not detected in dataset.

For every feature, label.txt file indicates activity mode for every sensor data. This means 500 samples has 500 labels, respectively. Most of the rows in label.txt file has all identical values across one row meaning the user did not change transportation-locomotion mode in 5 seconds. Some rows had multiple values meaning user changed transportation-locomotion mode in 5 seconds. Frames (rows in label.txt file) which changes transportation mode during 5 seconds (500 sample) were removed from the dataset. In total, we ignored 0,296% of (581 row) of training data and 0,361% (104 row) of validation data. Thus, our prediction model only focus on continuous activities within the same frame. Since all the samples in the same frame have the same activity, we reduced these 500 labels in one row to one label.

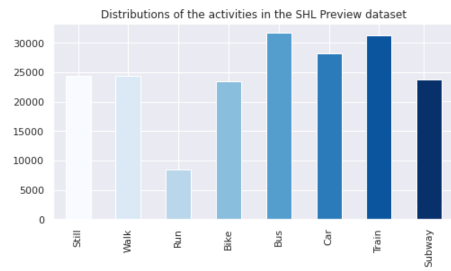

Figure 1: Distributions of the activities in the SHL preview train dataset.

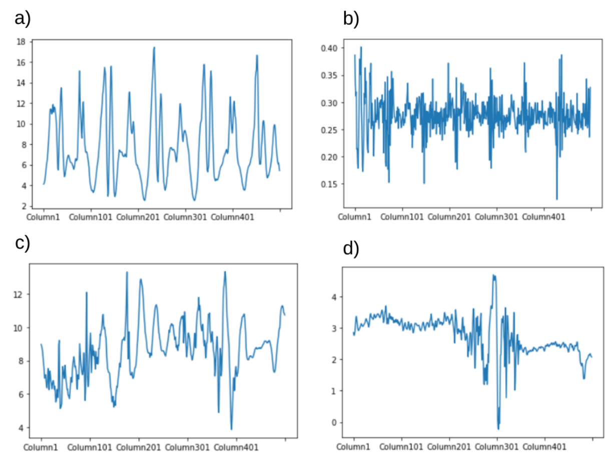

Figure 2: 1 frame 500 sample (a:Walk, b:Bike, c:Bus, d:Train) data frequency overview for Acceleration (Acc).

3.2 Feature Extraction

Human activities are carried out for relatively long periods (in seconds or minutes) compared to the sampling rate of the sensors. A single instance at a given moment does not provide enough information to describe the activity performed. It is almost impossible for the incoming signals to be exactly same as the actual performed movement. For this reason, the activities should be evaluated as a whole, not as an individual data sample. Statistical and structural approaches are preferred approaches for extracting features from time series data [7].

SHL dataset sampled at 5 seconds and 100 Hz sampling rate. Each dataset (train, validation,test) contains 500 samples and varying number of frames. The distribution of the activities for in the SHL training dataset is shown in Figure 1. All techniques are designed to support the high variability of the signals.

Feature extraction was performed by calculating the mean, Mean Absolute Deviation (MAD), Standard Deviation (STD), and minimum and maximum value of each 500 samples for every row. These methods were used to reduce high dimensionality, and also to make the prediction algorithm sensitive to the spread of the data. Mean calculations occur in a dataframe by summing up each row sample value and dividing it to the column number. This mean operation is an efficient method to express average sensor value for each row (500 samples). Mean absolute deviation was calculated for SHL data to get insight to distinguish the activity modes with the positive average distance between each data and the mean in each row. If any user does different activities, collected data frequencies will be different from each other when with accelerometer, gyroscope, etc., in each frame as it is illustrated in Figure 2.

$$s_x = \sqrt{\sum_{i=1}^{n}(x_i - x_{avg})^2 \frac{1}{n-1}} \tag{1}$$

$$MAD = \sum |x + x_{mean}| \frac{1}{n} \tag{2}$$

$$x_{mean} = \sum_{i=1}^{N} x_i \frac{1}{N} \tag{3}$$

Standard deviation was calculated to see how the data is spread out for each frame as seen in Equation 1. Minimum and maximum values were added into the dataframe so it can be used when summarizing the 500 columns into five features: the Mean (Equation 3), MAD (Equation 2), STD (Equation 1), Min, and Max values. As a result, we have five features for each sensor data files (Acc_x, Acc_y, Acc_z, Gra_x, Gra_y, Gra_z, Gyr_x, Gyr_y, Gyr_z, Mag_x, Mag_y, Mag_z, LAcc_x, LAcc_y, LAcc_z, Ori_w, Ori_x, Ori_y, Ori_z, Pressure). As we have four smartphones (Bag, Hand, Hips, Torso) in different locations, all calculation processes are done for every phone's sensor data. Overall, 400 statistical features were extracted from the raw data (5 features for each 20 types of data).

