A Survey on Trajectory Data Mining

V. Tanuja *thanuja_in@yahoo.com*
Research Scholar, Dept of Computer Science
Sri Venkateswara University,
Tirupathi, Chittoor(Dt) AP, India

P. Govindarajulu *pgovindarajulu@yahoo.com*
Retd Professor, Dept. of Computer Science,
Sri Venkateswara University,
Tirupathi, Chittoor (Dt) AP, India

Abstract

The availability of advanced capturing and computation techniques delivered the way to study on trajectory data, which denote the mobility of a variety of moving objects, such as people, vehicles and animals. Trajectory data mining is of the research trend in data mining research to cope with the current demand of trajectory data analysis providing profit rich applications. Data clustering is one of the best techniques to group the community data. In this paper efforts are made to review the trendy research being done in trajectory data mining. The review is three fold, surveying the literature on location and community based trajectory mining, trajectory data bases and trajectory querying. The review explored the trajectory data mining framework. This review can help outline the field of *trajectory data mining*, providing a quick outlook of this field to the community. Trajectory clustering methods are discussed. The opportunities and applications of cluster based trajectory data mining are presented. The method of similarity based community clustering is adopted to go with the future work.
.

Keywords: Trajectory Data Mining, Trajectory Clustering, Trajectory Querying, Applications of Trajectory Data Mining, Uncertain Trajectory Data.

1. INTRODUCTION

Internet enabled mobile devices are primary sources for obtaining very large volumes of trajectory data that capture the movements of different types of objects such as people, vehicles, animals and vessels [36]. The increasing pervasiveness of location acquisition technologies has enabled collection of very large trajectory datasets for different types of moving objects. Useful patterns discovered from the movement behavior of moving objects are very valuable and forms a trajectory knowledgebase, and much useful to variety of real time applications. Ubiquitous amounts of trajectory data sets are being generated continuously with the rapid development of location acquisition technologies [6]. Trajectories of moving objects are useful in finding knowledge such as moving patterns, moving group patterns, finding location of a specific object or service etc. Many global positioning systems (GPS) enabled devices are pervasive [5]. Finding movement patterns in trajectory data stream is very useful in many real time applications [12].

The source of trajectory data is typically from mobile devices that capture the position of an object at specific time intervals. The encouragement for mining trajectory datasets is the possibility of realizing inherent information, helping to gain understanding of the fundamental phenomena of movement. The knowledge of object movements is helpful within many contexts. Increasing availability of data and the number of methods for utilizing this data is making trajectory data mining gain sufficient importance in various domains, including urban planning, traffic flow control, public health, wildlife protection and location aware advertising and many more.

It is unavoidable to extract movement patterns from trajectory data because of many reasons. Finding trajectory patterns is very useful in learning interactions between moving objects. Varieties of trajectory patterns have been proposed in the literature of trajectories [13]. Trajectory data are ubiquitous in the real world and there is an ever increasing interest in performing data analysis over trajectory data. Considerable efforts have been applied to discover trajectory patterns of moving objects by using variety of data mining techniques. There is no best clustering algorithms that can solve all the problems occurring in the universe [46]. Discovery of trajectory patterns is very useful in learning interactions between moving objects [64]. In the trajectory literature many methods have been proposed for obtaining only a specific type of trajectory patterns. Nowadays trajectory data is available abundantly [67], [68].

1.1 Trajectory Data

Trajectories are subsets of spatiotemporal data that represents movements of moving objects. Trajectories must be represented with time and space properties. A trajectory is defined as continuous mapping from R (time) to R^2 (the two-dimensional plane). Discrete trajectory is defined as a poly line in three dimensional spaces (one time dimension and two space dimensions). Trajectory data are generated by various moving objects and collected from multiple data sources [68].

