UPIC: User and Position Independent Classical Approach for Locomotion and Transportation Modes Recognition

Md Sadman Siraj[1,3], Md Ahasan Atick Faisal[1,4], Omar Shahid[1,5], Farhan Fuad Abir[1,6], Tahera Hossain[2,7], Sozo Inoue[2,8] and Md Atiqur Rahman Ahad[1,9]

University of Dhaka, Bangladesh[1]; Kyushu Institute of Technology, Japan[2]

sadman.siraj@ieee.org[3], atickfaisal@gmail.com[4], omarshahid232@gmail.com[5], farhan.fuad@ieee.org[6], taheramoni@gmail.com[7], sozo@brain.kyutech.ac.jp[8], atiqahad@du.ac.bd[9]

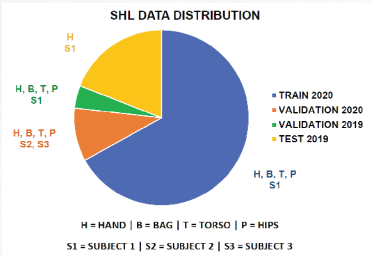
UBICOMP 2020

Introduction

- Presented a concise summary on Sussex-Huawei Locomotion-Transportation (SHL) dataset for locomotion and transportation activity analytics
- Used rotation matrix to make the collected data position invariant
- Extracted both time domain and frequency domain features
- Designed a feature selection technique by combining few existing methods
- Selected train set and validation set in a way that ensure the model turns into more general rather than performing better in a specific combination
- The whole designed system is more general and require less computation power to develop the learning model

Sussex-Huawei Locomotion-Transportation (SHL) Dataset

- Dataset included eight modes of locomotion and transportation - walk, run, bike, still, car, bus, train and subway.
- Data acquisition was done using smartphone sensors using a non-overlapping window of 5 seconds at 100 Hz sampling rate.
- Dataset included data from four smartphone positions - bag, hand, hips and torso.
- Dataset consists of 20 sensor channels in total from 7 sensors - accelerometer, gyroscope, magnetometer, linear acceleration, gravity, orientation, and ambient pressure sensors.



- The train set contained data of user-1 from all four positions whereas the validation set contained data of user-2 and user-3

Methodology

Train and Validation Set Modification

We modified the given train and validation set by adding additional data from 2019 challenge dataset. This resulted in a training set consisting of all 3 users data from all 4 positions.

Train Set	Validation Set
VALIDATION 2020 VALIDATION 2019 TEST 2019	TRAIN 2020

Derived Data Channel

- Magnitude is calculated acceleration, linear acceleration, magnetometer, gravity and gyroscope.
- Orientation sensor values are used to derive a rotation matrix which was used to rotate the given acceleration values into earth axis.
- Vertical and horizontal acceleration channels were calculated from linear acceleration and gravity
- Jerk was calculated by taking the derivative of the acceleration channels.
- Another data channel was derived by taking the derivative of pressure.

Feature Extraction

We have extracted in total **789** features in both time and frequency domain from **28** given and derived data channels.

Selected Features

Channels	Time Domain Features	Frequency Domain Features
Acceleration (x, y, z, mag), Linear Acceleration (x, y, z, mag), Gravity (x, y, z, mag), Magnetometer (x, y, z, mag), Gyroscope (x, y, z, mag), Vertical Acceleration, Horizontal Acceleration, Jerk (Total, Body, Horizontal and Vertical), Pressure, Derivative of pressure	Min, Max, Peak to Peak Range, Average, Standard Deviation, Variance, Max Rate of Change, Average Rate of Change, Mean Absolute Deviation, Interquartile Range, Correlation Coefficient, Mean Crossing Rate, Mutual Correlation (X-y), Covariance (X-y), Signal Magnitude Area, Root Mean Square, Energy, Linear Velocity	Max Spectral Power, Center Frequency, Dominant Frequency, Entropy, Spectral Energy, Skewness, Kurtosis, Number of Peaks, First 10 FFT Coefficients

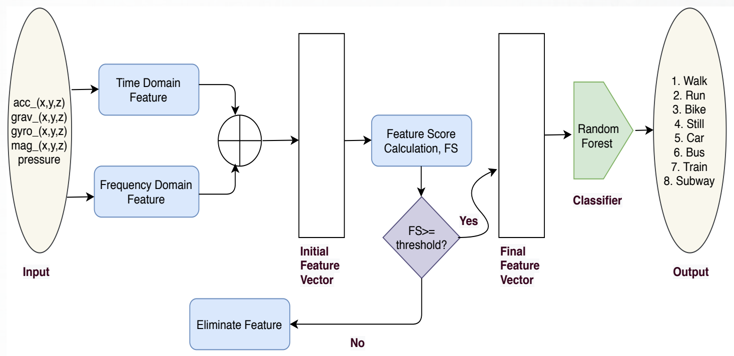
Feature Selection

- Six different feature selection techniques - Mutual Information, Chi-square Test, Tree-based Selection, Pearson Correlation Coefficient, Spearman Correlation Coefficient and ANOVA F-value were at first used to determine the importance score.
- Individual features scores from all six methods were averaged to create the final feature importance list.
- Importance score of each feature from the averaged feature importance list was compared with a threshold value.
- Lastly 349 features were found from the threshold comparison

Classifier

Random Forest Classifier - Maximum depth of trees = 70, Number of trees = 300

Block Diagram of the Proposed Method

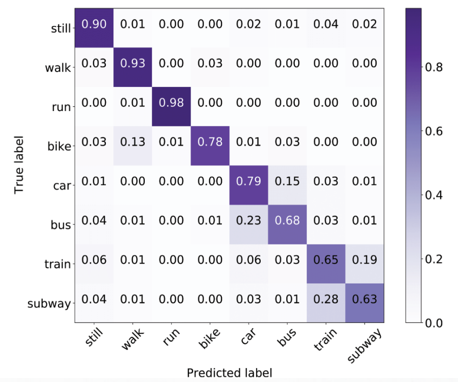


Result and Analysis

Training and validation accuracy of based on the modified training and validation sets are as shown. The trained model shows comparable accuracy in individual positions as well as the combined data. It is due to the combination of the all positions and all users data in the training set.

Position	Training Accuracy	Validation Acccuracy
Bag	92.69%	77.81%
Hand		77.57%
Hips		74.91%
Torso		77.85%
All		77.04%

Confusion Matrix



From the confusion matrix it can be interpreted that the the model makes a few wrong predictions between bus-car and train-subway. This is due to the subtle similarities between these two sets of transports.

Reference

[1] 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.

[2] 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.