Table 1: Feature extraction methods applied to the data

	Torso/ Hips/ Bag/ Hand
Acc-X,Y,Z	Mean, Std, Min, Max, Mad
Gyr-X,Y,Z	Mean, Std, Min, Max, Mad
Gra-X,Y,Z	Mean, Std, Min, Max, Mad
Mag-X,Y,Z	Mean, Std, Min, Max, Mad
LAcc-X,Y,Z	Mean, Std, Min, Max, Mad
Ori-W,X,Y,Z	Mean, Std, Min, Max, Mad
Pressure	Mean, Std, Min, Max, Mad

Table 1 summarizes the feature extraction methods for acceleration (Acc), gravity (Gra), rate of turn (Gyr), linear acceleration (LAcc), orientation of the device in quaternions (Ori), and atmospheric pressure (Pressure) signals. These feature extraction methods applied to the data for each phone position to be used in machine learning models. They are applied for training, validation, and testing data.

3.3 Classification Algorithms

The recent development of sensor devices has simplified the data collection process. However, discovering information requires more effort than merely collecting data. Information discovery is very important in sensor data, and raw data is often useless for this step. Therefore, HAR systems make use of machine learning tools that are helpful in building patterns to describe, analyze, and predict data. Since a human activity recognition system should return a label such as walking, running; most HAR systems work in a supervised fashion. In a machine learning context, patterns are to be discovered from a set of given examples or observations denominated instances. Supervised learning is one of the critical fields, bringing a great number of algorithms that have been used in Human Activity Recognition. Most classifiers are capable of learning complex class structures.

Random Forest, Gradient Boosting, Gaussian Naive Bayesian methods gave the best accuracies during our implementation. Therefore we decided to use a Voting classifier to combine them. The main purpose of the Voting Classifier is to merge conceptually different machine learning classifiers and use a majority vote or the average predicted probabilities (soft vote) to predict the class labels [6]. Figure 3 depicts the soft voting system. The calculation for smart voting is shown in Equation 4, where P is the predicted probability for each classifier c [6]. Such a classifier is useful for models that have equally well performing results to balance out their individual weaknesses.

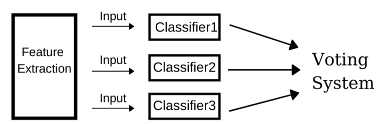

Figure 3: Schematic diagram of our voting system.

$$y = argmax \frac{1}{N_c} \sum_c (P_1, P_2, P_3 .... P_n) \tag{4}$$

4 RESULTS

In this section, we briefly show the results of the most successful Machine Learning models. All supervised machine learning algorithms are implemented and the methods with the highest accuracies are selected.

Python programming language was used at every stage of the study. Pandas, numpy, ggplot libraries were used for data related operations. In addition, we ran all machine learning calculation experiments with the scikit-learn library. We used a dual processor machine with 16 cores and 256GB of RAM for computation. We have saved the trained model of 410 MB size on the hard disk and it takes approximately 3.36 hours to train the model. It took to evaluate the test dataset approximately 3.06 seconds.

Figure 4 shows validation accuracies in different scenarios, where we used different modalities. These results were achieved before the merging of validation and training data, only train data is used for training. We started training our model with only two modalities which are Acceleration and Gyroscope. After getting our first accuracy, we continue adding modalities to observe model performance. Adding all sensor modalities made improvement in model performance except Ori files (Ori_w,Ori_x,Ori_y,Ori_z). So, we decided not to include Ori sensor data in any part of the training phase.

When using only the training set for model training, we achieve an accuracy of 61.2% with the validation set. The model performance has been improved effectively by incorporating the validation data

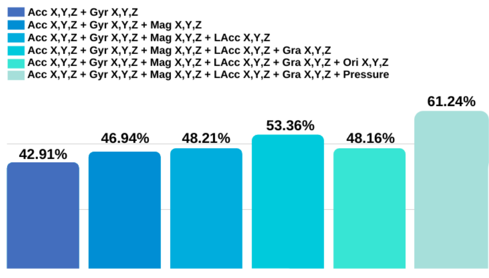

Figure 4: Validation accuracy based on used modalities.

Table 2: The success rates before and after merging of validation data with test data.

Train Set	Validation Set	Random Forest Classifier	Gradient Boosting Classifier
Train	Validation (100%)	61%	63%
Train + Validation (20%)	Validation (80%)	73%	71%

for model training. We took 20% of validation data into training data to evaluate the improved accuracy score. An equal number of samples from each locomotion-transportation mode category were taken from validation data to avoid overfitting. Rest of the validation data is used for testing the model. When both the training data and some part of the validation data is used for model training, we achieve an accuracy of 73% as it is shown in Table 2.

Table 3: Machine learning models success rates.

Machine Learning Model	Validation Accuracy	Validation Precision	Validation Recall	Validation F1 Score
Voting System (Random Forest + Gradient Boosting + Gaussian NB)	73%	74%	66%	68%
Voting System (Random Forest + Gradient Boosting + Decision Tree)	71%	74%	63%	64%
Random Forest	73%	74%	65%	65%
Gradient Boosting	71%	70%	66%	68%
Gaussian Naive Bayes	53%	59%	48%	50%
AdaBoost	47%	47%	50%	47%

After both the training and validation set is used for model training, model accuracy, precision, recall, and F1, scores are calculated as shown in Table 3. As can be seen in the table, the most successful machine learning algorithms are Random Forest, Gradient Boosting and Gaussian Naive Bayesian. Therefore, we used those algorithms in the voting classifier system. The ensemble model takes advantage of the different algorithms and yields better performance than a single one. Voting classifier can be a good choice when a single strategy is not capable of reaching the desired accuracy. Voting system is used for combining the predictions from multiple machine learning algorithms and performance improvement.

