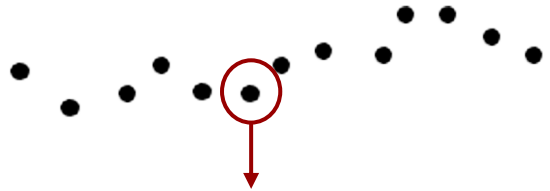
MAT-builder: a System to Build Semantically Enriched Trajectories

Chiara Pugliese - PhD student in Computer Science at University of Pisa
April 1st, 2022

chiara.pugliese@phd.unipi.it

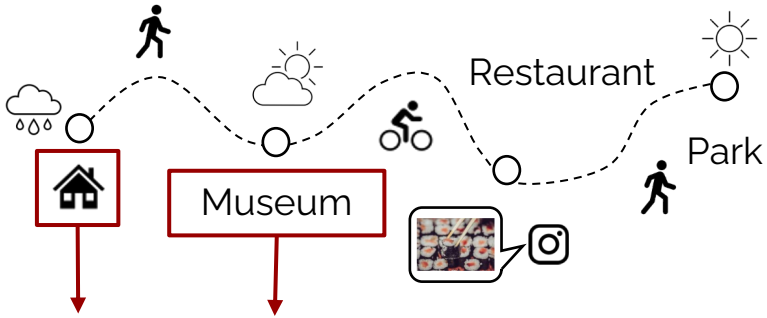
Raw Trajectory vs Multiple-Aspect Trajectory

Raw trajectory



latitude, *longitude*, *timestamp*

Multiple-Aspect Trajectory



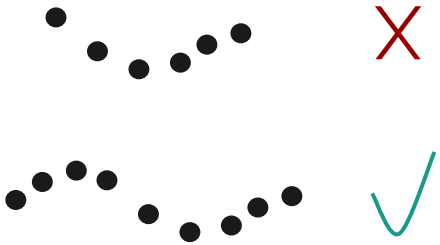
semantic aspects

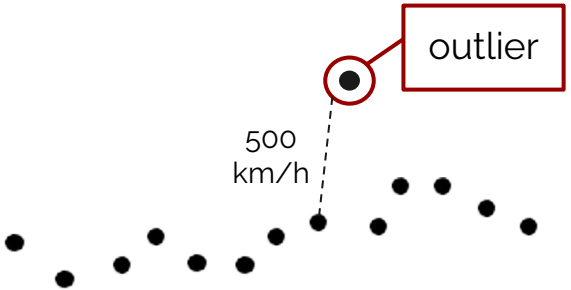
Semantic enrichment process

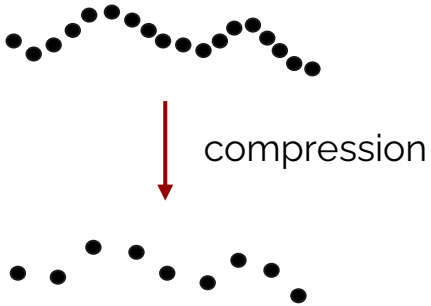
1. Trajectory preprocessing
2. Trajectory segmentation
3. Segment enrichment

Trajectory preprocessing

- **Discard** trajectories with # points < **threshold**
- Filter out **outliers**
- **Compression**

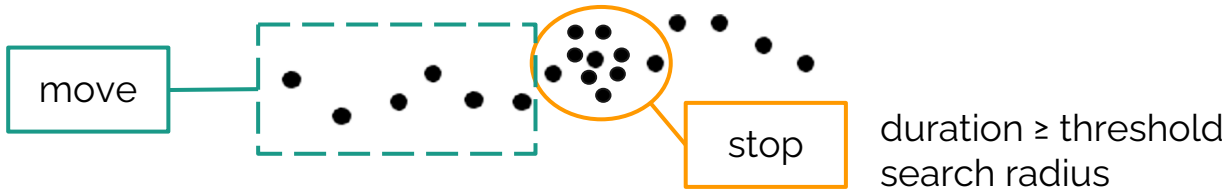






Trajectory segmentation

- We use **stop and move** method (... but there are a lot of methods)



Semantic enrichment

- We can choose the entity to enrich:
  - moving object
  - trajectory
  - segment
  - point
- In this case, we enrich segments: **stop** and **move**

Move enrichment

- Add new information to moves:
  - max **speed,** max **acceleration,** average **speed,** average **acceleration, total distance**
- **Transport means prediction**
  - train a classifier (in our case a Random-Forest) with **GeoLife**

