A general methodology for building multiple aspect trajectories

Francesco Lettich
ISTI-CNR, Pisa, Italy
francesco.lettich@isti.cnr.it

Chiara Pugliese
ISTI-CNR, Pisa, Italy
chiara.pugliese@isti.cnr.it

Chiara Renso
ISTI-CNR, Pisa, Italy
chiara.renso@isti.cnr.it

Fabio Pinelli
IMT Lucca, Lucca, Italy
fabio.pinelli@imtlucca.it

ABSTRACT

The massive use of personal location devices, the Internet of Mobile Things, and Location Based Social Networks, enables the collection of vast amounts of movement data. Such data can be enriched with several semantic dimensions (or *aspects*), i.e., contextual and heterogeneous information captured in the surrounding environment, leading to the creation of *multiple aspect trajectories* (MATs). In this work, we present how the MAT-Builder system can be used for the semantic enrichment processing of movement data while being agnostic to aspects and external semantic data sources. This is achieved by integrating MAT-Builder into a methodology which encompasses three design principles and a uniform representation formalism for enriched data based on the Resource Description Framework (RDF) format. An example scenario involving the generation and querying of a dataset of MATs gives a glimpse of the possibilities that our methodology can open up.

CCS CONCEPTS

• **Information systems** → **Geographic information systems**.

KEYWORDS

Trajectory enrichment, semantic enrichment, multiple aspect trajectory, knowledge graph

ACM Reference Format:
Francesco Lettich, Chiara Pugliese, Chiara Renso, and Fabio Pinelli. 2023. A general methodology for building multiple aspect trajectories. In *The 38th ACM/SIGAPP Symposium on Applied Computing (SAC '23), March 27-March 31, 2023, Tallinn, Estonia.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3555776.3577832

1 INTRODUCTION AND MOTIVATIONS

Movement data is nowadays continuously produced in large amounts thanks to the enormous diffusion of tracking technologies and devices that keep track of the location evolution of human beings, vehicles, and animals. We also observe an increasing availability of contextual and semantic data that is shared through various media such as social media, web APIs, semantic web technologies, and so on. This, in principle, enables practitioners to enrich movement data with a large amount of *aspects*, i.e., semantic dimensions that are heterogeneous and complex in nature. When synergically combined, movement data and aspects lead to the generation of multiple aspect trajectories [3] (MATs), which opens up the possibility to conceive innovative applications concerning the analysis of mobility behaviours. While there is an increasing interest related to the modelling and analysis of MATs [3, 6], we notice that there is no common established methodology that can guide users through semantic enrichment processes involving dynamic and heterogeneous aspects. Such problem becomes even more relevant if we consider that information on aspects must be typically retrieved from multiple *external semantic data sources* (i.e., the various media mentioned before). In general, existing approaches that implement semantic enrichment processes are monolithically tied to specific datasets, aspects, and applications, thus resulting unsuitable for different or more dynamic scenarios. This, in turn, causes a scarcity of MAT datasets available to the community, which affects the development of novel analysis methods.

In this work, we consider an extended version of the MAT-Builder system [5] within the context of a *general methodology* that aims to overcome the above limitations. More specifically, we extended MAT-Builder to adopt three design principles which we believe are necessary to provide systems capable of instantiating arbitrary semantic enrichment processes. We also extended the system to adopt the Resource Description Framework (RDF) formalism, coupled with a customization of the STEP ontology [4], to store datasets of MATs into knowledge graphs, as this enables to query and analyse enriched movement data in an *uniform* way.

The rest of the paper is organized as follows. Section 2 presents the extended MAT-Builder system. Section 3 provides an example scenario in which we generate a knowledge graph containing a dataset of MATs, and then query it to conduct analyses on the movement behaviours of an individual. Section 4 draws the conclusions.

2 THE MAT-BUILDER SYSTEM

In this section we introduce the extended version of the MAT-Builder system, whereby we focus on the novel contributions targeting the problems presented in the introduction.

