Mobility Data Representations for Spatiotemporal Tasks

Cristiano Landi[1, 3]

Supervisors: R. Guidotti[2, 3], A. Monreale[2, 3], M. Nanni[3]

[1]{name}.{surname}@phd.unipi.it; [2]{name}.{surname}@unipi.it; [3]{name}.{surname}@isti.cnr.it

IN SUPREMÆ DIGNITATIS 1343 UNIVERSITY OF PISA

OCTO

Consiglio Nazionale delle Ricerche

ISTI ISTITUTO DI SCIENZA E TECNOLOGIE DELL'INFORMAZIONE "A. FAEDO"

Introduction

Mobility data (MD) are everywhere. Smartphones and connected cars, as well as tracking devices with GPS capabilities, produce enormous amounts of spatiotemporal data. The most similar field in the literature is time series (TS), which involves streams of observations over a finite period. TS research is more extensively explored, particularly in classification tasks, where a wide variety of methods exist [1].

Comparing TS with mobility literature, we can observe that the former tends to focus more on report-style publications, emphasizing the results of the analysis rather than the methodologies employed.

Another key challenge in MD analysis is achieving geographic transferability of models. A model trained on data from one region may perform poorly when applied to another due to differences in the road network patterns or population behavior.

During my PhD, I'm focusing on developing fast, reusable, and effective trajectory representations suitable for multiple machine learning tasks, with an emphasis on geographic transferability and interpretability.

Methodologies

Raw GPS Trajectories

Trajectory Interval Forest

Geographic Shapelets

Bag of words

Conv. NN

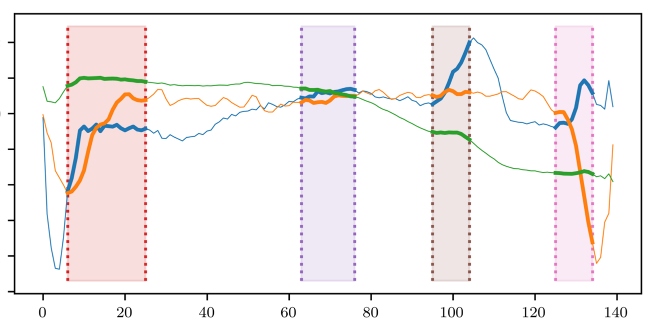
XAI

Classification, Regression and Clustering methods

Benchmark Geographic Transferability

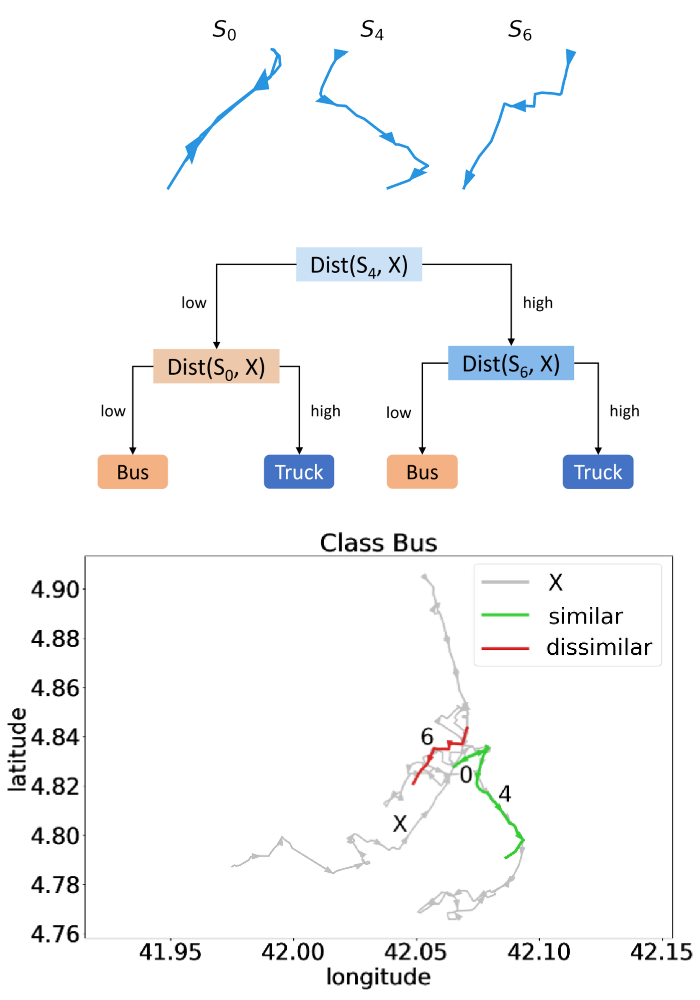
Preliminary Results

TIF [2]

