

Multiple Tree Model Integration for Transportation Mode Recognition

Yan Ren, Renmin University of China

Introduction

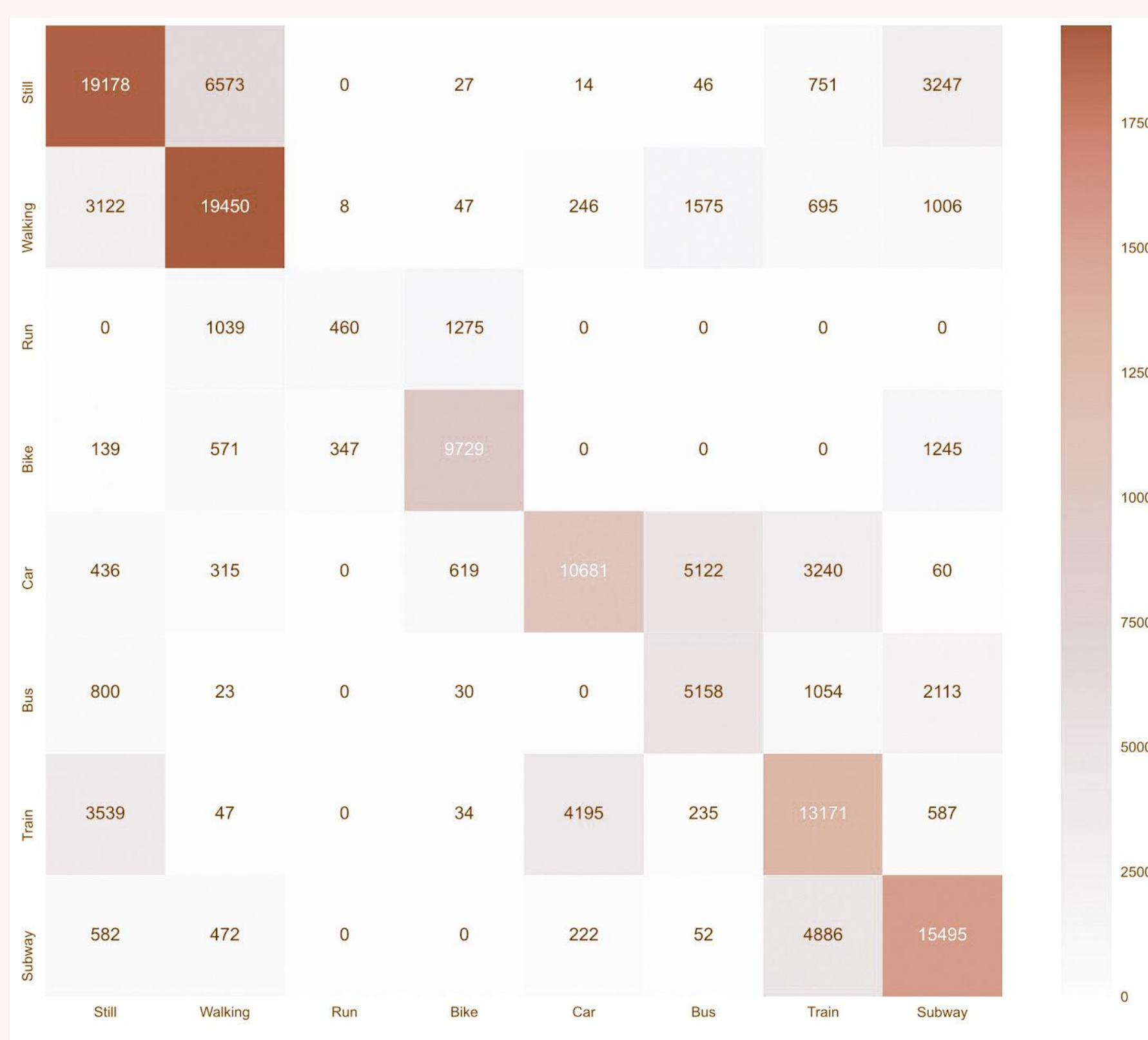
Sussex-Huawei Locomotion Transportation (SHL) recognition challenge 2021 aims at differentiating 8 transportation modes with 4 mobile phone signal sensor data, including Wifi, Cells, Location, GPS access and detail.

1	2	3	4	5	6	7	8
Still	Walking	Run	Bike	Car	Bus	Train	Subway

Result & Discussion

Overview

The method gets 0.65 accuracy score on validation dataset, which can set a baseline for the problem. The confusion matrix is shown as below.



Discussion

1. More semantic features should be derived with specialized knowledge in signal field. Currently, only conventional statistical methods are applied.
2. A more intelligent post-processing method is needed.
3. The method does not include deep learning framework. The accuracy is not as high as expected.
4. The model can not handle individual differences well.

Pipeline

Data Pre-Processing

Data pre-processing includes timestamp alignment, settling NaN value etc. A 0-1 feature is added for each signal source to discriminate between no signal received.