The confusion matrix of our model is presented in Figure 5. Our model can classify still, bike, walk, and car activities better than the other categories. We believe this is because the other activities are considered to have faster movement; thus they have similar sensor patterns.

5 CONCLUSION

We recognized the locomotion and transportation modes of user activity from the inertial sensor data of a smartphone by using

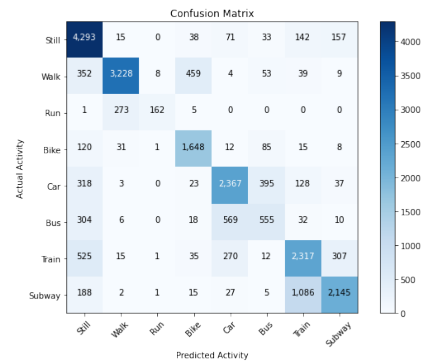

Figure 5: Confusion matrix of voting system prediction.

supervised Machine Learning models with the Sussex-Huawei Locomotion Dataset. Random Forest, Gradient Boosting, and Gaussian Naive Bayesian algorithms performed the best between the developed models. We used these models with the Voting System (Classifier). Our findings show that activities which have slower movement are better recognized by our model. Finally, recognition result for the testing dataset will be presented in the summary paper of the SHL recognition challenge [9].

REFERENCES

[1] Liming Chen, Jesse Hoey, Chris D Nugent, Diane J Cook, and Zhiwen Yu. 2012. Sensor-based activity recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 42, 6 (2012), 790–808.

[2] Diane Cook, Kyle Feuz, and Narayanan Krishnan. 2013. Transfer Learning for Activity Recognition: A Survey. *Knowledge and information systems* 36 (09 2013), 537–556. https://doi.org/10.1007/s10115-013-0665-3

[3] Ayşenur Gençdoğmuş, Şeref Recep Keskin, Gulustan Dogan, and Yusuf Ozturk. 2019. A Data-Driven Approach to Kinematic Analytics of Spinal Motion. 2222–2229. https://doi.org/10.1109/BigData47090.2019.9006164

[4] Hristijan Gjoreski, Mathias Ciliberto, Lin Wang, Francisco Javier Ordonez Morales, Sami Mekki, Stefan Valentin, and Daniel Roggen. 2018. The university of sussex-huawei locomotion and transportation dataset for multimodal analytics with mobile devices. *IEEE Access* 6 (2018), 42592–42604.

[5] Şeref Recep Keskin, Ayşenur Gençdoğmuş, Buse Yildirim, Gulustan Dogan, and Yusuf Ozturk. 2020. DNN and CNN Approach for Human Activity Recognition. 254–258. https://doi.org/10.1109/ICEEE49618.2020.9102624

[6] Sanchita Mangale. 2019. Voting Classifier. https://medium.com/@sanchitamangale12/voting-classifier-1be10db6d7a5,

[7] Robert T Olszewski. 2001. *Generalized feature extraction for structural pattern recognition in time-series data.* Technical Report. Carnegie-Mellon Univ Pittsburgh Pa School Of Computer Science.

[8] Lin Wang, Hristijan Gjoreski, Mathias Ciliberto, Paula Lago, Kazuya Murao, Tsuyoshi Okita, and Daniel Roggen. 2019. Summary of the Sussex-Huawei locomotion-transportation recognition challenge 2019. In *Adjunct Proceedings of the 2019 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2019 ACM International Symposium on Wearable Computers.* 849–856.

[9] Lin Wang, Hristijan Gjoreski, Mathias Ciliberto, Paula Lago, Kazuya Murao, Tsuyoshi Okita, and Daniel Roggen. 2020. Summary of the Sussex-Huawei locomotion-transportation recognition challenge 2020. In *Proceedings of the 2020 ACM International Joint Conference and 2020 International Symposium on Pervasive and Ubiquitous Computing and Wearable Computers, 2020.*

[10] Lin Wang, Hristijan Gjoreski, Mathias Ciliberto, Sami Mekki, Stefan Valentin, and Daniel Roggen. 2019. Enabling reproducible research in sensor-based transportation mode recognition with the Sussex-Huawei dataset. *IEEE Access* 7 (2019), 10870–10891.

[11] Lin Wang, Hristijan Gjoreskia, Kazuya Murao, Tsuyoshi Okita, and Daniel Roggen. 2018. Summary of the sussex-huawei locomotion-transportation recognition challenge. In *Proceedings of the 2018 ACM international joint conference and 2018 international symposium on pervasive and ubiquitous computing and wearable computers.* 1521–1530.