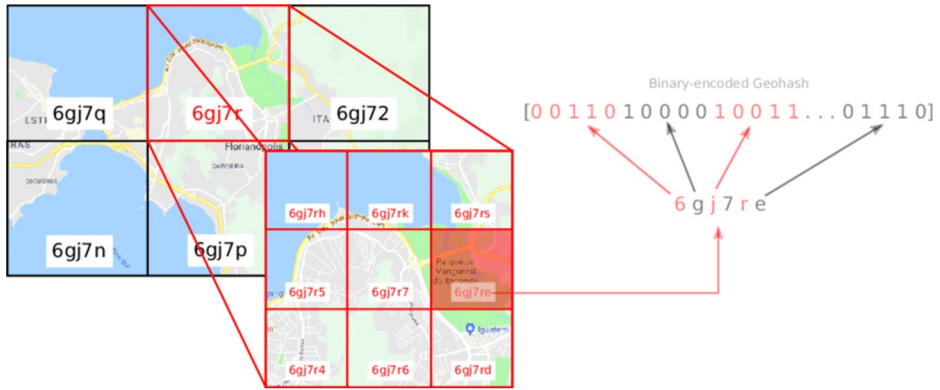
Stop enrichment

- **Systematic stops**

Stops falling within the same area more than a given number of times

- **Occasional stops**

Systematic stop detection



Geohash used to detect points falling in the same area

May Petry, Lucas & Leite da Silva, Camila & Esuli, Andrea & Renso, Chiara & Bogorny, Vania. (2020). **MARC: a robust method for multiple-aspect trajectory classification via space, time, and semantic embeddings.** International Journal of Geographical Information Science. 34. 1-23. 10.1080/13658816.2019.1707835.

Systematic stop enrichment

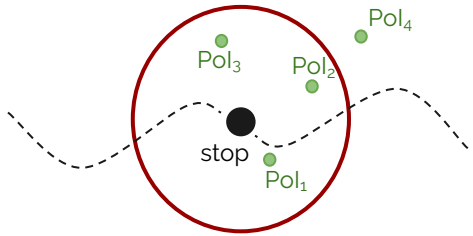
A systematic stop is labeled as:

- **home** (during evening and night)
- **work** (during morning and afternoon)
- **other**

based on the frequency during time slots

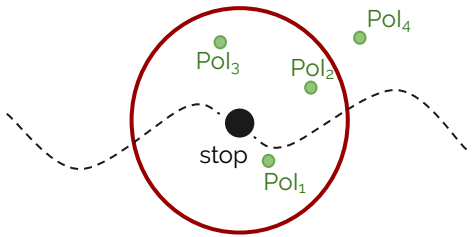
Occasional stop enrichment (I)

- Harder to characterize than **systematic** ones
- Associate a list of **Points of Interest (PoIs)**
  - **match** based on spatial features



Occasional stop enrichment (II)

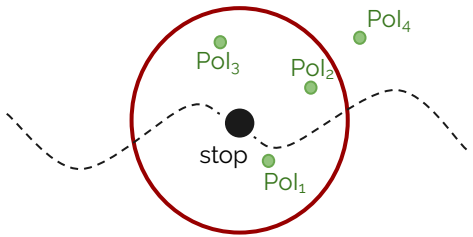
- **Rank** the PoIs matching with the stop, filtering them by some rules (e.g., opening hours)



stop (10 p.m)		
Point of Interest	**Distance**	**Opening Hours**
PoI_1	4 m	8:00-18:00
PoI_2	15 m	18:00-02:00
PoI_3	17 m	19:00-23:00

Occasional stop enrichment (II)

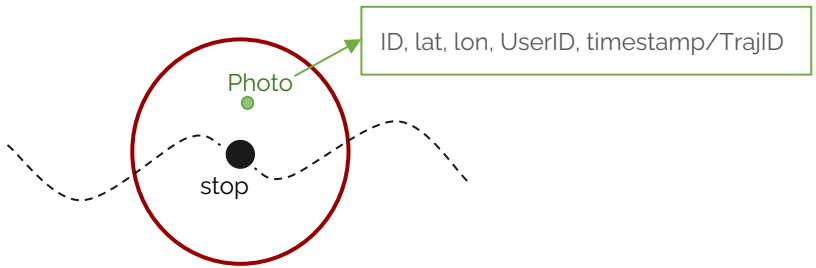
- **Rank** the PoIs matching with the stop, filtering them by some rules (e.g., opening hours)



