Measuring Similarity between Groups of Trajectory Data with Multiple Aspects

Vanessa Lago Machado[1]**, Tarlis Tortelli Portela**[2]**,**
Lucas Vanini[1]**, Ronaldo dos Santos Mello**[3]

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

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

[3]Universidade Federal de Santa Catarina (UFSC)
Florianópolis – SC – Brasil

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

lucasvanini@ifsul.edu.br, r.mello@ufsc.br

Abstract. *This paper proposes an approach to measure similarity between groups of trajectory data using representative trajectories. By summarizing each group into a representative trajectory and comparing these, we address the challenges in group trajectory analysis. This offers a versatile solution for understanding group behaviors and interactions. Our approach is demonstrated using the Foursquare NYC dataset, which shows its potential in social behavior analysis and highlights its diverse applications, such as urban planning, transportation optimization, and animal migration analysis. The results show that our approach provides meaningful insights into group trajectory patterns, significantly advancing the field of trajectory data analysis.*

1. Introduction

The trajectory data analysis has become increasingly significant across urban planning, transportation, and social behavior analysis. As technology advances, the volume and complexity of trajectory data have grown, necessitating sophisticated methods to derive meaningful insights. Traditionally, trajectory analysis has focused on individual trajectories, but there is a growing need to understand group movement patterns to gain deeper insights into collective behaviors and interactions [Wiratma et al. 2017].

A trajectory refers to the sequence of movements an object follows through space over time that can be enriched with contextual information to form *multiple-aspect trajectories (MATs)*. While trajectory data mining has yielded valuable insights from individual movement patterns, analyzing group patterns remains challenging [Alowayr et al. 2021, Gupta et al. 2013]. Group pattern mining seeks to understand the movement of objects traveling together, essential for applications like optimizing transportation routes, studying animal behaviors, and designing efficient urban spaces.

One key challenge in trajectory analysis is measuring similarity among groups of trajectories, especially when considering spatial, temporal, and semantic attributes. Existing approaches often fail to consider multiple aspects, limiting their real-world applicability [Wiratma et al. 2017]. Understanding how groups of objects move together and

interact is essential for applications such as vehicle convoys, studying animal migratory patterns, and analyzing social interactions in urban spaces.

This paper addresses this gap by proposing an approach to measure similarity between groups of trajectory data using representative trajectories. By summarizing each group into a representative trajectory and comparing these, we offer a versatile solution for analyzing group interactions and identifying similarities between groups of trajectories. Our approach effectively captures the essential characteristics of group movements, providing meaningful insights into group trajectory patterns.

2. Background and Related Work

The concept of trajectory has evolved significantly over time, from *raw trajectory*, sequential movements [Erwig et al. 1999], to *semantic trajectory* enriched with contextual data [Parent et al. 2013]. The Internet of Things (IoT) and social media have further expanded this to multiple aspect trajectories (MATs), integrating spatial, temporal, and various heterogeneous semantic aspects [Mello et al. 2019], such as the type of movement, purpose, and context of the travel.

Trajectory data analysis enables the extraction of valuable insights, such as path discovery, pattern recognition, and mobility prediction. Critical to these analyses are similarity measures, which help to identify common movement patterns and compare two trajectories based on spatial, temporal, and semantic aspects [Machado et al. 2023b]. Various similarity measures have been developed, each tailored to different aspects of trajectory data. For raw trajectories, measures like the *Discrete Fréchet* distance (DF) [Eiter and Mannila 1994] and *Uncertain Movement Similarity (UMS)* [Furtado et al. 2018] focus on spatial and temporal dimensions. For semantic trajectories, measures such as *Multidimensional Similarity Measure (MSM)* [Furtado et al. 2016], and *Multiple aspect trajectory similarity (MUITAS)* [Petry et al. 2019] provide a holistic view by integrating multiple aspects. Different similarity measures focus on distinct aspects, and the choice of which to use depends on the intended purpose of the analysis.

Analyzing groups of trajectories involving sets of objects traveling together is crucial for applications like animal behavior studies and vehicular traffic monitoring [Wiratma et al. 2017]. Several methods have been proposed to identify and analyze group movements, including concepts like *flocks*, *clusters*, and *convoys*. A flock is defined as a set of objects traveling together for a certain duration, while a cluster identifies trajectories that overlap significantly over time [Kalnis et al. 2005]. Convoys use density-connected spatial clustering to define groups [Jeung et al. 2008]. These methods offer various approaches to identifying and analyzing groups within trajectory data, each with its own powers and limitations.

