POPAyI: Muscling Ordinal Patterns for low-complex and usability-aware transportation mode detection

---

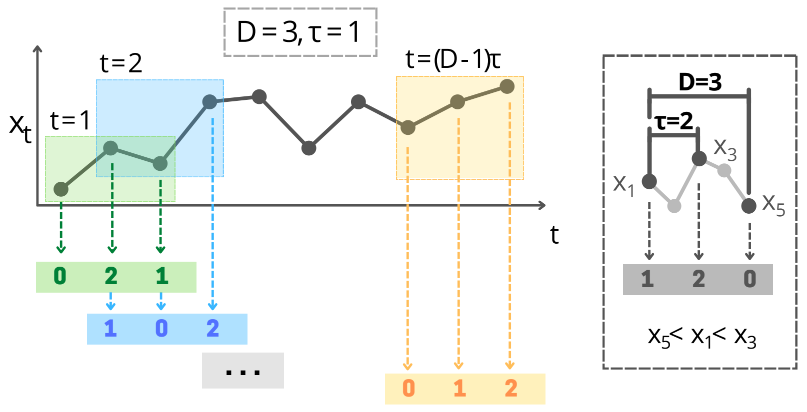
Isadora Cardoso-Pereira, João B. Borges, Aline C. Viana, Antonio A. F. Loureiro, **Heitor S. Ramos** - heitor@dcc.ufmg.br

Federal University of Minas Gerais (UFMG), Brazil, Federal University of Rio Grande do Norte (UFRN), Brazil, Inria, France

Introduction

- Transportation Mode Detection (TMD) involves classifying mobility traces to identify the corresponding transport mode.
- Most TMD literature relies on hand-crafted features (reaching hundreds of features) or computationally- and data-intensive DL methods, which pose challenges for resource-constrained Internet of Things (IoT) scenarios
- We present POPAyI, an innovative strategy based on Ordinal Pattern (OP) transformation applied to mobility-related time series that is computationally efficient, scalable, and provides accurate transportation mode detection.

Ordinal Patterns



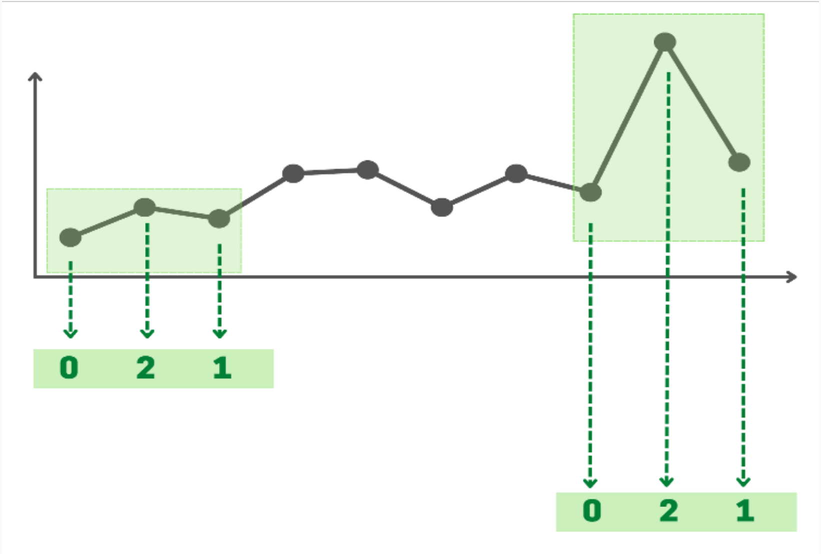
A lightweight solution well-suited for edge computing applications, enabling on-device analytics: preserving privacy

Ordinal Patterns Limitations

Traditional OPs face limitations when dealing with multidimensional data (e.g., mobility)