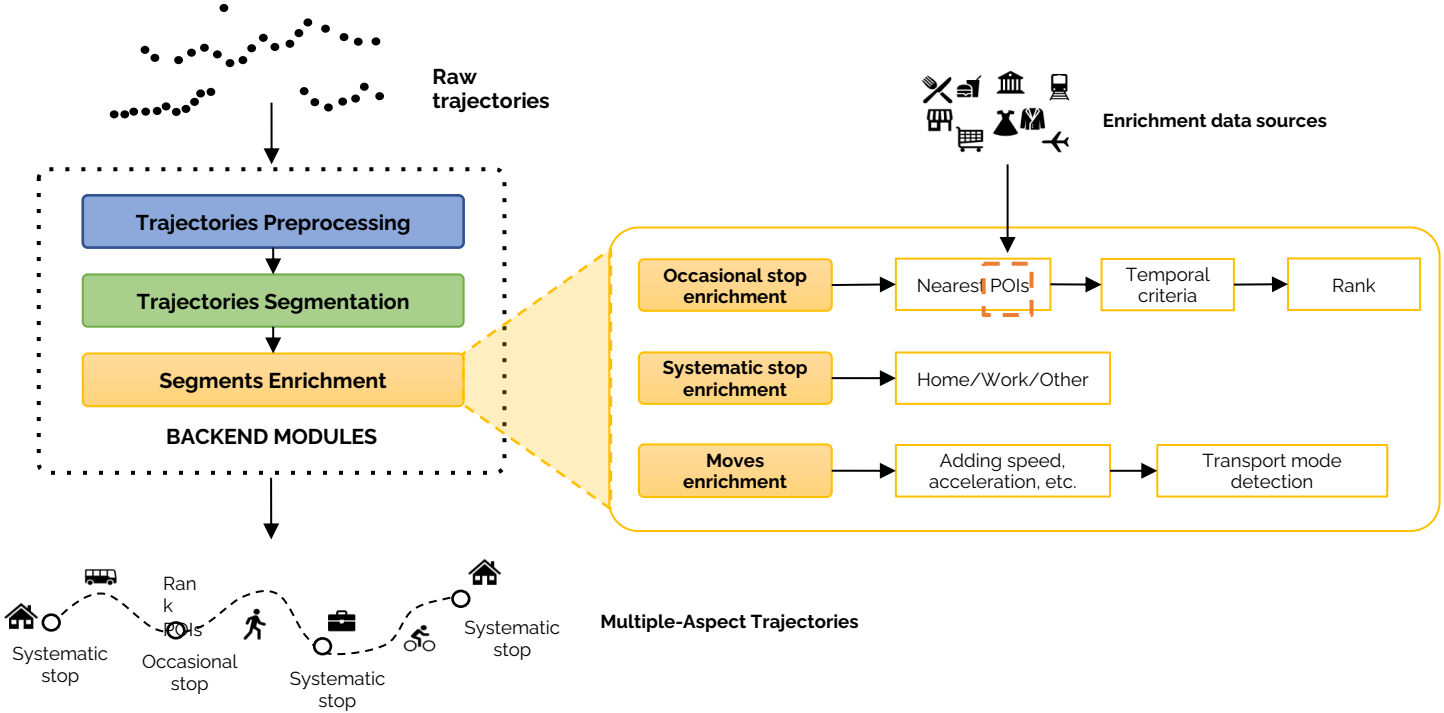
stop (10 p.m)		
Point of Interest	**Distance**	**Opening Hours**
~~Pol_1~~	~~4 m~~	~~8:00-18:00~~
Pol_2	15 m	18:00-02:00
Pol_3	17 m	19:00-23:00

Occasional stop enrichment (III)

We use the same method to enrich trajectories with other datasets (they need to have some specific attributes)



MAT-Builder


Raw trajectories
Trajectories Preprocessing
Trajectories Segmentation
Segments Enrichment
BACKEND MODULES
Enrichment data sources
Occasional stop enrichment
Nearest POIs
Temporal criteria
Rank
Systematic stop enrichment
Home/Work/Other
Moves enrichment
Adding speed, acceleration, etc.
Transport mode detection
Rank POIs
Systematic stop
Occasional stop
Systematic stop
Systematic stop
Multiple-Aspect Trajectories

MAT-Builder - functionalities (I)

1. Users can **customize** parameters based on their **needs** for each step
2. Users can **add** new modules and new methods to do preprocessing, segmentation, or enrichment

MAT-Builder - functionalities (II)

3. Users who do **not** have a **PoI dataset**, they could provide trajectories **bounding box** (PoIs downloaded from OpenStreetMap (OSM))
4. Users can provide the *semantic granularity* of data from OSM

