Unified Approach to Trajectory Data Mining and Multi-Aspect Trajectory Analysis with MAT-Tools Framework

Tarlis Tortelli Portela[1], **Vanessa Lago Machado**[2], **Chiara Renso**[3]

[1]Instituto Federal do Paraná – Palmas, PR – Brazil

[2]Instituto Federal Sul-Rio-Grandense – Passo Fundo, RS – Brazil

[3]ISTI-CNR – Pisa – Italy

tarlis@tarlis.com.br, vanessalagomachado@gmail.com

chiara.renso@isti.cnr.it

Abstract. *Multiple-aspect trajectory (MAT) data mining requires sophisticated tools to handle the complexity and volume of complex data. This paper introduces MAT-Tools*[1]*, a comprehensive Python framework for MAT data mining. The framework consists of five main packages: mat-data, which supports data preprocessing and synthetic dataset generation; mat-model, offering model classes tailored for MAT data; mat-similarity, providing methods for similarity measurement; mat-view, visualization tools for MAT data, experimental preparation, and results exploration on a web interface; and mat-classification and mat-clustering, which includes advanced classification and clustering algorithms. Each package addresses specific challenges in MAT data analysis, from preprocessing to modeling and classification. MAT-Tools facilitates efficient and accurate trajectory data analysis, making it invaluable for diverse tasks since exploratory data analysis, anomaly detection, and predictive modeling applications. This framework's integration and extensibility empower researchers and practitioners to gain deeper insights and achieve more reliable results in trajectory data mining.*

1. Introduction

With the rapid increase in information availability and the proliferation of mobility devices, mobility data has become more complex and abundant with new data sources. Mobility data is a series of spatio-temporal information organized chronologically that records moving objects' movements, known as *moving object trajectories*. Trajectories are often linked to heterogeneous data sources leading to the concept of Multiple-Aspect Trajectories [Mello et al. 2019], where different semantic dimensions, called *aspects*, are associated with its trajectory parts.

Trajectory data mining involves extracting valuable patterns and insights from large movement datasets. Analyzing and representing mobility data has numerous real-world applications, including studying the movement of people, vehicles, hurricanes, drones, and animals [Leite da Silva et al. 2019]. Techniques used in this process often involve clustering, classification, and pattern recognition to identify common routes, predict future movements, detect anomalies, and enhance overall decision-making processes.

[1]Demonstration Video: https://youtu.be/HMymPh66Anc

The escalating complexity of trajectory data, with its high dimensionality and diverse enrichment sources, has created the need for a standard way to analyze this data in a single repository. To meet this need, MAT-Tools has been developed. MAT-Tools is a comprehensive set of tools designed to address the challenges associated with trajectory analysis and data mining. Related works include MAT-Builder [Lettich et al. 2023], a system for semantic enrichment processes that allow for the generation of synthetic multi-aspect trajectories, and pactus [Viera-López et al. 2023], a python framework for trajectory classification. The pactus framework aligns well with traditional trajectory classification methods, handling raw trajectories. While MAT-Builder focuses on generating synthetic MAT data and does not prioritize data analysis, pactus is focused solely on classification tasks of raw trajectories. In 2023, the Automatise framework [Portela et al. 2022a] was proposed, which involves data preprocessing, MAT classification, and visualization in web applications. However, this application is primarily focused on data preprocessing and classification, with the absence of other analysis tasks

In this way, the MAT-Tools framework is an innovative Python framework for the analysis and mining of MATs. It provides a range of analysis tools and visualization options, created to simplify the process of preparing, analyzing, and interpreting multidimensional trajectory data. These tools are particularly useful for researchers and developers in trajectory analysis, machine learning, and data science.

Besides this introduction, the text is organized into two more sections. Section 2 provides the framework overview and subsections detail package descriptions. Section 3 provides the conclusion and future works.

2. Framework Overview

The Multiple-Aspect Trajectory Data Mining Framework introduces a platform designed to address the challenges posed by the increasing complexity of trajectory data, which is now high-dimensional and associated with heterogeneous semantic information. This platform, named *MAT-tools*[2], offers a set of Python libraries to support trajectory and multidimensional data mining tasks, in an open-source mode. The framework is designed to be scalable, modular, and user-friendly. It provides a set of tools and algorithms for preprocessing, analyzing (such as measuring similarity), and classifying trajectory data. Figure 1 presents an overview of the framework structure.