A trajectory T is defined as an ordered list of spatiotemporal sample points p_1, p_2, p_3, …, p_n where each $p_i = (x_i, y_i, t_i)$ and x_i, y_i, are the spatial coordinates of the sampled point and t_i is the timestamp at which the position of the point is sampled such that $t_1 < t_2 < t_3 < t_4 < \ldots < t_n$ [68]. Discrete representation of movements of continuously moving objects inherently associated with uncertainty because locations between two sampled positions are unknown. Move bank is a public moving object database that contains real world trajectory data. There exist many synthetic trajectory data sets.

A trajectory of a moving object is a discrete trace that the moving object travels in geographical space. Generally, it is a sequence of geo-locations with corresponding timestamps in spatiotemporal space [68]**.** A moving object can be a person, an animal, a vehicle, a mobile device, or even a phenomenon. A trajectory of a person records one's trace for a period of time. A trajectory of an animal describes its trace generated by daily activities such as sleeping, eating, walking, and running. Trajectory data is categorized as trajectory data of people, transportation, animal and natural phenomena and so on.

When trajectories are associated with contextual information they are called semantic trajectories. Trajectories can also be classified based on the type of moving objects, activities, modes of transportation (such as driving, walking, biking, going by flight, going by bus) and so on. Individual trajectories as well as groups of trajectories contain interesting patterns and features. There exist many methods for representing trajectories and space time cube is one such method for representing trajectories. Spatio-time cube combines time and space in a single display.

Trajectory training data sets are useful for location based recommendation applications [12] such as personalized location prediction, group based location recommendation, user mobility modeling, vehicle mobility modeling, ship mobility modeling and animal mobility modeling. Many spatiotemporal training data sets that are publicly published are useful for predicting social relationships among the people. It is very difficult to analyze and mine trajectory training data that contains complex human mobility characteristics.

1.1.1 Trajectory Databases

Trajectory Data Sampling: Generally, the sizes of trajectory databases are very large in many real time applications and operations on such trajectory databases are very expensive, complicated, inefficient, and inaccurate and time consuming. Trajectory data sampling means that selecting only the representative samples from the original set of trajectory data tuples. It is difficult to compare two trajectories derived with different sampling strategies by directly utilizing spatial proximity based similarity measures like Euclidean distance [68]. Nowadays trajectory database is an important state-of-the-art research area that is increasing rapidly. The main objective of trajectory databases is to extend the already available database technology principles to support the representation and querying of movement behaviors of moving objects and their technology.

- **Trajectory Pattern Mining**

Another trend of research has considered constraint database models to represent Trajectory Pattern Mining [13]. The main goal of trajectory pattern mining is to find interesting trajectory patterns from trajectory data sets. Close locations frequently visited by the same group of people in a given time period are clustered into regions. Scalable algorithms are needed for trajectory data mining [56].

- **Link Prediction Mining**

Main goal of link prediction mining is to find the friendship relation between two persons, similarly social movements between or among the people. Tree or graph based models are also available for determining social relationships. Many statistical measures are also available for finding link based relationships between two people, between two groups of people and among many groups of people. Human mobility factors are also used for link prediction.

- **Local Recommendation**

Main aim of local recommendation is to design and develop an intelligent computer based system that recommends locations that a user is likely to visit based on historical trajectory information of a user. Nowadays learning methods are popularly used in many real time applications. Distributed representation learning is successfully applied in many applications areas such as natural languages processing, speech recognition, robot working technique, game solving and signal processing.

1.1.2 Trajectory Data Modeling

Trajectory data modeling is receiving a lot of attention from research people who are applying existing spatiotemporal data models to trajectory data and researchers who are proposing new models for clear and accurate understanding of movement behaviors of movement objects. Many of the existing spatiotemporal models can be used to model trajectories. To model and manage adequately trajectory uncertainty different interpolation methods and modeling concepts have been proposed. A trace of a moving object in geographical space is continuous while a trajectory is only a sample of location points that the moving object passes [68].