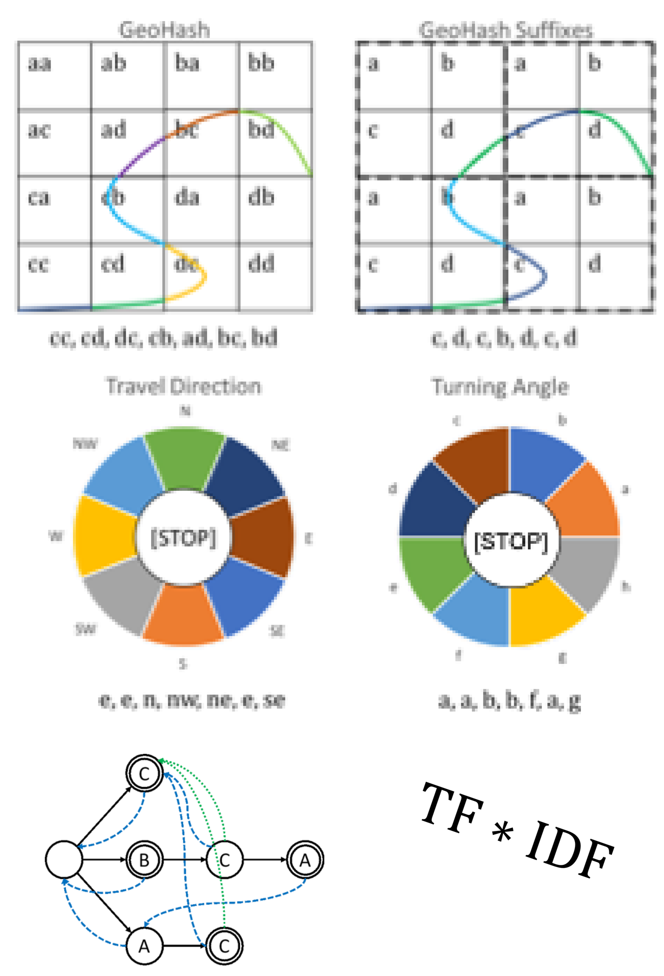


	Name	Equation
$\mathcal{P}$	distance	$dist_i = \|\|\vec{x}_i - \vec{x}_{i+1}\|\|$
	speed	$speed_i = \frac{dist_i}{t_{i+1}-t_i}$
	acceleration	$accel_i = \frac{dist_i}{(t_{i+1}-t_i)^2}$
	direction	$dir_i = arctan2\left(lat_{i+1} - lat_i, long_{i+1} - long_i\right)$
	turning angle	$turn_i = dir_{i+1} - dir_i$
$\mathcal{A}$	statistics	sum, max, min, mean, std, cov, var
	rate upper	$rate_{upper} = \frac{\|\{p>\theta \mid p\in\mathcal{P}(X)\}\|}{sum(\{dist_i \| i\in k\})}$
	rate below	$rate_{below} = \frac{\|\{p\leq\theta \mid p\in\mathcal{P}(X)\}\|}{sum(\{dist_i \| i\in k\})}$
$\mathcal{I}$	mean squared displacement	$msd = var(X_{lat}) + var(X_{long})$
	straightness	$str = \frac{\|\|\vec{x}_0-\vec{x}_m\|\|}{\sum_i dist_i}$
	intensity use	$iu = \frac{\sum_i dist_i}{\sqrt{Area\ of\ movement}}$
	sinuosity	$sin = 2\left[p\left(\frac{1-c^2-s^2}{(1-c)^2+s^2} + b^2\right)\right]^{-0.5}$

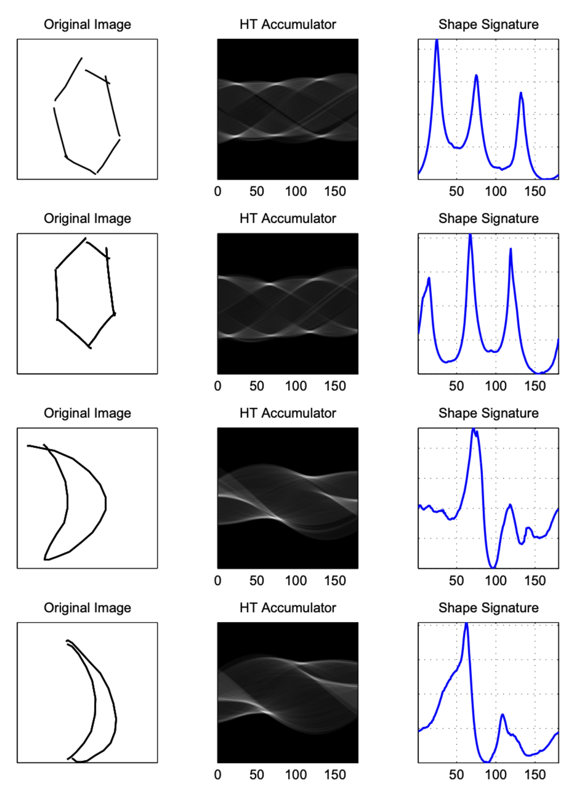
Geolet [3]



