Spatio-Temporal Trajectory Similarity Measures: A Comprehensive Survey, Benchmark, and Evaluation

Danlei Hu, Lu Chen, Hanxi Fang, Ziquan Fang, Yunjun Gao, *Member, IEEE*

Abstract—Trajectory data analytics has inspired a broad range of real-life applications across different fields. Trajectory similarity measure, which aims to evaluate the distance between two trajectories, is a fundamental functionality of trajectory analytics. In this paper, we propose a comprehensive survey that summarizes almost all the representative trajectory measures in terms of three hierarchical perspectives (i.e., Non-learning vs. Learning, Free Space vs. Road Network, and Standalone vs. Distributed). Moreover, we provide an evaluation benchmark (publicly available at https://github.com/ZJU-DAILY/TSM) by designing five real-world transformation scenarios. Based on this benchmark, extensive experiments are conducted to study the effectiveness, robustness, efficiency, and scalability of each measure, which offers a reference for trajectory measure selection among traditional similarity measures, deep learning models, and distributed processing technologies.

Index Terms—Trajectory Similarity Measure, Distributed Similarity Search, Deep Representation Learning, Experimental Evaluation

1 Introduction

With the proliferation of GPS-equipped devices and mobile computing services, massive trajectory data of moving objects such as people, vessels, and vehicles are being captured [36], [81]. For example, people tend to share visited places (e.g., POIs) on social networks by their smartphones to generate "check-in" trajectories; according to the AIS project [43], millions of vessels connected with the AIS services continuously report locations to ensure the sailing safety; the world's largest ridesharing company Uber collects up to 17 million vehicle trips daily [58]. An original trajectory T of a moving object is typically denoted as a time-ordered sequence of spatio-temporal locations: $T = \langle p_1, p_2, ..., p_n \rangle$.

Trajectory data with its analytics benefit a broad range of real-life applications across different fields such as urban computing [35], transportation [82], behavior study [33], and public security [30], to name but a few. A fundamental functionality of most trajectory analysis is to evaluate the relationship/distance between two trajectories, i.e., trajectory similarity/distance measurement. With an accurate and efficient trajectory similarity measure, downstream trajectory analytics involving retrieval [46], [53], clustering [15], [28], classification [19], [27], and mobility pattern mining [24], [40], [55] tasks can be well-supported to serve upper applications. For instance, Xie et al. [65] propose two distance measures to support similarity queries and joins in a large-scale trajectory dataset. Wang et al. [61] design a road network oriented distance measure for vehicle trajectory similarity computation, based on which, they [63] proceed to study k-means clustering to detect hot/popular traveling paths in a city. In both cases, the trajectory similarity measure plays a fundamental role, and a different measure selection may result in totally different query results and clustering quality. More trajectory similarity based analyses tasks can refer to [36], [60].

Unlike isolated spatial points or one-dimensional time series where the distance definition is straightforward, it is non-trivial to define the distance between continuous and two-dimensional trajectories while considering trajectory characteristics: (i) Different data sources, i.e., free space vs. road network space. In the latter case, a proper trajectory measure should take the road topology into account, as people and vehicles cannot travel like vessels without spatial constraints [50]. (ii) Various sampling rates and lengths. Unlike time series that generally feature constant and high sampling rates [32], trajectory data shows varying samplings, resulting in variable trajectory lengths. (iii) The effect of noise. The noise points commonly exist, especially due to strength attenuation and interference in urban cities [77]. (iv) Complex shapes. Compared to private-car trajectories that are usually inaccessible caused by privacy principles, taxi trajectories are widely studied in the community [12], [56], [62]. However, taxi trajectories exhibit much more diverse, complex, and flexible geometric, because of various pick-pop demands. To deal with these spatio-temporal characteristics, there are tremendous amounts of research efforts on designing dozens of trajectory similarity measures in the literature.

But being faced with a huge amount of trajectory measures, researchers are often too exhausted to select a proper one. On the one hand, there are too many trajectory measures, which were proposed under different scenarios, e.g., learning based or non-learning based, free-space oriented or road network oriented, as well as standalone or distributed processing. In each scenario, various measures also exist. Consequently, users need to spend tons of time and efforts

- *D. Hu, L. Chen (Corresponding Author), Z. Fang and Y. Gao are with the College of Computer Science, Zhejiang University, Hangzhou 310027, China, E-mail:{dlhu, luchen, zqfang, gaoyj}@zju.edu.cn.*
- *H. Fang is with the School of Earth Science, Zhejiang University, Hangzhou 310027, China, E-mail:hanxif@zju.edu.cn.*