The system. MAT-Builder[1] has been written in Python and is made up of two components, the *backend* and the *user interface* (UI). The **backend** constitutes the core component of our system, as it implements the MAT-Builder's processing engine according to *three* fundamental *design principles*. The backend is (1) *modular*,

[1]Source code available at: https://github.com/chiarap2/MAT_Builder

SAC '23, March 27-March 31, 2023, Tallinn, Estonia

ACM ISBN 978-1-4503-9517-5/23/03.
https://doi.org/10.1145/3555776.3577832

since it requires to have the *processing functionalities*, i.e., functionalities that concern the processing and enrichment of movement data, partitioned into modules, each addressing a specific task. The backend has also been designed to be easily (2) *extensible* which, coupled with the open source nature of MAT-Builder and Python's vast ecosystem of libraries, we believe will easily allow researchers and practitioners to contribute with novel or additional processing functionalities (e.g., management of additional aspects or external data sources). Finally, the backend has been designed to be (3) *configurable*, i.e., users can pick the desired modules from those available to the system and arrange them in a suitable order to instantiate their own semantic enrichment processes. The **user interface** (UI) exposes the functionalities of the modules making up a semantic enrichment process. More precisely, the UI enables the modules of a process to get input from and provide feedback to users, and dynamically adapts the content shown on screen to the needs of the modules making up any semantic enrichment process.

Representing datasets of MATs. Knowledge graphs [1] represent a natural choice for MATs, since they can support the representation of multitudes of aspects regardless of their heterogeneity and complexity. Moreover, knowledge graphs can be leveraged to conduct powerful analyses once they are imported in some triplestore of choice. The adoption of a schema that gives proper structure to the information is therefore one of the main problems when considering knowledge graphs for storage. In the context of our work, the choice fell on the STEPv2 ontology [4], which we lightly customized to suit our needs[2]. We refer to the original paper for the full details of the ontology. In the following, we summarily report on the main customizations we made. The first customization lets each instance of the *Agent* class (which in the ontology represents the concept of moving object) to be in relationship with instances of *FeatureOfInterest*: this allows to enrich also the moving objects. We also extended the *QualitativeDescription* class, which can be subclassed to support arbitrary aspects. More specifically, we added subclasses modelling the four aspects used in the semantic enrichment process considered in Section 3.

3 EXAMPLE SCENARIO

In this section, we illustrate an example scenario in which we first execute a semantic enrichment process to generate a RDF knowledge graph containing a dataset of MATs. The dataset is then imported in a triplestore and finally queried to show some interesting analyses to find out the mobility behaviours of an individual.

Background and semantic enrichment process execution. In the scenario, we consider a dataset of publicly available trajectories and a few external semantic data sources used to gather information concerning the aspects. The trajectory dataset covers the area of the province of Rome, Italy, and has been downloaded from OpenStreetMap (OSM). It contains 26395 trajectories from 3181 distinct users between March 2007 and July 2021. The semantic data sources are OSM, which we also used to retrieve 28787 POIs that may augment the individuals' stops, and Meteostat[3], which we used to obtain historical weather information. The social media

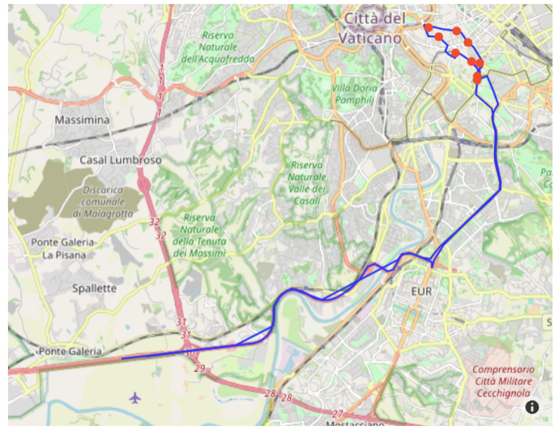

Figure 1: Plot of the considered MAT. The *red dots* represent the *stops*, while the *blue curve* represents the *moves*.

posts were generated synthetically and represent posts on Twitter. To produce a dataset of MATs, we instantiate the semantic enrichment process described in [2] and execute its modules via the MAT-Builder UI. The modules are executed in this order: trajectory pre-processing, trajectory segmentation, and enrichment. In the pre-processing module we set the minimum number of samples to *1500*, and the maximum speed threshold to *300 km/h*. This yields a set of 620 pre-processed trajectories from 575 users. In the segmentation module, we set the minimum duration of a stop to *10 minutes*, while the maximum spatial radius a stop can have is set to *0.5 km*. This yields a set of 666 stops and 1076 moves. Finally, the enrichment module enriches the segmented trajectories with the stops and their regularity, moves with transportation means estimation, weather information, and social media information. The stops have been augmented with POIs found to be less than *50 meters* far from the centroids of the stops. This yields a dataset of enriched MATs, which is then stored in a RDF knowledge graph according to the customized STEPv2 ontology. The dataset is finally imported in the GraphDB[4] triplestore, which is then used to perform some movement behavior analysis with the SPARQL 1.1 query language.