GASM [4]



Picture a TRJ [5]



From left to right: image (trajectory), Hough accumulator, shape signature by vertical summation of accumulator

On Geographic Transferability

Transfer a Local Models

Framework:

1. Source domain identification
2. Source-target domain linking
3. Target Domain refining

City A → City B

Transfer Dataset A→B

Ensemble of Local Models

Ensemble of Local models:

1. Most similar city transfer,
2. Weighted model ensemble
3. Weighted data sampling

Takeaway messages:

1. The transfer is better between similar cities
2. Accuracy(A→B) != Accuracy(B→A)

Source Dataset (fit)		Rome No-train	Rome Train D	Rome Train OD	New York No-train	New York Train D	New York Train OD	Athens No-train	Athens Train D	Athens Train OD	All cities No-train	All cities Train D	All cities Train OD
Rome	KNN	0.900	-	-	-0.033	0	-0.041	-0.058	-0.075	-0.075	-0.159	0.017	0.034
	DT	0.967	-	-	-0.484	0.025	-0.530	-0.564	-0.008	-0.481	-0.042	-0.009	-0.468
	RF	0.975	-	-	0.025	0	0	-0.017	-0.017	-0.008	0.008	0	0.017
NY	KNN	0.025	0.033	0.033	0.942	-	-	-0.025	-0.033	-0.017	0.017	0.008	0.008
	DT	-0.041	0.008	0.016	0.950	-	-	-0.033	-0.017	-0.025	-0.017	-0.025	0
	RF	0	0.008	0	0.958	-	-	-0.017	-0.033	-0.025	0	0.017	0.025
Athens	KNN	0.025	0.033	0.033	0.025	0.008	0.025	0.975	-	-	0.067	0.025	0.059
	DT	-0.041	0.008	0.016	0.033	-0.016	0.042	1.000	-	-	-0.034	-0.017	0.008
	RF	0	0.008	0	0.034	0.017	0.025	1.000	-	-	0.017	0.025	0.025
All	KNN	0	0.008	-0.025	-0.024	-0.024	-0.008	-0.066	-0.041	-0.057	0.933		
	DT	0.025	0.016	0.025	0.042	0.050	0.050	-0.025	-0.017	-0.008	0.992		
	RF	0.008	0.008	0.017	0.017	0.034	0.025	-0.017	-0.025	-0.017	0.975		

Destination Dataset (predict)

Table 1: F1-score deltas of geographically transferred models. Negative values indicate that the model trained and tested on the destination city performs better.

Conclusions

To sum up, we proposed TIF based on a survey of MD analysis, which served as one of the baselines for my work. Then, we introduced Geolet, the first shapelet-based method for raw Trajectories, and began investigating its capabilities and limitations.

We plan to integrate the developed trajectory transformations with other methods we are collaborating on [6], creating interpretable pipelines for Mobility Data Analytics.

Additionally, we plan to integrate deep learning techniques, for example generative models, to produce the discriminative sub-trajectories used by Geolet in the transformation.

References

[1] Middlehurst, Matthew, Patrick Schäfer, and Anthony Bagnall. "Bake off redux: a review and experimental evaluation of recent time series classification algorithms." Data Mining and Knowledge Discovery (2024): 1-74.

[2] Landi, Cristiano, et al. "The Trajectory Interval Forest Classifier for Trajectory Classification." Proceedings of the 31st ACM International Conference on Advances in Geographic Information Systems. 2023.

[3] Landi, Cristiano, et al. "Geolet: An Interpretable Model for Trajectory Classification." International Symposium on Intelligent Data Analysis. Cham: Springer Nature Switzerland, 2023.

[4] Landi, Cristiano, and Riccardo Guidotti. "A Shape-Based Map Matching Approach for Geographic Transferability of Discriminative Subtrajectories." EDBT/ICDT Workshops. 2024.

[5] Vlachos, Michail, et al. "Rotation invariant indexing of shapes and line drawings." Proceedings of the 14th ACM international conference on Information and knowledge management. 2005.

[6] Lusito, Salvatore, Andrea Pugnana, and Riccardo Guidotti. "Solving imbalanced learning with outlier detection and features reduction." Machine Learning 113.8 (2024): 5273-5330.

[7] Guidotti, Riccardo, et al. "Interpretable Data Partitioning Through Tree-Based Clustering Methods." International Conference on Discovery Science. Cham: Springer Nature Switzerland, 2023.