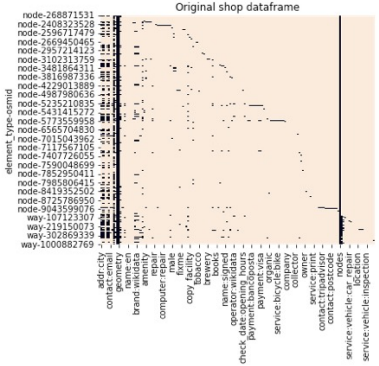
ID	**Name**	**Category**	**Opening hours**	**…**	**Website**
0	Da Martino	Restaurant	19-23	…	*missing value*
1	Al Gusto 129	Restaurant	*missing value*	…	*missing value*
2	Baronetto	Restaurant	*missing value*		*missing value*
…	…	…	…	…	…

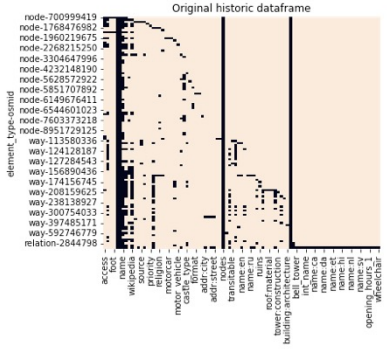
MAT-Builder - functionalities (II)

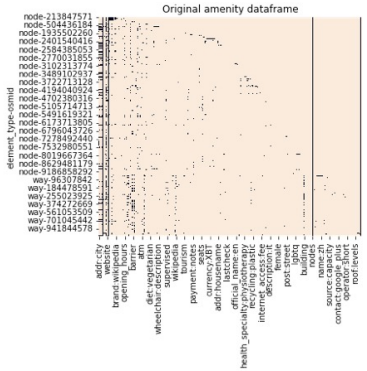
3. Users who do **not** have a **PoI dataset**, they could provide trajectories **bounding box** (PoIs downloaded from OpenStreetMap (OSM))
4. Users can provide the *semantic granularity* of data from OSM

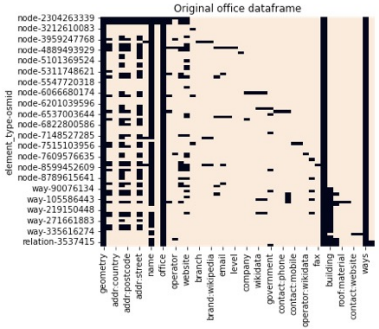
ID	**Name**	**Category**	**Opening hours**	**…**	**Website**
0	Da Martino	Restaurant	19-23	…	*missing value*
1	Al Gusto 129	Restaurant	*missing value*	…	*missing value*
2	Baronetto	Restaurant	*missing value*		*missing value*
…	…	…	…	…	…