Group pattern mining focuses on understanding collective movement behaviors [Wiratma et al. 2017]. Summarizing trajectories can offer a more compact representation that maintains crucial information. However, addressing the challenge of effectively summarizing trajectories while minimizing data loss is important. Representative data can characterize a group of trajectories, facilitating tasks of trajectory analysis [Machado et al. 2023a]. Additionally, identifying similarities between different groups is important to understand these collective movement behav-

iors [Machado et al. 2023b, Wiratma et al. 2017], providing insights into group interactions and dynamics, which are valuable for various domains, including ecology, where it can provide valuable insights into the interactions between animals [Li et al. 2013].

According to Wiratma et al. (2017), most measures for single trajectories can be directly extended to groups, focusing on three main types: *similarity*, *closeness*, and *centrality* with raw trajectories focused. Similarity measures compare groups by averaging the similarity of individual trajectories or by using a many-to-many matching approach. Closeness measures assess the proximity of groups over time, and centrality measures evaluate a group's importance relative to neighboring groups, providing insights into group dynamics and interactions. For these measures, the authors assume that all groups start and end simultaneously, allowing insights like the average number of other groups that exist during the lifetime of a specific group.

Despite progress in extracting movement patterns and identifying close-knit groups, understanding group interactions remains an open issue [Gupta et al. 2013]. Leveraging representative data to comprehend patterns within MATs presents a robust solution to the challenges posed by the volume and complexity of trajectory data. However, it is essential to align computation methods with the specific objectives and requirements of the intended use case.

3. Methodolody

To address the gap in measuring similarity between groups of trajectories, we propose an approach that uses Representative Trajectories (RT). Our methodology comprises four main steps: (i) data collection and preprocessing, (ii) representative trajectory computation, (iii) similarity measure calculation, and (iv) evaluation, as illustrated in Figure 1.
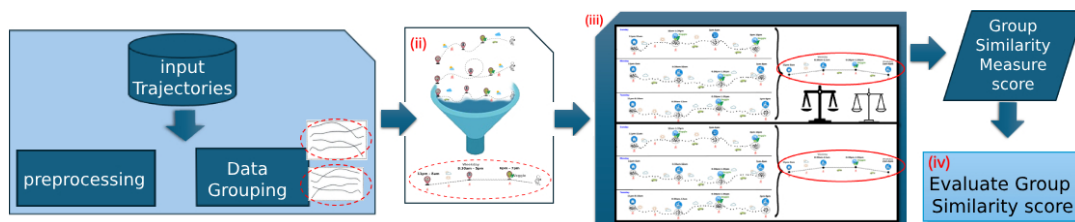

Figure 1. Group trajectory similarity proposed

The first step involves collecting trajectory data from various sources, such as GPS traces, animal movement data, and social media check-ins. After collection, the data are preprocessed to ensure consistency and meaningful segmentation. This involves cleaning the data to remove errors caused by GPS inaccuracies or missing entries and normalizing them, standardizing spatial and temporal dimensions. Additionally, groups of trajectories are defined using algorithms or predefined criteria to identify sets of trajectories that travel together.

In the second step, a RT is computed for each group, summarizing the trajectories into each. This can be done using methods such as *centroid* trajectory, *medoid* trajectory, or *specific methods* that compute RT's like those proposed in [Machado et al. 2023a].

In the third step, the similarity measure between RT's is computed using similarity measures to estimate their similarity. Examples include the DF distance and UMS for

raw trajectories. The MSM could be used for semantic trajectories, while the MUITAS measure could be employed for MATs. These measures provide a similarity score that reflects the closeness between different groups of trajectories.

Our proposed approach offers a robust solution for measuring the similarity between groups of trajectories. By addressing spatial, temporal, and semantic dimensions, our approach provides valuable insights into movement patterns and group interactions, including MATs. This will help in dealing with complex data and understanding when different groups are close to one another, applicable across various domains such as urban planning, transportation optimization, and social behavior analysis.

4. Preliminary Experimental

Dataset. We used the Foursquare NYC dataset, which includes check-in records from April 2012 to February 2013 in New York City. This dataset is enriched with contextual information such as *weekday*, *category*, *price*, *rating* of the POIs, and *weather conditions*, comprising 3079 MATs from 193 users.

Experimental Setup. To compute the similarity measure between groups of trajectories, we defined the following elements: (i) grouping criterion, where each defines the criterion for grouping trajectories as the ground truth; (ii) representative trajectory computation, we used the MAT-SGT method, which is designed specifically for MATs; and, (iii) similarity measure, we employed the state-of-the-art MAT similarity measure, MUITAS, to establish trajectory similarity.

