Transition-points-based Segmentation and Hierarchical Classification for Locomotion and Transportation recognition on Radio-data

Nhat Tan Le, Nazmun Nahid, Sozo Inoue
Sozolab, Kyushu Institute of Technology

Introduction

- The 2021 Sussex-Huawei Locomotion-Transportation (SHL) recognition challenge aims to handle 8 transportation and locomotion modes (Still, Walk, Run, Bike, Car, Bus, Train, and Subway) based on Radio-data (GPS reception, GPS location, Wifi reception, and GSM cell tower scans).
- Commonly, transportation mode recognition is achieved by analyzing the accelerometer and gyroscope data. However, the combination of GPS, GIS, Wi-Fi, and Cellular data can also solve that problem and enrich information about location and mobile connection.
- Our team (HuBi) proposed a pipeline that includes transition-points-based segmentation and hierarchical classification based on our deep analysis of this type of dataset, which achieved an overall accuracy of 75% on the Validation dataset.

Methodology

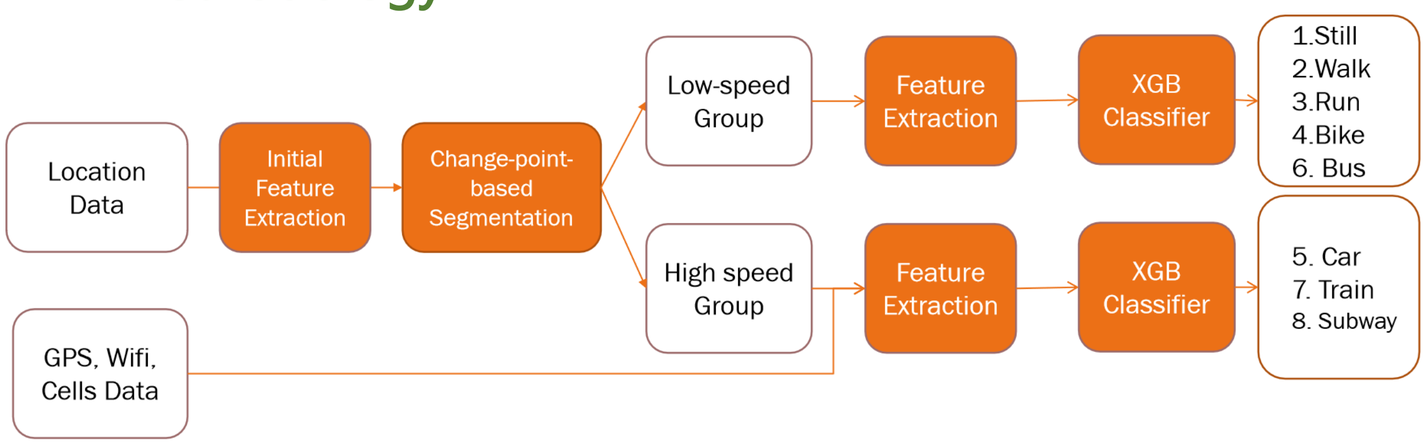

Figure 1: The process of two classification layers and transition-points-based segmentation.

Our pipeline contains 2 main stages:

- First classifier along with the transition-point-based segmentation: classifying GPS points into 2 groups by **SVM model** trained on **velocity and acceleration features**, then relying on point-clusters of 2 groups (low-speed and high-speed) to handle the segmentation.
- Second classifier: each group has different feature extraction strategies to train the specific XGBoost model.