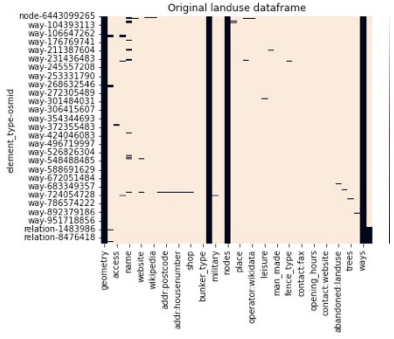
OSM – Pisa PoIs

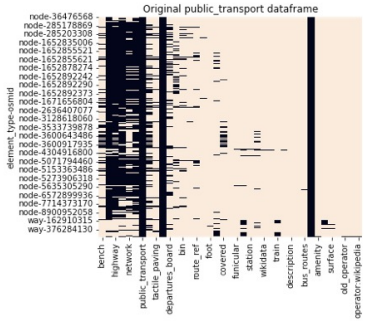






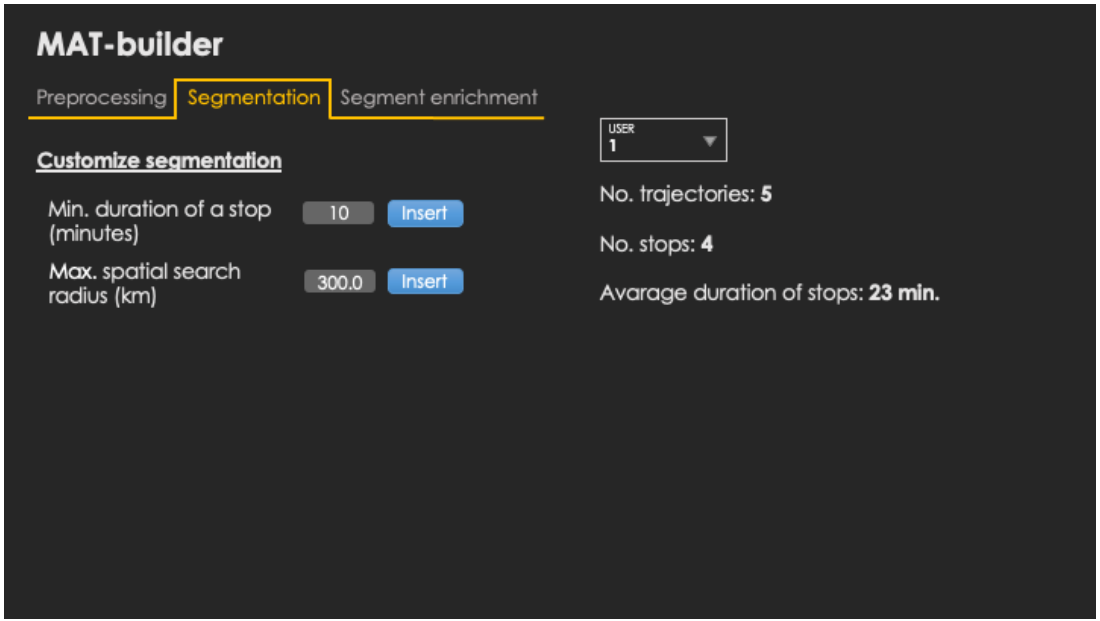




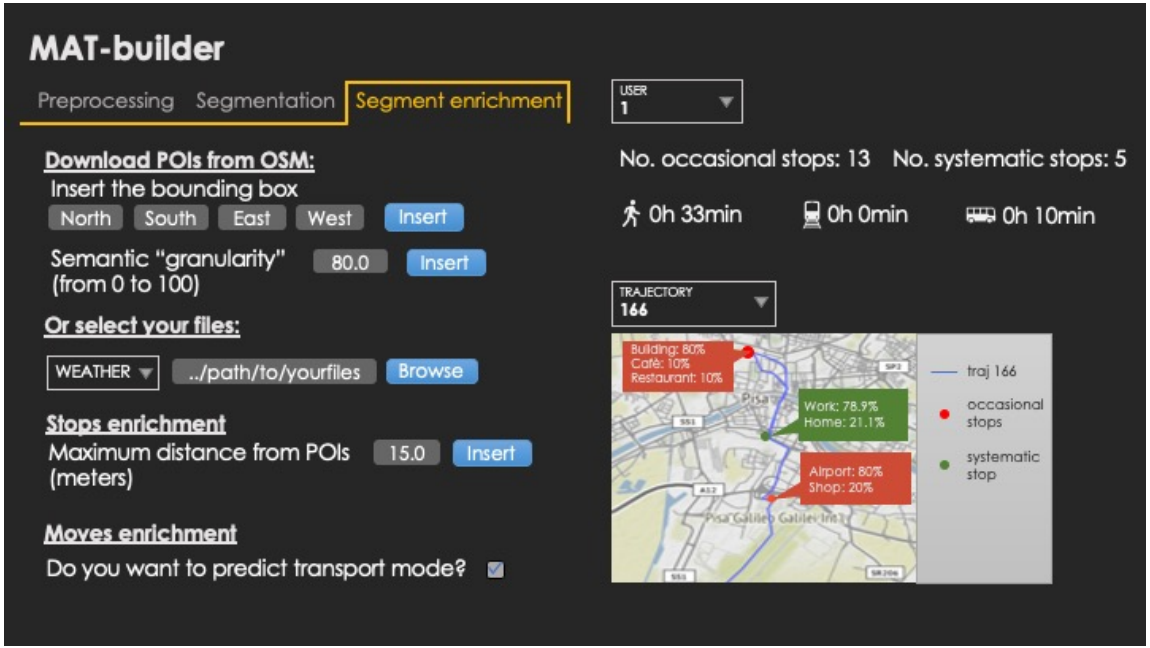


MAT-Builder – UI (I)

MAT-Builder – UI (II)



MAT-Builder – UI (III)



Application scenarios

- analysis of **transportation** usage to reduce pollution
- analysis of **movements** and **interests** of **tourists** for a more sustainable tourism
- **multivariate time series**

… now?!?

Multiple-Aspect Trajectories **characteristics**:

- high-dimensional
- variable-length
- multimodal attributes

Conclusion and future works

- MAT-Builder is the step *zero* because of the **lack** of **available datasets**
- how can we **improve** the **activity detection**?
- how can we **represent** multiple-aspect trajectories?
- how can we **capture** the **latent correlation** between **aspects**?
- how can we compute the multiple-aspect trajectories **similarity**?

THANK YOU :)

Chiara Pugliese
email: chiara.pugliese@phd.unipi.it