For the case study, we use the proposed technique to identify the nearest user pattern for a determined user. We define as a sample the top-10 RT's identified for the Foursquare dataset, along with the user and parameter configuration of each computed RT. This rank was presented by the RMMAT measure work, a representativeness measure that computes the representativeness score of a RT regarding its original trajectories [Machado 2024]. This rank helps identify users who follow high representativeness by RMMAT score [Machado et al. 2023b].

MAT-SGT uses two thresholds, τ_{rv} and τ_{rc}, representing a rate of representativeness value for ranking values in RT and a rate of a minimum proportion of all input MAT points to define if a cell is considered relevant in RT computation. We used the same parameter configuration defined in the top 10 to ensure reproducibility.

To compute the similarity measure between RT's using MUITAS, some proximity functions could be defined to assess the similarity across each dimension: *(i) spatial*, using Euclidean distance measure; *(ii) temporal*, using a match function based on interval overlap; and *(iii) semantic*, evaluating attribute matching for *numeric* types ($\leq 10\%$ of the maximum value) and *categorical* types (equality or overlap). We consider $w = 1/3$ for each dimension to balance them, ensuring that no single dimension (spatial, temporal, or semantic) dominates the similarity calculation, leading to a more holistic comparison between trajectory groups. However, this paper presents an approach to measuring the similarity between the groups. Then, the parameter configuration should be defined by the analyst according to the need in each case.

Results.   Table 1 presents the top 10 RT identified in the Foursquare dataset, along with the user and parameter configuration of each computed RT. For each user, we provide information on the two nearest users who follow similar patterns, including the user and the computed similarity. Notably, in all computations, the most similar user refers to the self-user with a similarity score 1.0. Therefore, the next users in the list are considered the most similar to each one.

Table 1. The top 10 computed RT in Foursquare dataset *Vs* the more similar user to each one

user	τ_{rv}	τ_{rc}	RMMAT	1st Nearest User		2nd Nearest User	
				User	Similarity	User	Similarity
730	0.10	0.05	0.94	1054	0.698	842	0.694
895	0.10	0.05	0.87	916	0.714	65	0.700
207	0.05	0.10	0.87	99	0.788	73	0.784
754	0.05	0.25	0.87	891	0.718	635	0.705
365	0.05	0.05	0.84	951	0.720	702	0.705
647	0.05	0.25	0.84	754	0.690	857	0.685
69	0.10	0.05	0.82	718	0.739	371	0.731
440	0.05	0.05	0.81	916	0.736	484	0.729
438	0.05	0.10	0.80	889	0.702	787	0.700
673	0.10	0.05	0.80	432	0.756	834	0.742

Discussion.   The results showed the most similar users and their similarity scores, with an average of 70% similarity in patterns (similarity scores hover around 0.700), demonstrating the effectiveness of our proposed approach in identifying similar group movement patterns. However, the study noted a limitation in the analysis, as MUITAS did not consider the temporal sequence between the points in the MATs, which may affect the accuracy of the representativeness in certain scenarios.

5. Comparative Analysis

While trajectory analysis has been the subject of extensive research, the challenge of effectively measuring similarity between groups of trajectories remains under-explored, especially in MATs. Previous methods have primarily focused on single trajectories or used simplistic models that fail to account for the complexity of group interactions and the multidimensional nature of MATs.

Traditional approaches, such as the DF distance and UMS, have effectively analyzed single trajectories based on spatial and temporal dimensions. Recent work in trajectory analysis has extended the consideration of semantic dimensions with methods like the MSM and MUITAS. However, these approaches fall short in capturing movement behaviors within groups. By summarizing group trajectories into RT's, our approach allows for assessing group patterns, thus providing a more holistic understanding of group behavior.

Techniques like flocks, clusters, and convoys have been proposed to identify and analyze groups of trajectories. However, these methods often rely on density or temporal overlap and do not provide a nuanced similarity measure that integrates multiple aspects. Our approach leverages a method to summarize trajectories with multiple aspects. In our experiment, we use MAT-SGT to summarize MATs and the MUITAS as similarity measures for MAT, identifying groups and quantifying their similarity across multiple dimensions, offering a more detailed and flexible analysis.

Limitations. While our approach advances the state-of-the-art, we acknowledge certain limitations, such as the current approach's inability to fully account for the temporal sequence of MAT points, as we do not identify similarity measures for MATs that consider it. This presents opportunities for future research to integrate temporal alignment techniques, potentially leading to even more accurate similarity measures.