1.2 Trajectory Data Mining Framework

With the help of many modern technologies such as global positioning systems (GPS), smart phone sensors and mobiles it is very easy to collect and store very large scale trajectory data of tracking traces of moving objects - vehicles, persons, animals, and vessels etc [36]. Examples for moving objects are - people vehicles, animals, and hurricanes etc. Many trajectory data mining based applications are beneficial to the universities, telecommunication industries, government, common people, business people and many commercial organizations.

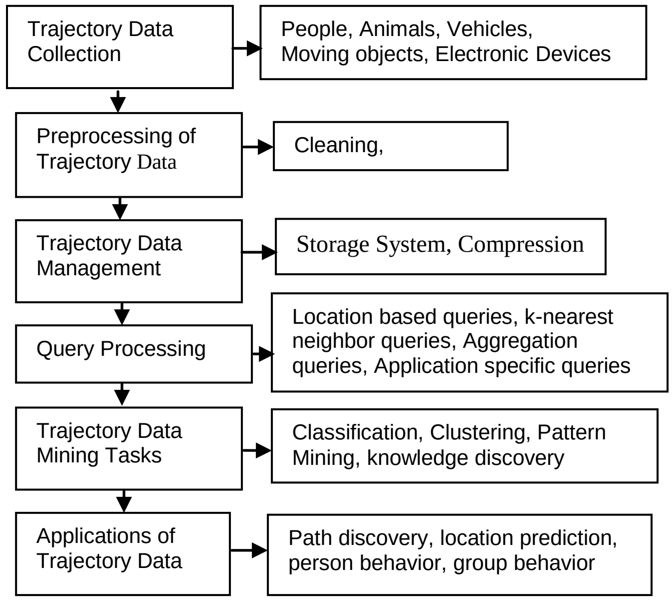

FIGURE 1: A Framework for Trajectory Data Mining.

Storing and managing very large trajectory training data sets is very difficult. Comparing trajectories of two different objects is the fundamental function in trajectory data mining and as such it is very difficult to find a similarity metric for comparing trajectory sets of two different users because trajectories of different objects are constructed with different sampling strategies and sampling rates. Constructing a similarity metric for comparing trajectories of two different objects is very difficult because trajectories belonging to two different objects are not homogeneous. Important layers in the trajectory data mining framework are - *data collection*, *trajectory data mining techniques*, *applications*. Specifically, the layers of trajectory data mining techniques contain six components listed in Figure 1[68].

There exist a wide spectrum of applications driven and improved by trajectory data mining and these applications significantly benefit the common people, commercial organizations and government agencies. Trajectory data mining techniques are directly related to the collected trajectory data and the requirements of the applications. Trajectory data mining techniques are divided into six categories:

a) Trajectory data collection
b) Preprocessing of trajectory data
c) Trajectory data management
d) Query processing
e) Trajectory data mining tasks
f) Applications of trajectory data mining

a) Trajectory Data Collection

Nowadays large scale trajectory data sets of moving objects are generated with the help of state-of-the-art location acquisition technologies such as Mobile Phones, Satellites, Sensors, GPS, Radio Frequency Identification (RFID), infrared rays, Video, GMS Network, Smart phones sensors, Infrared systems, Ultrasonic systems, GSM beacons, WLAN networks, Location estimation, Satellite imaginary, Robotics, Web traffic Monitoring Devices [14], [20].

b) Preprocessing of Trajectory Data

Preprocessing is the vital phase in data mining. When trajectory datasets are collected at very first time they are called raw trajectory datasets and raw trajectories contain uncertainty and noise. In general, raw trajectory datasets are of very huge sizes containing millions of trajectory tuples. Similarity measures designed by using raw trajectory datasets must be preprocessed so that noise and uncertainty is reduced to the maximum extent and also trajectory database sizes are reduced in such a way that trajectory contains only important representative location details of trajectory paths. Preprocessing is a basic step that performs at the beginning and it aims at improving quality of trajectory data and generating sub-trajectories [68]. Trajectory preprocessing prepares or modifies raw trajectory data into a convenient and efficient useful trajectory data sets or subsets. Preprocessing reduces the raw trajectory data sets by pruning and selecting only important sampled points.