Movement behavior analysis. In the following, we aim to extract the mobility behaviours of an individual by analysing the MAT obtained by semantically enriching one of their trips. From the dataset of MATs, we selected an individual (ID 2115) who produced a trajectory (ID 2652) that originates close to the Fiumicino Rome Airport in the early morning, then spends half of the day within the centre of Rome, and then goes back to the same airport in the early afternoon (Figure 1). The overall duration of the trajectory is 6 hours. All such evidence might hint that the individual is some kind of tourist passing by the city. Accordingly, we want to find out (1) which transportation means the individual has likely used during their trip, (2) the POIs the individual may have visited while staying in Rome, and (3) the weather conditions and social media posts related to the trip.

The transportation means the individual has possibly used can be found by executing the SPARQL query below. The query first finds out the trajectory of interest (lines 3-4), then retrieves all its aspects

[2]The customized version is available in the MAT-Builder GitHub repository.
[3]https://meteostat.net/
[4]https://graphdb.ontotext.com/

(lines 5-8), and finally filters out those that are not a *move* (line 9). With the occurrences of the *move* aspect, the query first retrieves for

```
1  SELECT ?type_move ?t_start ?t_end (ofn:asMinutes(?t_end - ?t_start) AS ?duration_mins)
2  WHILE {
3     ?traj ^step:hasTrajectory / foaf:name "2115" ;
4              step:hasID "2652" ;
5              step:hasFeature ?feat .
6     ?feat step:hasEpisode ?ep .
7     ?ep step:hasSemanticDescription ?move ;
8           step:hasExtent ?ex .
9     ?move rdfs:subClassOf step:Move ;
10             rdf:type ?type_move .
11    ?ex step:hasStartingPoint / step:atTime / time:inXSDDateTime ?t_start .
12    ?ex step:hasEndingPoint / step:atTime / time:inXSDDateTime ?t_end .
13 } ORDER BY ASC(?t_start)
```

each occurrence the estimated transportation means (line 10), and then determines its starting and ending instants (lines 11-12). The SELECT finally returns a list of tuples, each representing a move with the estimated transportation means, its starting and ending instants, and its duration. The query finds 10 move occurrences from which the individual appears to go from the airport to the Rome's city centre by train, then mostly walked and partially used the bus while moving in the city, and finally went back to the airport by bus. Next, we want to find out the POIs the individual has possibly visited during their trip. Accordingly, the query below focuses on the individual's stops. The query first finds out the trajectory of

```
1  SELECT ?t_start ?t_end (ofn:asMinutes(?t_end - ?t_start) AS ?duration) ?poi_name
     ?poi_category
2  WHERE {
3     ?traj ^step:hasTrajectory / foaf:name "2115" ;
4              step:hasID "2652" ;
5              step:hasFeature ?feat .
6     ?feat step:hasEpisode ?ep .
7     ?ep step:hasSemanticDescription ?stop .
8     ?stop rdf:type step:Stop .
9     ?ep step:hasExtent / step:hasStartingPoint / step:atTime / time:inXSDDateTime ?t_start ;
10          step:hasExtent / step:hasEndingPoint / step:atTime / time:inXSDDateTime ?t_end .
11    ?stop step:hasPOI ?poi .
12    ?poi step:hasOSMCategory ?poi_category ;
13          step:hasOSMName ?poi_name .
14 } ORDER BY ASC(?t_start)
```