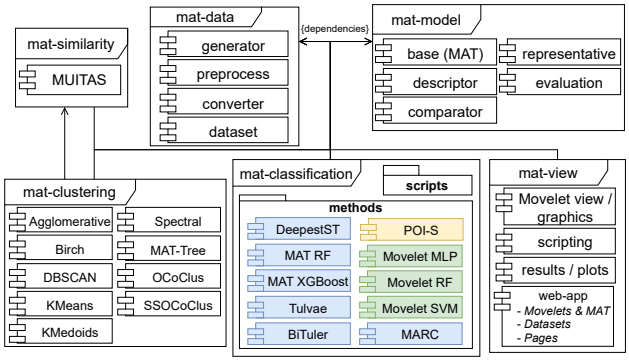


Figure 1. Framework structure.

[2]Sources: `https://github.com/mat-analysis/mat-tools`

The platform offers various visualization options and hides intricate coding details to expedite the creation of interactive visualizations essential for data analysts. The framework is designed to easily be extended to accommodate other data types, feature visualizations, and new methods. Following is a short description of the available packages.

- *MAT-Data*: package to support the user in the task of data preprocessing of multiple-aspect trajectories, converting between different data formats, and generating synthetic datasets;
- *MAT-Model*: library of model classes to the representation of MAT. Provides data description structures and evaluation functions as distance metrics for complex data types of MAT;
- *MAT-Similarity*: focused on similarity measurement, this package includes methods and functions specifically designed for evaluating the similarity of trajectory data. It supports a range of similarity metrics essential for clustering, classification, and anomaly detection tasks;
- *MAT-Classification*: designed for the analysis and classification of MAT. The package integrates various classification methods tailored for multidimensional sequence data mining, including movelet-based, feature-based, and trajectory-based classifiers;
- *MAT-Clustering*: provides similarity-based and MAT clustering algorithms;
- *MAT-View*: a set of web tools to support the user in visualizing MATs and movelets. Also, it presents a detailed list of available datasets on a public repository, descriptions of MAT data mining methods, experimental results exploration, and environmental and shell script generators.

2.1. *MAT-Data*: Trajectory Data Preprocessing

The Trajectory Data Preprocessing Package (*MAT-Data*) is designed to simplify the handling, transformation, and manipulation of trajectory datasets. The package includes a variety of functions that facilitate the conversion between different data formats, dataset splitting for k-fold cross-validation experiments, random sampling and scaled data generation, dataset statistics, and more. Furthermore, this project provides access to a repository of publicly available preprocessed datasets[3] specifically curated for research experiments.

Key features include (i) Data conversion methods to convert datasets between various formats such as CSV, Parquet, and ZIP; (ii) Dataset splitting and stratification for transforming datasets into train and test sets using holdout and k-fold methods ensuring balanced class distributions; (iii) Data loading and reading from various sources including public repository (e.g. Brightkite, Gowalla, Animals); (iv) Synthetic trajectory data generation based on real or random data; (v) Dataset organization and statistics to standardize trajectory datasets, sorting, computing statistics, and utility functions.

2.2. *MAT-Model*: Multiple-Aspect Trajectory Model

The *MAT-Model* (Multiple-Aspect Trajectory Model) is a package designed for handling complex trajectory data with multiple aspects. This package supports the creation, manipulation, and modeling of multidimensional trajectory data for classification, summarization, similarity measuring, and clustering. This package provides a structured framework

[3]Datasets Repository: `https://github.com/mat-analysis/datasets`.

for managing multidimensional trajectory data, offering tools for data manipulation, comparisons, feature extraction, and evaluation.

The package enables representations of aspects, such as spatial, temporal, and semantic dimensions. Also, provides a variety of distance functions. Key features include support for various distance metrics, aspect-based data handling, and integration with the other *MAT-Tools* framework packages.

2.3. *MAT-Similarity*: Similarity Measures for Multiple-Aspect Trajectory

Trajectory data analysis enables the extraction of valuable insights such as path discovery, pattern recognition, and mobility prediction. Central to these analyses are similarity measures, which help identify common movement patterns and compare trajectories based on spatial, temporal, and semantic aspects. Various similarity measures have been developed, each tailored to different aspects of trajectory data.