Trajectory data preprocessing includes trajectory data cleaning and it is the most important and prerequisite requirement for any trajectory data mining technique. The performance of trajectory data mining decreases as the number of outliers increases in the trajectory data.. Main goal of trajectory data cleaning is to find and remove outliers at the beginning itself [14]. Uncertain trajectory data sets must be used with the help of other rules and constraints. Trajectory data completion is a way of improving the quality of uncertain trajectories before actually usage of them in any trajectory data mining [28].

c) Trajectory Data Management

Trajectory data management [15] requires efficient, effective, robust, accurate and scalable features for storing very large sizes of trajectory training data sets. Efficient storage of very large trajectory databases is the fundamental problem in trajectory management [15]. It is impossible to store and transmit very large trajectory databases generated by many location acquisition techniques such as GPS, mobiles etc. There exist many trajectory compression techniques and algorithms to reduce storage requirements and network transmission costs and loads. There exist many artificial neural network and machine earning techniques for compressing trajectory databases [16]. Achieving a reasonably good compression ratio is possible by sacrificing a small amount of accuracy. Parallel compression leads to efficient compression [31]. Some of the efficient storage methods for trajectory data storage are segment, sub-directory, column oriented, representative samples data sets, frame sets and pruned data sets [68]

A dynamic indexing structure [26] is needed to maintain tremendous amount of trajectory databases. B+ tree is the state-of-the-art indexing data structure for relational databases but it is one dimensional. R+ tree, R* tree and R tree are spatial indexing data structures. Similarly a state-of-the-art dynamic indexing data structure is needed for efficient trajectory data management. An efficient Trajtree tree indexing data structure is developed for trajectory data management particularly for data retrieval, for query answers.

d) Query Processing

Main goal of query processing system is to retrieve desired data conveniently from the specified trajectory data storage system efficiently with high speed. Whenever a trajectory data query is posed, the underlying trajectory data storage and management system is responsible for providing answers to such queries and data must be retrieved efficiently and effectively [68].

Location Based Queries: A query component system must provide answers to various queries.. The main application of location based queries is finding the best route for a trip from the given multiple places. For a given location query, the query component of the trajectory data management system returns a set of trajectories each of which contains all given query locations in a specific order or in random fashion [36], [39].

Range Rueries: A range query [55] always gives two bounds - lower and upper bounds. There exist many applications of trajectory data management that involves range queries. One example for range query is to find all trajectories that contain a set of cities in the path trajectory such that starting from the 9^{th} city to 18^{th} city. Range queries are particularly suitable for handling uncertain trajectories. In many real life applications, data may be uncertain. One way to manage and control uncertain trajectories is with the help of probabilistic data model involving Probability

Density Function (PDF). Range queries also retrieve trajectories that cover a given area with the specific threshold value of the probability at a specific interval of time [66].

K-nearest Neighbor Queries: The trajectory management system returns k-nearest neighbor trajectories with respect to the specific given trajectory. K-NN query is an integral part of any spatial temporal trajectory data mining. Particular type of nearest neighbor queries called probabilistic nearest neighbor queries are particularly suitable for handling uncertain trajectories in the uncertain trajectory databases [17], [39].

Top-k Queries: Top-k queries mean finding the most important k-similar trajectories for a given trajectory by considering uncertain trajectory database. There exists many probabilistic top-k nearest neighbor queries. A scalable indexing data structure is needed to support efficient query retrievals in trajectory data mining tasks [9].

Pattern Queries: In order to apply query patterns, trajectory management system must support a way for obtaining results to certain pattern based queries. There must be a provision for writing a query that specifies a set of predicates that must be satisfied in a specific order. The predicate may be a range condition or a k-nearest neighbor condition or a top-k query or a pattern query or combination of two or more of the above. A regular language consisting of regular expressions is particularly suitable for applying pattern queries [49], [50].