Feature Extraction

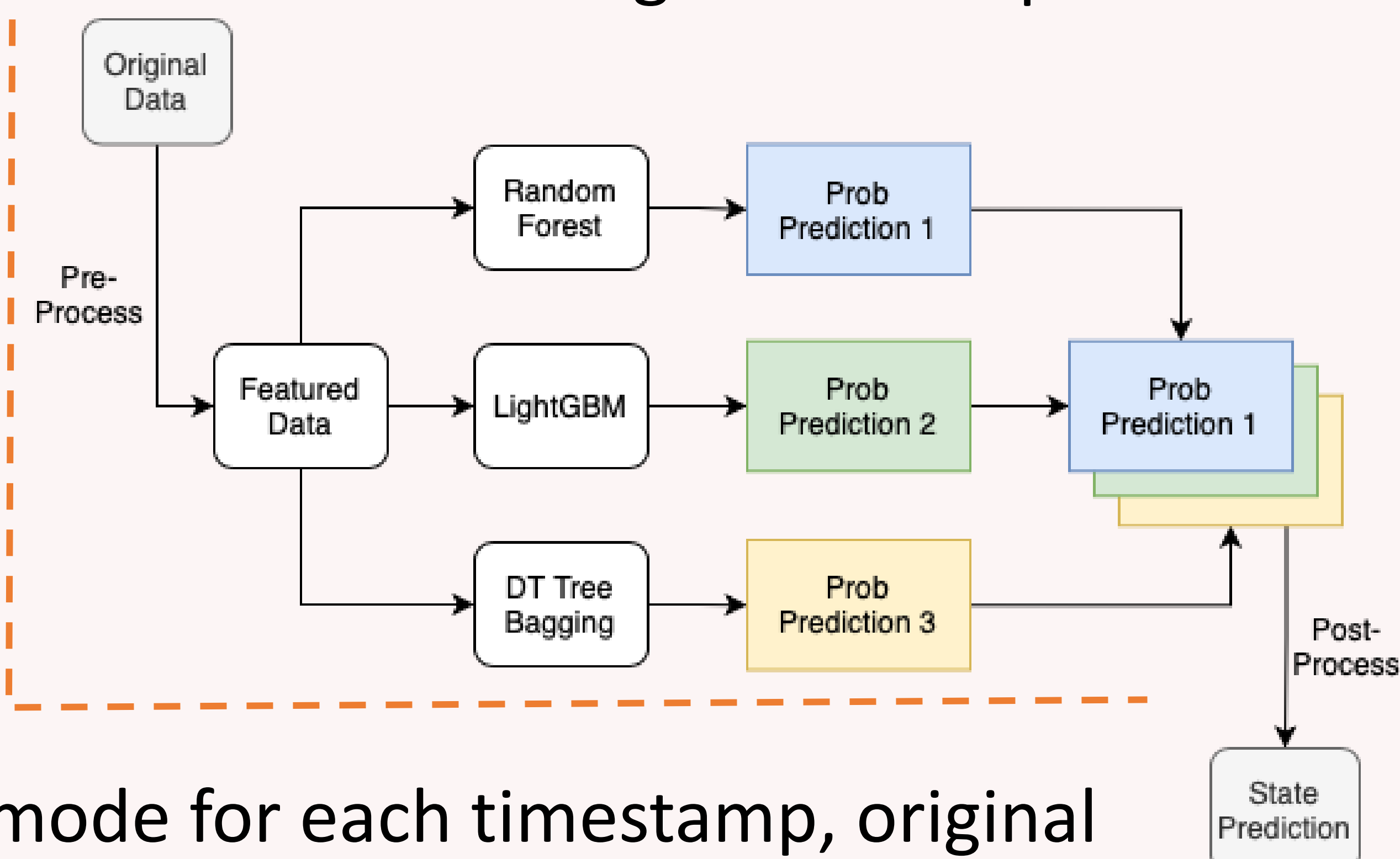
1. Basic features are classified into point features and window features. Both of them are in statistical domain.
2. Then a series of descriptive statistics are applied to screen out features that do not apparently discriminating.
3. Possible data transformations are also explored. e.g. log transformation is used for 'speed' to handle right-skew problem.

Finally, 32 discriminative features are extracted as listed in table.

File	Original Features	Derived Number	Method
Location	accuracy	1	original value
	speed	4	log transformation, window std, log of window maximum, std of window maximum
	acceleration	2	original value, window std
GPS	number	1	entry number
	snr	4	mean, min, max, std
WiFi	number	1	entry number
	rssi	5	mode, mean, min, max, std
Cells	frequency	1	mean
	number	1	entry number
	isRegistered	1	mean
	asuLevel	4	mean, min, max, std
	dBm	4	mean, min, max, std
	level	3	mean, min, max

Model Integration

Three tree based models (Random Forest, LightGBM, DT Tree Bagging) are trained separately. The results of them are combined to get the raw prediction.



Data Pre-Processing

To keep the continuity of transportation mode for each timestamp, original point predictions are replaced by mean values within the window (size of 120s).

Reference

- [1] H. Gjoreski, M. Ciliberto, L. Wang, F.J.O. Morales, S. Mekki, S. Valentin, and D. Roggen, "The University of Sussex-Huawei locomotion and transportation dataset for multimodal analytics with mobile devices," IEEE Access 6 (2018): 42592-42604. [2] L. Wang, H. Gjoreski, M. Ciliberto, S. Mekki, S. Valentin, and D. Roggen, "Enabling reproducible research in sensor-based transportation mode recognition with the Sussex-Huawei dataset," IEEE Access 7 (2019): 10870-10891.

Contact

Github: Ryanna-github
Email: renyan_ruc@163.com