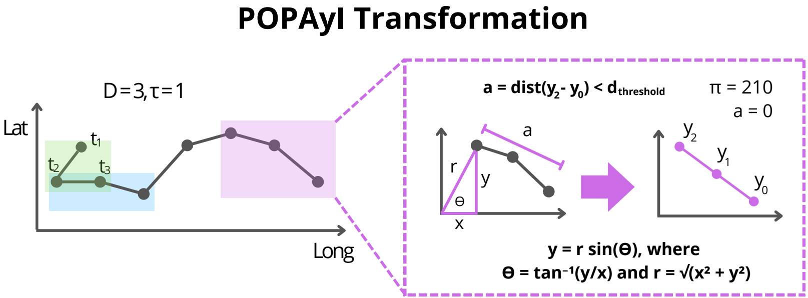
1. Absence of amplitude information
2. Originally defined for univariate time series

Ordinal Patterns Limitations

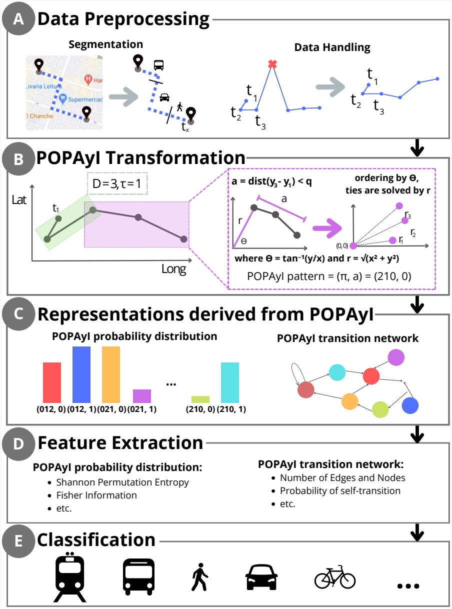


Our approach: PoPayl

POPAyI (Polar Ordinal Patterns with Amplitude Information) is the first work to leverage amplitude information in a multivariate OP approach.



Methodology



A **Data Preprocessing**

B **POPAyl Transformation**

C **Representations derived from POPAyl**

D **Feature Extraction**

POPAyl probability distribution:

- Shannon Permutation Entropy
- Fisher Information
- etc.

POPAyl transition network:

- Number of Edges and Nodes
- Probability of self-transition
- etc.

E **Classification**

Experiments

Goals:

ℹ Note

- Evaluate the effectiveness of POPAyI in classifying different transportation modes. - Compare POPAyI to state-of-the-art Machine Learning (ML) and Deep Learning (DL) methods.

Experiments

Setup:

ℹ Note

- We used GeoLife dataset, which contains 182 users' GPS trajectories over 5 years and various transportation modes: walking, biking, car, bus, subway, train.
- Fixed-length segments of 500 points to ensure comparability across all models
- Classification using cross-validation with 5 folds and stratified classes

Experiments

Hyperparameters selection

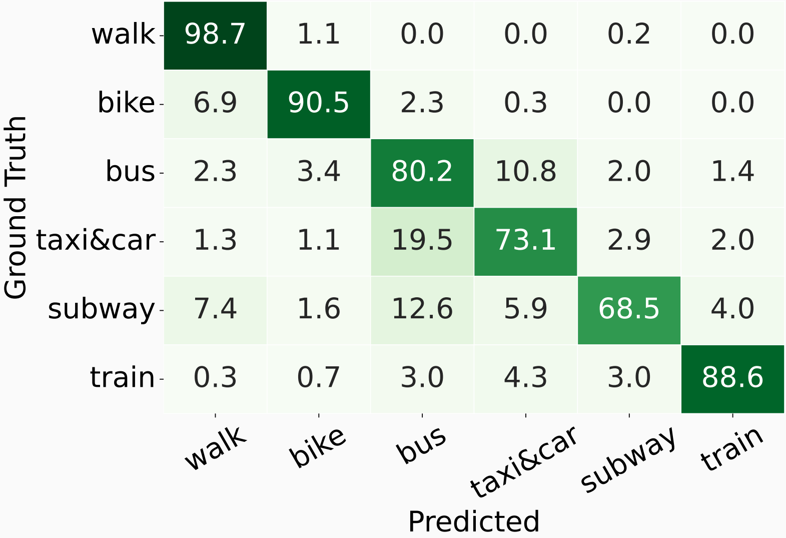
- Explored values:
    - Embedding Dimension: $D = 3, 4, 5$
    - Embedding Delay: $= 1, 2$
    - Amplitude Threshold: $q = 0.0005$ to 3 (0.5 m to 3 km)
- Selection Process:
    - Used Successive Halving: progressively eliminated weaker configurations based on F1-score
    - $\tau = 2$ consistently achieving the best results suggests that non-consecutive points capture more meaningful transport dynamics
    - Difficulty of establishing a straightforward correlation between hyperparameter values