Aggregate Queries**:** Aggregate queries are useful to find statistical details of trajectories stored in the trajectory databases. Total or average number of persons travelled on a specific road segment in a given time interval is an example for aggregate query. Average velocity, maximum and minimum velocity of a set of vehicle trajectories on a particular road segment is another aggregate query [31].

Application Specific Queries**:** Keyword queries for semantic trajectory are very useful in many location based applications for example; a person wants to visit a temple, a university, a city during his tour. Three keywords-temple, university, and cities are extracted from the given query and then a search is initiated in order to find keyword based search results from the trajectory databases. Keyword search may be exact keyword search, approximate keyword search, fuzzy keyword search and so on [63], [67].

e) Trajectory Data Mining Tasks

Most important and frequently used data mining tasks [68] are: classification, clustering, pattern mining, knowledge discovery, frequent pattern mining and so on. Trajectory data mining tasks are classified as:

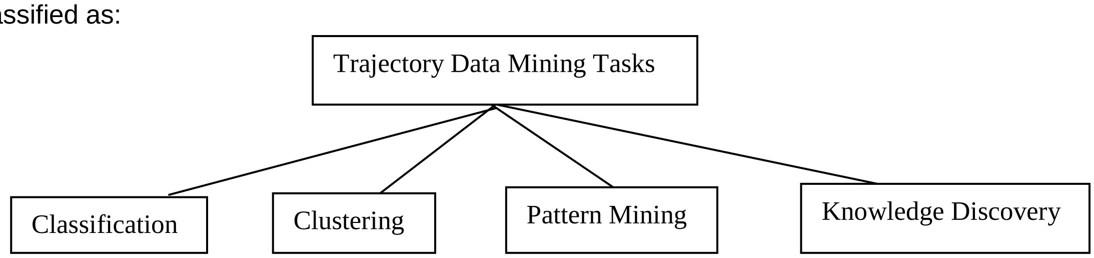

FIGURE 2: Trajectory Data Mining Tasks.

- **Classification**: Classification is a two-step process. In the first step a model is constructed using trajectory training data set. In the second step constructed model is used to predict the class labels of test trajectories. Various types of new trajectory properties are applied during trajectory data classification [10].

- **Clustering**: Trajectory data clustering is useful to cluster trajectory data into groups of trajectories based on similar movement patterns. Moving objects are grouped based on similarity movement measures such as movement velocity, movement direction, spatial distance, spatial dispersion, temporal duration and syntax and semantic meaning of locations. Finding only important locations of a trajectory is an example for clustering and the mobility-based clustering is less sensitive than the density-based clustering to the size of trajectory dataset [68]**.**

- **Pattern Mining**: Pattern mining means analyzing and finding movement patterns of moving objects or group of moving objects for efficient and effective decision making. Different types of pattern mining are – individual patterns, group patterns, sequential patterns, periodic patterns, frequent patterns, rare patterns and so in. A trajectory is a sequence of locations. A subsequence is a set of common locations shared by at least minimum number of trajectory where minimum number is the specific threshold value that depends upon on a specific application. This subsequence is called sequential pattern. Sequential pattern can be coarse patterns or fine grained patterns and can be represented by using a tree data structure for efficient processing. A periodic trajectory pattern is very useful to understand the behavior of moving objects [64].

- **Knowledge Discovery:** Knowledge that is mined in trajectory data mining may be used in many places and in many applications. Some of the possible potential applications are clustering of moving vehicles in a city, classification of buying patterns, finding sequential patterns in market basket analysis, share market share trend prediction and so on [60].