First Classifier

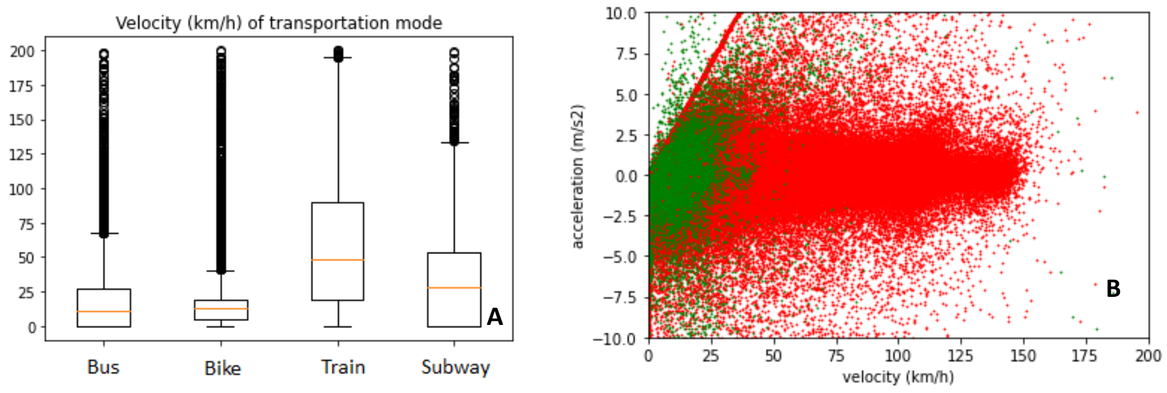

Figure 2. A: The bus frequently moves with a velocity not higher than the non-motorized transportation (such as Run, Bike) as the distance between two stops is not long enough to keep the high velocity within this short time. Consequently, **we considered the bus as a low-speed transportation mode**.

B: The velocity and acceleration of low-speed (green) vs. high-speed modes (red). With this distribution, we applied the **SVM** to handle the first classifier.

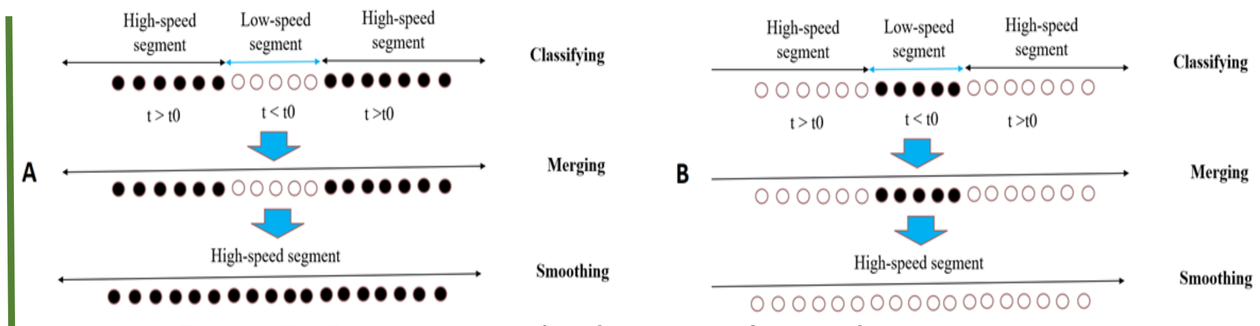

Figure 3: Segmentation by **detecting the mode transition point**. (black: high-speed point, white: low-speed point).

The **smoothing process reduces the fake transition points** in some cases:

- The train or subway trips after classifying with SVM have many small low-speed segments due to their stop time. These problems can be solved by merging and smoothing (A).
- Similar to this concept, in bus trips the short high-speed segments are merged into the whole low-speed segment (B).

Second Classifier

Features	Low-speed	High-speed	Features	Low-speed	High-speed
Distance*	X	X	GPS satellites change rate (CR)		X
Latitude**		X	Wifi connection points CR		X
Longitude**		X	Cell Entries CR		X
Velocity*	X	X	Stop rate	X	X
Acceleration*	X	X	Velocity CR	X	X
GPS satellites**		X	Heading CR	X	X
Wifi connection points **		X			
*Cell Entries***		X			

* Including max, mean, standard deviation, percentile 50, and 75.
** Including max, mean, standard deviation

Table 1: All training features

- In high-speed group, features are **directly extracted** from the high-speed window after transition-point segmentation step. Because high-speed mode can not transit to each other. Some **advance features related to connection points changes** per distance/time were applied in this case.
- In low-speed group, **a time-window segmentation (4 minutes)** is implemented before extracting features.
- XGBoost model was used to train these two group models because of its high performance.