Focused on similarity measurement, *MAT-Similarity* includes methods and functions specifically designed to evaluate the similarity of MAT data. The first release supports the MUITAS similarity metric, a state-of-the-art measure essential for clustering, classification, and mining tasks [Petry et al. 2019]. *MAT-Similarity* offers several key features, such as similarity metrics and its integration with Other MAT-Tools Packages. By providing a comprehensive and flexible toolkit, *MAT-Similarity* is crucial for accurately measuring similarity in MATs, supporting a wide range of applications from exploratory data analysis.

2.4. *MAT-Classification*: Multiple-Aspect Trajectory Classification Methods

The *MAT-Classification* is an advanced toolkit developed to facilitate the classification of MAT. Moreover, it is compatible with any multidimensional sequential data. The package incorporates a variety of publicly available classifiers from the literature, each tailored to different aspects of trajectory data, ensuring robust and flexible data analysis capabilities.

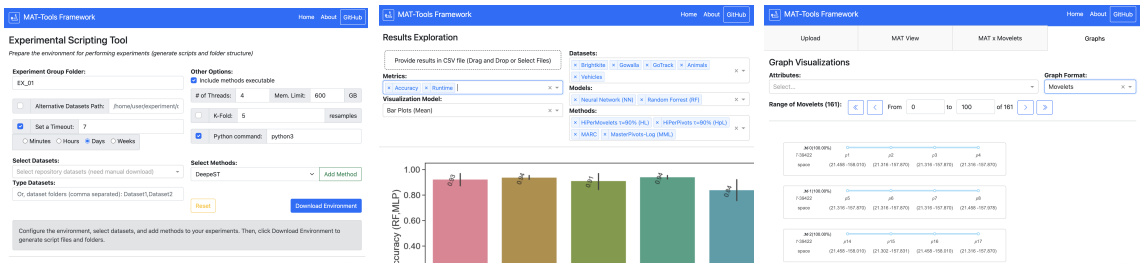
The package makes available MAT classifiers using three different approaches: movelet-based for classifiers that receive the movelets (the discriminant subtrajectories of each class label), trajectory-based for MAT as input, and feature-based for other features. *Movelet-based* approaches, as implemented in [Ferrero et al. 2020], are: MMLP (Movelet Multilayer-Perceptron), MRF (Movelet Random Forest), MRFHP (Movelet Random Forests with varying parameters), MDT (Movelet Decision Tree), and MSVN (Movelet Support Vector Machine). Additionally, there are movelet extraction methods such as HiPerMovelets (high-performance movelet extraction) [Portela et al. 2022b] and UltraMovelets (Efficient Movelet Extraction) [Portela et al. 2024], both designed for MAT Classification. *Trajectory-based* approaches are MARC [Petry et al. 2020] (trajectory embedding based neural network), TRF (Trajectory Random Forest), TXGBost (Trajectory XGBost), Tulvae (recurrent neural network with embedding sequence of POIs), BiTuler (enhancement of Tulvae), and DeepeST (deep learning for sub-trajectory classification, as implemented in [de Freitas et al. 2021]. *Feature-based* approach: POI-S [Vicenzi et al. 2020] (an extension of the POI-Frequency method that employs a TF-IDF approach for feature extraction).

2.5. *MAT-Clustering*: Multiple-Aspect Trajectory Clustering Methods

MAT-Clustering is created to unify traditional clustering algorithms for MATs while introducing innovative methods within a single framework. This package streamlines the process of clustering trajectory data by incorporating well-known algorithms like K-Means, DBSCAN, Agglomerative, K-Medoids, and Spectral clustering[4] enhances them with the capabilities of the similarity measures for MAT.

2.6. *MAT-View*: Web Interface for Analysts

MAT-View provides MAT analysis tools. It aims to simplify the process of running experiments, reading results, computing statistics, and visualizing trajectories with its movelets. Figure 2 presents four screens from *MAT-View* web interface. **Scripting**: the scripting module provides functions to generate command-line scripts for running data mining methods (Figure 2a). Available methods to generate scripts include classification methods. However, its structure can be easily extended with any other methods; **Results**: offers functions to read and compile statistics from experimental result files with any provided metrics (Figure 2b). The web interface can read compiled results for plotting a variety of graphs. Additionally, this functionality can be easily extended to accommodate other visualizations; **MATs and Movelets Visualization**: a tool that provides visualization schemes for movelets as multidimensional trajectory and subtrajectory (Figure 2c). Additionally, provides movelets representaions as Sankey diagrams[5], Markov chain[6], a tree visualization and as plain multiple-aspect sequence.