- **Privacy Protection:** Privacy protection is one of the important tasks in trajectory data mining. It is very difficult to maintain both trajectory data mining and privacy protection simultaneously. Always trajectory data must be privacy protected. There is a need for applying encryption techniques to maintain privacy preserving of trajectory data [43].

f) Applications of Trajectory Data Mining

Potential applications of data mining are retailing, banking, credit card management, insurance, telecommunications, telemarketing and human resource management. There exist many applications of similarity search on spatio-temporal trajectories. A suitable distance function is used for finding similarity between uncertain trajectories. Trajectory clustering algorithms for moving objects are very useful in finding traffic jams, important location identification, and facilities available at a particular location at particular time etc. Some of the applications of trajectory data mining are Route Recommendation, Animal Migration, Transportation Management, Real time Traffic Information Details of Transport Organization, and Tourism [69]. Some of the useful trajectory data mining applications are path discovery, shortest path discovery, individual behavior prediction, group behavior prediction, location prediction, service prediction and so on. Applications of trajectory data mining can be classified based on domain of application as follows:

Path Discovery**:** Path discovery is also known as route discovery. There exist many ways for finding a path between any two given nodes. Sometimes there is a need to find most frequent path between two locations in a certain time period. In many real time applications there is a need to find an efficient and correct path. For some applications, shortest path is needed, some applications require shortest path, another set of applications need cheapest or most popular path. Path discovery must find at least one path. Most frequent paths are better than the fastest paths or shortest paths in many real time applications [6].

In terms of public transportation, people's real demand for public transportation are employed to identify and optimize existing flawed bus routes, thus improving utilization efficiency of public transportation [68]**.** Dai *et al.* [11] proposed a recommendation system that chooses different routes for drivers with different driving preferences. This kind of *personalized route recommendation* avoids flaws of previous unique recommendation and improves quality of user satisfaction. Previous experiences showed that human mobility as extraordinary regular and thus predictable.

Location/destination Prediction: Location based services are particularly useful to the people in urban areas. When locations aware services are equipped with technologies such as GPS, smart phones, mobiles and so on, it is very easy to find location of a person, or a vehicle or a flight or a train or service or process and so on [20]. Destination prediction is similar to the path discovery [41]. Many location based applications require location prediction or destination prediction to send advertisements to targeted consumers, to recommend tourist spots or restaurants, or to set destinations in navigation systems [68]. Hidden Markov Model (HMM) is useful for predicting the next place that a user will visit by analyzing and exploring human mobility patterns [37].

Movement Behavior Analysis: A set of Trajectory may represent movement behavior of a single person or movement behavior of a group of persons. Movement behavior of moving objects is represented in the form of trajectories and then analyzed and processed before taking crucial decisions. Finding movement patterns is very important in order to understand human behavior clearly. Predicting human behavior accurately under emergency is a crucial issue for disaster alarming, disaster management, disaster relief and societal reconstruction after disasters. Human mobility behavior can be studied from spatial, temporal and social aspects [68].

Gao *et al.* [15] proposed a comprehensive analysis of temporal effects in modeling mobility behavior. It has been studied that human mobility exhibits strong temporal cyclic patterns in the period of an hour; day or week. Liu *et al.* [33] proposed a method to model trajectories in terms of user decision on visiting a point of interest (POI) and conduct rationality analysis upon trajectory behavior. Rationality of trajectory behavior is explored through several impacting factors.

***Group Behavior Analysis*:** Moving objects, especially people and animals, sometimes tend to form groups or clusters due to their social behavior. For instance, movement of a person is affected by not only personal activities, but also social ties with that of the groups he belongs to [63]. Besides, a gathering pattern, as a novel modeling of trajectory patterns, describes movement pattern of a group of moving objects. Examples include celebrations, parades, traffic congestion, large-scale business promotions, protests etc [68].

In trajectory data mining a group is defined as a cluster that has at least *m* moving objects being densely connected for at least certain duration of time. Gupta *et al.* [18] first address a problem of efficiently modeling individual and group behavior and then present a simulation framework that simulates people's movement behavior in order to generate spatiotemporal movement data and the simulation is of great significance since a large amount of movement data in public domain are limited and unavailable in reality [68].

Urban Service: The knowledge discovered after applying trajectory data mining techniques is useful in many ways for urban area people. Large scale trajectory data are collected by using electronic techniques [60]. People living in a city need the knowledge to assist their decision on buying or renting a house, choosing a job. Meanwhile, the knowledge helps urban planners to make decisions on future development of the city and to estimate effects of previous policies. There exist a great many studies to facilitate interpretation of raw trajectory data [68].