6. Conclusion

This paper presents an approach to measuring similarity between groups of trajectories using representative trajectories, addressing a significant gap in the literature, mainly regarding MATs, and enabling more efficient analysis of group movement patterns. Our methodology consisted of four key steps involving data collection and preprocessing demonstrated by the Foursquare NYC dataset, representative trajectory computation with MAT-SGT, similarity measurement using MUITAS, and the evaluation step given by our case study.

The evaluation results indicated that our approach allows for identifying similar patterns among groups, showcasing the power of using representative trajectories for similarity assessment. However, the current similarity measure, MUITAS, does not account for the temporal sequence between MAT points, which may impact representativeness in certain scenarios. Additionally, to accurately compute the similarity measure between groups, it is essential to match both the representative trajectory computation and the similarity measure purpose, ensuring that the result score reflects the relevant information.

Future work should focus on integrating temporal alignment techniques to enhance accuracy, exploring hybrid approaches, or augmenting MUITAS with temporal alignment techniques, which could be beneficial. Additionally, since we lack a baseline technique to compute the similarity between groups of trajectories, especially for MATs, and due to limited space for further exploration in this work, we could potentially explore traditional techniques such as computing the average score between all measures or comparing with other baseline methods in future experiments. This approach identifies similar group patterns and holds potential applications in various domains, including urban planning, transportation, and behavioral analysis, where understanding group mobility patterns is crucial.

References

Alowayr, A. D., Alsalooli, L. A., Alshahrani, A. M., and Akaichi, J. (2021). A review of trajectory data mining applications. In *2021 International Conference of Women in Data Science at Taif University (WiDSTaif)*, pages 1–6.

Eiter, T. and Mannila, H. (1994). Computing discrete frechet distance. Technical report cd-tr 94/64, Christian Doppler Laboratory for Expert Systems, TU Vienna - Austria.

Erwig, M. et al. (1999). Spatio-temporal data types: An approach to modeling and querying moving objects in databases. *GeoInformatica*, 3(3):269–296.

Furtado, A. S., Alvares, L. O. C., Pelekis, N., Theodoridis, Y., and Bogorny, V. (2018). Unveiling movement uncertainty for robust trajectory similarity analysis. *Int. J. Geogr. Inf. Sci.*, 32(1):140–168.

Furtado, A. S., Kopanaki, D., Alvares, L. O., and Bogorny, V. (2016). Multidimensional similarity measuring for semantic trajectories. *Trans. GIS*, 20(2):280–298.

Gupta, A., Mishra, A., Vadlamudi, S., Chakrabarti, P., Sarkar, S., Mukherjee, T., and Gnanasambandam, N. (2013). A mobility simulation framework of humans with group behavior modeling. In *2013 IEEE 13th International Conference on Data Mining*, pages 1067–1072.

Jeung, H., Yiu, M. L., Zhou, X., Jensen, C. S., and Shen, H. T. (2008). Discovery of convoys in trajectory databases. *Proc. VLDB Endow.*, page 1068–1080.

Kalnis, P., Mamoulis, N., and Bakiras, S. (2005). On discovering moving clusters in spatio-temporal data. In *Advances in Spatial and Temporal Databases: 9th International Symposium, SSTD 2005*, pages 364–381. Springer.

Li, Z., Wu, F., and Crofoot, M. (2013). Mining following relationships in movement data. In *Proc. 2013 Int. Conf. Data Mining (ICDM'13)*.

Machado, V. L. (2024). *On Computing Representative Data for Summarizing Multiple Aspect Trajectories*. Tese (doutorado), Universidade Federal de Santa Catarina, Florianópolis.

Machado, V. L., Portela, T., Machado, A. d. L., Schreiner, G. A., and Mello, R. d. S. (2023a). A method for computing representative data for multiple aspect trajectories based on data summarization. In *XXIV Brazilian Symposium on Geoinformatics*.

Machado, V. L., Portela, T., Renso, C., and Mello, R. d. S. (2023b). Towards a representativeness measure for summarized trajectories with multiple aspects. In *XXIV Brazilian Symposium on Geoinformatics*.

Mello, R. d. S. et al. (2019). MASTER: A multiple aspect view on trajectories. *Trans. GIS*, 23(4):805–822.

Parent, C. et al. (2013). Semantic trajectories modeling and analysis. *ACM Comput. Surv.*, 45(4):42:1–42:32.

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

Wiratma, L., van Kreveld, M., and Löffler, M. (2017). On measures for groups of trajectories. In *20th AGILE conference on Geographic Information Science*, pages 311–330. Springer.