String-Based Feature Representation for Trajectory Clustering

B. A. Sabarish, Department of Computer Science and Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, India
Karthi R., Department of Computer Science and Engineering, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, India
Gireesh Kumar T, TIFAC CORE in Cyber Security, Amrita School of Engineering, Coimbatore, Amrita Vishwa Vidyapeetham, India

ABSTRACT
A trajectory is the spatial trail of a moving object as a function of time. All moving objects such as humans, robots, cloud, taxis, animals, mobile phones generate trajectories. Trajectory clustering is grouping of trajectories that have similar moving patterns, and the formed clusters depend on feature representation, similarity metrics, and clustering algorithm used. In this article, trajectory features are generated after mapping trajectories onto grids, as this smoothens the variations that occur in spatial coordinates. These variations occur due to differences in how GPS points at varying intervals are generated by the device, even when they follow the same path. The main motivation for the article is to devise an algorithm for trajectory clustering that is independent of the variations from GPS devices. A string-based model is used, where trajectories are represented as strings and string-based distance metrics are used to measure the similarity between trajectories. A hierarchical method is applied for clustering and the results are validated using three metrics. An experimental study is conducted and the results show the effectiveness of string-based representation and distance metrics for trajectory clustering.

KEYWORDS
Clustering, Similarity, Spatial Grid, Spatial Points, String Representation, Trajectory

INTRODUCTION
All geo-location services are related to the position of the user using the service. Spatial data provides information pertaining to the position of user or device and is of importance to provide geo-location-based services. GPS enabled devices provides a set of spatial coordinates over a period of time which form a trajectory of user or device. Analysis of trajectories is an active area of research as they are used in a wide range of applications to improve user experiences such as path finding, recommendation systems, traffic management and analysis, identifying hotspots, behaviour classification and anomaly detection.

Many researchers have developed algorithms to extract patterns from spatial datasets using clustering, classification and frequent mining methods. Spatial analysis is difficult due to the large volume of spatial data and the complexity in the integration of spatial parameters with other attributes considered for analysis is a challenge. Clustering is an important unsupervised learning technique where the goal of clustering is to determine the intrinsic grouping in a set of unlabeled data as

DOI: 10.4018/IJERTCS.2019040101

Copyright © 2019, IGI Global. Copying or distributing in print or electronic forms without written permission of IGI Global is prohibited.
specified by John, Nair and Kumar (2017). Cluster contains objects which are similar to one another and dissimilar objects belong to different clusters as defined by Feng and Zhu (2016). Trajectory clustering algorithms consider three aspects during algorithm development. They are a) how features are extracted from trajectories and represented as a feature vector b) finding a suitable metrics for measuring similarity c) finding an algorithm that helps in grouping these spatial data.

This paper focuses on development of an algorithm for trajectory clustering by finding a suitable representation for trajectory. Spatial locations are mapped to a grid structure and are used for trajectory feature representation. Trajectories are clustered and validated using standard metrics. The paper is organized as follows: first section discusses the literature review followed by a description of the problem of trajectory clustering. The proposed algorithm for trajectory clustering is discussed in the second section. Experimental evaluation, results and conclusions are briefed in the last sections.

LITERATURE REVIEW

Trajectory data is regarded as time sequenced spatial data and a number of trajectory clustering methods have been proposed and a review of these methods is discussed by Zheng (2015) and Sabarish, Karthi and Gireeshkumar (2015). Trajectory clustering approaches analyse trajectories as a whole as a single unit or as multiple segments. Gaffney and Smyth (1999) proposed a model-based clustering approach where trajectories are considered to be generated from mixture models and EM algorithm is used to assign data to clusters. Trajectories are considered as a whole and assigned to clusters based on cluster parameters. Lee, Han, Li and Gonzalez (2008) proposed trclus algorithm that generates many line segments for a trajectory using characteristic points in partition phase and clusters line segments using dbscan algorithm in grouping phase. Trajectories are considered as multiple sub trajectories and helps in discovering patterns in these sub trajectories.

Tripathi, Debnath and Elmasri (2016) proposed a framework where directional orientation of the trajectories is used for clustering. The algorithm does smoothing, filtration, segmentation and clustering of segments to obtain overall directional patterns in the data set. Mao, Zhong, Qi, Ping and Li (2017) proposed an adaptive dbscan trajectory clustering method where trajectory are segmented into sub trajectories using modified Minimum Description Length. The clustering space is divided into grids and trajectories are mapped into grids and an adaptive dbscan algorithm is proposed for clustering. A survey of trajectory clustering algorithms is discussed by Yuan, Sun, Zhao, Li and Wang (2017) where the trajectory algorithms are classified into categories as spatial based, time based, group and partition, uncertainty, semantic based, road network and optimization based methods for clustering.

Many traditional clustering approaches cannot be applied for trajectory data as traditional algorithms work for point data. Extending the traditional algorithms require methods to capture series of data as a single entity and propose a suitable distance metrics for clustering. Vlachos, Kollios and Gunopulos (2002), analyzed various similarity between trajectories using edit, hausdorff, Dynamic Time Warping distance (DTW), longest common subsequence (LCSS) and Fréchet distance. A comprehensive review of different trajectory distance metrics with its limitations is discussed by Besse, Guillouet, Lubes and Royer (2016). A new method called symmetrized segment-path distance is proposed by the authors for trajectory clustering. Clustering of trajectories are tested for the new similarity measure using hierarchical clustering and affinity propagation method. Wang, Su, Zheng, Sadiq and Zhou(2013), studied six distance measures used for trajectory clustering where trajectories are subjected to shifting, sampling and noise addition. Effects of similarity measures on these transformed trajectories are evaluated and discussed.

Researchers have applied several techniques for transformation of trajectories and represent them with properties that capture the hidden features with reduced dimensionally and numerosity of data. Werner and Kiermeier (2016) proposed an alignment free representation of trajectories, where features are derived from orientation and length of trajectories and represented using character strings. Encoding of trajectory into string is done by discretizing the angles and lengths aiding in capturing
16 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the product's webpage:

www.igi-global.com/article/string-based-feature-representation-for-trajectory-clustering/225485?camid=4v1


www.igi-global.com/e-resources/library-recommendation/?id=2

Related Content

Implementation of Multi-Hop Cognitive Radio Testbed using Raspberry Pi and USRP
www.igi-global.com/article/implementation-of-multi-hop-cognitive-radio-testbed-using-raspberry-pi-and-usrp/188438?camid=4v1a

Optimal Green Virtual Machine Migration Model
www.igi-global.com/article/optimal-green-virtual-machine-migration-model/95722?camid=4v1a
A Trust Based Secure and Privacy Aware Framework for Efficient Taxi and Car Sharing System
[www.igi-global.com/article/a-trust-based-secure-and-privacy-aware-framework-for-efficient-taxi-and-car-sharing-system/203683?camid=4v1a](www.igi-global.com/article/a-trust-based-secure-and-privacy-aware-framework-for-efficient-taxi-and-car-sharing-system/203683?camid=4v1a)

Optimizing the Traffic of Voice Core Domain in UMTS Network through RNC Re-Homing
[www.igi-global.com/chapter/optimizing-traffic-voice-core-domain/62768?camid=4v1a](www.igi-global.com/chapter/optimizing-traffic-voice-core-domain/62768?camid=4v1a)