The huge volume of spatial trajectories enables opportunities for analyzing the mobility patterns of moving objects, which can be represented by an individual trajectory containing a certain pattern or a group of trajectories sharing similar patterns [69]. Mobility of people: people have been recording their real-world movements in the form of spatial trajectories, passively and actively, for a long time [69]. Mobility of transportation vehicles: A large number of GPS-equipped vehicles (such as taxis, buses, vessels, and aircrafts) have appeared in our daily life. Mobility of animals: Biologists have been collecting the moving trajectories of animals like tigers and birds, for the purpose of studying animals’ migratory traces, behavior, and living situations [69]. Mobility of natural phenomena: Meteorologists, environmentalists, climatologists, and oceanographers are busy collecting the trajectories of some natural phenomena, such as hurricanes, tornados, and ocean currents.

Many previous research works have elaborated on discovering user communities from user location history and such a community is thus referred to as a movement-based community. In light of movement-based communities, many novel location-based services and application systems can be developed [53]. To discover movement-based communities, one should first formulate the similarity of users in terms of their trajectory profiles and a trajectory profile should capture two kinds of movement information: a) the sequential patterns of hot regions, and b) The transition probabilities among hot regions [53].

Wen-Yuan Zhu et al. [53] proposed the concept of trajectory profile as a tree structure called sequential probability tree (SP-tree) to characterize one user's movement behavior from trajectories in terms of sequential patterns and transition probabilities to the next movement and based on the trajectory profiles.

Q. Li et al. [32] proposed a framework called hierarchical-graph-based similarity measurement (HGSM) for geographic information systems to consistently model each individual's location history and effectively measure the similarity among users. Seeking to find common behaviors in trajectory data, some methods group whole trajectories into clusters while others attempt to mine regional patterns that trajectories follow in a sub-segment of their evolutions and many of the existing trajectory clustering methods approach the problem by first defining a similarity measures (DTW, LCSS) for trajectories and then using one of the well-established clustering procedures [36]
Vehicle Motion Trajectory Dataset: This dataset contains 1500 trajectories gathered by tracking vehicles at a traffic intersection. These trajectories are annotated manually; each trajectory is assigned to one of 15 trajectory classes. The Greek Trucks Dataset: This dataset contains 1100 trajectories from 50 different trucks delivering concrete around Athens, Greece [36]. Atlantic Hurricane Dataset (HURDAT2): This dataset is provided by the National Hurricane Service (NHS) and contains 1740 trajectories of Atlantic Hurricanes from 1851 through 2012 [36].

2. RELATED WORK

Location Acquisition Technologies: In the trajectory data literature object representative locations are called hot regions or popular regions or selected regions or modified regions or pruned regions or transformed trajectories. Frequently visiting locations by the users are called transformed locations and only those locations are represented in trajectory datasets. Modified locations contain features that are useful to describe movement behavior of users. A transformed trajectory is a sequence of hot regions [53].

Users' movement behavior details are captured in hot regions by using sequential pattern mining techniques. Frequent sequential patterns are extracted from sequential patterns. Frequent sequential patterns are subsequences of the modified trajectories. Time complexity of deriving similarity measures is very high due to the very large size of frequent sequential patterns [32].

A trajectory comprises of a sequence of time stamped sample points, each containing a time, latitude and longitude. Line generalization represents a trajectory using a series of linear segments created by joining selected trajectory points. Compression is achieved by discarding unselected trajectory points. There are two main approaches to compressing trajectories: line generalization and delta compression [1].

Aiden Nibali et al. [1] proposed a new trajectory compression system called **Trajic** system.. Authors also proposed an algorithm, SQUISH, is a one pass trajectory compression algorithm that is based on the line generalization approach of discarding data points while minimizing the increase in compression error. AssawerZekri et al. [3] proposed a new similarity function and many frameworks with respect to tree structured data, namely Extended Sub-tree (EST). EST generalizes the edit based instances by providing new rules for sub-tree mapping. An empirical runtime analysis demonstrates that the new approach is one of the best tree distance functions in terms of runtime efficiency [2].