Final choice: $D = 3$, $\tau = 2$, $q = 0.0005$

Experiments

TM sets	Methods	F1	D. Size	# feat.	# params.
walk, bike car&taxi, bus	DT	74.77%	10-8000	13	-
	POPAyI	86.34%	10-8000	19	7.8×10^1
	Light. CNN	87.10%	500	500	1.1×10^4
	POPAyI	86.86%	500	19	7.8×10^1
	POPAyI	87.41%	1000	19	7.8×10^1
walk, bike, car, train bus&taxi, subway	XGBoost	87.40%	10-39120	111	-
	POPAyI	82.40%	10-39120	19	7.8×10^1
walk, bike, car&taxi, bus, train	Best CNN	74.80%	200	200	2.6×10^6
	7 CNNs	83.90%	200	200	1.8×10^7
	LSTM	91.90%	200	200	8.1×10^6
	AE + CNN	88.28%	200	200	4.1×10^7
	Light. CNN	83.90%	500	500	1.1×10^4
	POPAyI	85.68%	500	19	7.8×10^1
	POPAyI	85.77%	1000	19	7.8×10^1
	POPAyI	86.71%	10-17000	19	7.8×10^1

Confusion Matrix



	walk	bike	bus	taxi&car	subway	train
walk	98.7	1.1	0.0	0.0	0.2	0.0
bike	6.9	90.5	2.3	0.3	0.0	0.0
bus	2.3	3.4	80.2	10.8	2.0	1.4
taxi&car	1.3	1.1	19.5	73.1	2.9	2.0
subway	7.4	1.6	12.6	5.9	68.5	4.0
train	0.3	0.7	3.0	4.3	3.0	88.6

Confusion Matrix

- Imbalanced classes might significantly impact classification
- Certain transportation modes may misclassify due to shared traffic characteristics and similar temporal dynamics
- Classifications are clearer between motor and non-motor transports due to distinct speed and distance behaviors
- Some misclassifications result from data quality issues
- Impact of False Positives/Negatives varies across ITS applications
  - Acceptable mixing of transport types in urban planning.
  - Critical applications require minimal false negatives for safety.
  - POPAyI Advantage: Fewer false negatives between motorized and non-motorized transports enhance safety in critical applications

Conclusions

- POPAyI demonstrates effective TMD by leveraging multivariate ordinal patterns and amplitude information
- Achieves high accuracy with minimal computational resources, making it suitable for resource-constrained environments (e.g., IoT)
- Outperforms traditional methods and DL approaches while maintaining a lightweight framework

Future Directions

- **Expand Dataset Applications:** Test POPAyI on more diverse mobility datasets to validate generability
- **Integrate with Smart City Infrastructure:** Explore the use of POPAyI in smart city projects for traffic management and urban planning
- **Address Misclassifications:** Investigate methods to further reduce misclassifications, particularly between similar transportation modes
- **Incorporate Additional Features:** Enrich the model by integrating more features from mobility data, such as environmental factors affecting transport dynamics