interest and keeps only the occurrences of *stops* (lines 3-8). For each of these occurrences, the query first finds the starting and ending instants (lines 9-10), then gathers information concerning the occurrences that have at least one POI, and finally retrieves the names and categories of the POIs involved (lines 11-13). From the results, we report that the query finds 12 stops. By looking at the associated POIs, we report that the individual appears to have spent a good part of the morning visiting various monuments: the individual briefly stayed in the area surrounding the *Palatino* and then went to the *Tempio della Pace*. The moving object then proceeded to stay for more than an hour in the area surrounding the *Altare della Patria*, and then stayed nearby the *Pantheon* for around half an hour until lunchtime. After that, the individual appears to have stayed at a restaurant for almost one hour. Finally, the individual appears to have stayed again in the vicinity of the *Palatino*, and then went back to the airport. All in all, the moving object has been repeatedly observed nearby famous monuments and appears to have walked and used public transportation means, thus reinforcing the initial impression they were indeed a tourist. In the final part of the running analysis, we aim to find out the weather conditions and the social media posts that the individual has respectively experienced and published during the trip. Let us focus on the weather conditions, as the strategy for the other aspect is the same. To this end, we can insert in either of the two queries shown before, right before the end of the WHILE loop, the SPARQL fragment below. The OPTIONAL keyword serves the

```
OPTIONAL {
   ?traj step:hasFeature / step:hasEpisode ?ep_w .
   ?ep_w step:hasSemanticDescription / rdf:type step:Weather ;
            step:hasWeatherCondition ?weather_conditions ;
            step:hasExtent / step:hasStartingPoint / step:atTime / time:inXSDDateTime
      ?tw_start ;
            step:hasExtent / step:hasEndingPoint / step:atTime / time:inXSDDateTime
      ?tw_end .
   FILTER((?t_start <= ?tw_end) && (?tw_start <= ?t_end)) }
```

purpose of not filtering out from the final results the occurrences of move (stop) for which no weather information is available. The FILTER keyword ensures that each move (stop) occurrence gets associated with an occurrence of the weather aspect only if they have a non-empty temporal overlap. Finally, the *weather_conditions* variable can be integrated into the SELECT to report the weather conditions. Overall, we observe that the individual experienced a sunny day during their trip.

4 CONCLUSIONS

This paper proposes a general methodology to instantiate arbitrary semantic enrichment processes leading to the generation of datasets of multiple aspect trajectories (MAT). In this context, we introduce an extended version of the MAT-Builder system that enriches movement data with dynamic and heterogeneous semantic dimensions from external semantic data sources and represents the enriched data with a uniform formalism. In the example scenario, we aim to show the potential of our methodology. Here, we first use MAT-Builder to execute a semantic enrichment process leading to a dataset of MATs stored into a RDF knowledge graph and then provide a glimpse of the kinds of powerful analyses that can be conducted on such datasets once imported into a triplestore.

Acknowledgements. This work has been supported by the EC H2020 projects MOBIDATALAB (GA 101006879), MASTER (GA 777695), and the scheme "INFRAIA-01-2018-2019 – Integrating Activities for Advanced Communities", Grant Agreement n.871042, "SoBigData++: European Integrated Infrastructure for Social Mining and Big Data Analytics" (http://www.sobigdata.eu).

REFERENCES

[1] L. Ehrlinger and W. Wöß. Towards a definition of knowledge graphs. In *Joint Proceedings of the Posters and Demos Track of SEMANTiCS2016 and SuCCESS'16*, volume 1695 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2016.

[2] A. Ibrahim, H. Zhang, S. Clinch, and S. Harper. From GPS to semantic data: how and why - a framework for enriching smartphone trajectories. *Computing*, 103, 12 2021.

[3] R. d. S. Mello, V. Bogorny, L. O. Alvares, L. H. Z. Santana, C. A. Ferrero, A. A. Frozza, G. A. Schreiner, and C. Renso. Master: A multiple aspect view on trajectories. *Transactions in GIS*, 23(4):805–822, 2019.

[4] T. P. Nogueira, R. B. Braga, C. T. de Oliveira, and H. Martin. Framestep: A framework for annotating semantic trajectories based on episodes. *Expert Systems with Applications*, 92:533–545, 2018.

[5] C. Pugliese, F. Lettich, C. Renso, and F. Pinelli. MAT-builder: a system to build semantically enriched trajectories. In *MDM 2022*, pages 274–277, 2022.

[6] C. Renso, V. Bogorny, K. Tserpes, S. Matwin, and J. A. F. de Macedo. Multiple-aspect analysis of semantic trajectories (master). *International Journal of Geographical Information Science*, 35(4):763–766, 2021.