Results

Evaluation Metrics	Without Smoothing & Label Correction			With Smoothing & Label Correction		
	G0	G1	Total	G0	G1	Total
Precision	89	90	89	92	99	95
Recall	96	77	86	99	82	91
F1 Score	92	83	87	95	90	93
Accuracy	89			94		

Table 2: Results of The First Classifier. **Smoothing process significantly increased the classification performance**. (G0: Low-speed group, G1: High-speed group)

	G0					G1		
	1	2	3	4	6	5	7	8
Precision	89	93	53	81	92	100	100	67
Recall	92	89	57	85	92	100	67	100
F1 Score	90	91	55	83	92	100	80	80

Table 3: Results of The Second Classifier. The overall accuracy of both classifiers in high and low-speed group was **88%.**

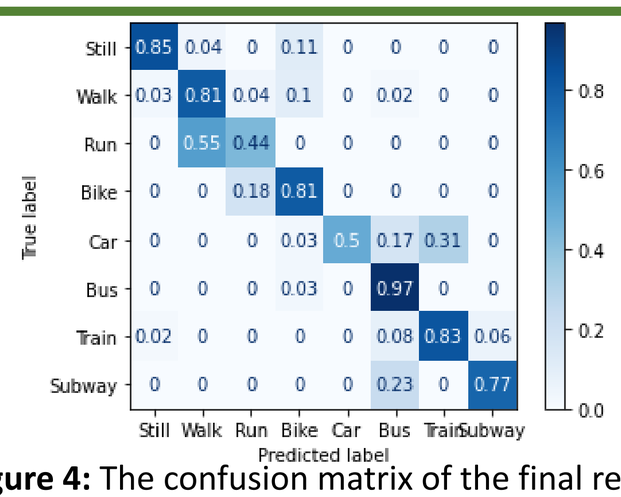

Figure 4: The confusion matrix of the final result in the validation data. **The overall accuracy was 75%.**

Analysis

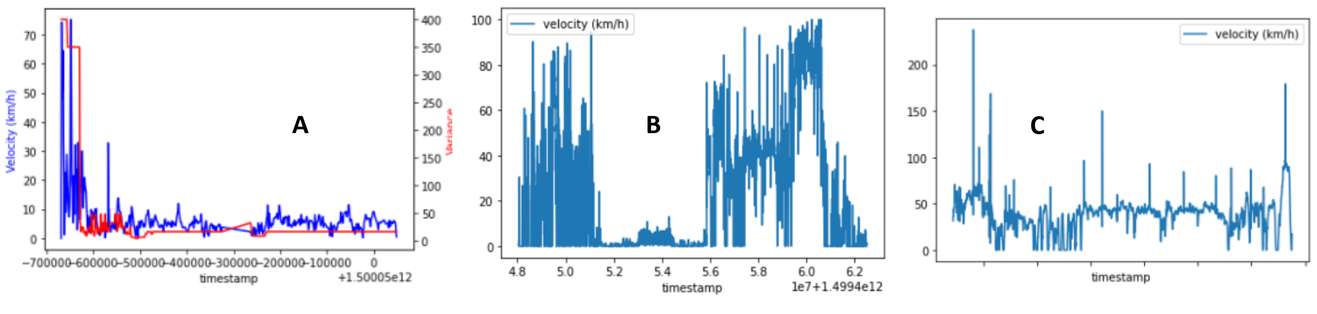

Figure 4: Some mis-prediction may be made due to **high error** in data collection (A), **mis-labeling** (B), and some **exceptional vehicle states** (C)

Conclusion

- The initial classifier and segmentation based on transition point significantly increase the final classification performance.
- The bus can be considered as low-speed transportation.
- Our work achieved an overall accuracy of 75% on the validation data, 94% in the first classifier, and 88% in the second classifier.
- Cases of false predictions are clearly analyzed in our study.