(a) Scripting tool. (b) Results exploration. (c) MAT and Movelets.

Figure 2. *MAT-View* web interfaces.

3. Conclusion and Future Works

The development of *MAT-Tools* represents a significant advancement in trajectory data mining, offering a robust platform for MAT analysis. Providing a user-friendly interface and versatile Python libraries simplifies the analyst work, from creating the experimental environment to visualizing the results, thereby enhancing the efficiency and effectiveness of data analysis tasks. Stands as a valuable tool for researchers and data analysts working with complex multidimensional datasets, paving the way for further advancements in trajectory data mining. Looking ahead, the potential for future enhancements includes implementing new similarity measures, classification, clustering methods, and the packaging of available summarization methods for MAT.

[4] https://scikit-learn.org/stable/modules/clustering.html
[5] https://en.wikipedia.org/wiki/Sankey_diagram
[6] https://setosa.io/ev/markov-chains/

References

[de Freitas et al. 2021] de Freitas, N. A., da Silva, T. C., de Macêdo, J. F., Junior, L. M., and Cordeiro, M. (2021). Using deep learning for trajectory classification. In *Proceedings of the 13th International Conference on Agents and Artificial Intelligence - Volume 2: ICAART*, pages 664–671. INSTICC, SciTePress.

[Ferrero et al. 2020] Ferrero, C. A., Petry, L. M., Alvares, L. O., Leite da Silva, C., Zalewski, W., and Bogorny, V. (2020). MasterMovelets: discovering heterogeneous movelets for multiple aspect trajectory classification. *Data Mining and Knowledge Discovery*, 34(3):652–680.

[Leite da Silva et al. 2019] Leite da Silva, C., May Petry, L., and Bogorny, V. (2019). A survey and comparison of trajectory classification methods. In *2019 8th Brazilian Conference on Intelligent Systems (BRACIS)*, pages 788–793.

[Lettich et al. 2023] Lettich, F., Pugliese, C., Renso, C., and Pinelli, F. (2023). Semantic enrichment of mobility data: A comprehensive methodology and the MAT-BUILDER system. *IEEE Access*, 11:90857–90875.

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

[Petry et al. 2019] Petry, L. M., Ferrero, C. A., Alvares, L. O., Renso, C., and Bogorny, V. (2019). Towards semantic-aware multiple-aspect trajectory similarity measuring. *Transactions in GIS*, 23(5):960–975.

[Petry et al. 2020] Petry, L. M., Leite da Silva, C., Esuli, A., Renso, C., and Bogorny, V. (2020). MARC: a robust method for multiple-aspect trajectory classification via space, time, and semantic embeddings. *International Journal of Geographical Information Science*, 34(7):1428–1450.

[Portela et al. 2022a] Portela, T. T., Bogorny, V., Bernasconi, A., and Renso, C. (2022a). Automatise: Multiple aspect trajectory data mining tool library. In *2022 23rd IEEE International Conference on Mobile Data Management (MDM)*, pages 282–285.

[Portela et al. 2022b] Portela, T. T., Carvalho, J. T., and Bogorny, V. (2022b). Hipermovelets: high-performance movelet extraction for trajectory classification. *International Journal of Geographical Information Science*, 36(5):1012–1036.

[Portela et al. 2024] Portela, T. T., Machado, V. L., Carvalho, J. T., Bogorny, V., Bernasconi, A., and Renso, C. (2024). Ultramovelets: Efficient movelet extraction for multiple aspect trajectory classification. In *Database and Expert Systems Applications (DEXA)*, pages 79–94, Cham. Springer Nature Switzerland.

[Vicenzi et al. 2020] Vicenzi, F., Petry, L. M., Silva, C. L. D., Alvares, L. O., and Bogorny, V. (2020). Exploring frequency-based approaches for efficient trajectory classification. *Proceedings of the ACM Symposium on Applied Computing*, pages 624–631.

[Viera-López et al. 2023] Viera-López, G., Morgado-Vega, J., Reyes, A., Altshuler, E., Almeida-Cruz, Y., and Manganini, G. (2023). pactus: A python framework for trajectory classification. *Journal of Open Source Software*, 8(89):5738.