Tree comparison is performed by tree distance/similarity functions. The applications include document clustering, natural language processing, cross browser compatibility, and automatic web testing. However, a group of mapped nodes should have a stronger emphasis on the similarity of trees when they form an identical sub-tree. That is, an identical sub-tree represents a similar sub structure between trees, whereas disjoint mapped nodes indicate no similar structure between the two trees [2].

AssawerZekri et al. [3] proposed a conceptual modeling for trajectory ETL process and trajectory data warehouse, and also proposed two algorithms in order to implement trajectory ETL tasks and to construct trajectories. Banerjee P et al. [4] proposed a technique called InferTra to infer uncertain trajectories from network constrained partial observations and rather than predicting the most likely route, the inferred uncertain trajectory produces many probable routes and for trajectory inference,

XinCao et al. [5] proposed a framework that encompasses new techniques for extracting semantically meaningful geographical locations from such data and for the ranking of these locations according to their significance. Chen Z. et al. [6] proposed a spatial indexing algorithm to achieve efficient K-BCT search, based on a simple Incremental K-NN based algorithm (IKNN) and the contributions mainly lie in adapting the best-first and depth first KNN algorithms to the basic IKNN properly. Chiang M. et al. [7] proposed a reachability based prediction model on Time constrained Mobility Graph (RTMG) to address the sparsity in low-sampling-rate trajectories and to predict locations for distant-time queries.

Christine Parent et al. [8] proposed new ideas and techniques related to the elaboration and analysis of semantic trajectories. Raw trajectory data must be combined with related contextual data and these enriched trajectories are called semantic trajectories. Weka-STPM is an extended version of Weka data mining tool kit and the first toolkit for multilevel mining of semantic trajectories. M-ATLAS is another tool that analyses semantic trajectories to infer behavior [8]. Chunyang Ma et al. [9] proposed a new measuring technique called p-distance that is able to quantify the dissimilarity between two uncertain trajectories using indexing technique called UTgrid. UTgrid based query processing is more scalable in terms of query execution time, input, output.

Hoyoung et al. [20] proposed a flexible trajectory modeling approach that takes into account model-inferred actual positions, time varying uncertainty, and non-deterministic uncertainty ranges. Many state-of-the-art positioning technologies are available for determining trajectory of moving objects across different locations over time. Data uncertainty is based on a common principle “location uncertainty is captured by an interval on the position recorded in the database”.

Kai Zheng et al.[28] proposed an intuitive model for uncertain trajectory that represents the motion along a road network, and provide a unified probability density function for the possible locations of a moving object at a given time-instant. An effective indexing structure as well as efficient processing algorithms for both snap-shot and continuous range queries is designed. Dynamic time wrapping (DTW) [43] and longest common subsequences (LCSS) are two popular similarity measuring techniques based on time series analysis. Those two similarity measuring techniques are suitable for discrete data but not for interval data. Expected Euclidean distance (ED) is another similarity measuring technique between two uncertain trajectories but it is not a reliable one because it is sensitive. When trajectories are associated with a discrete probability distribution then probability distribution is taken as Probability Mass Function (PMF). Probability density function (PDF) is used to represent trajectories with continuous probability distribution [9].

Cynthia Sung et al. [10] proposed a data driven approach to predicting motion. In the first method, trajectories are classified based on similarity between trajectories and is appropriate for short trajectories and it cannot discover common subsequences in long trajectories. The second method of trajectory classification aims to generate a spatial map of typical motion in a trajectory. The knowledge is the ultimate goal that will provide clues to decision makers who can tell the pattern of movement frequently conducted by mobile users to move from one base location to another base location through a set of intermediate locations [12].

ElioMasciari et al. [13] proposed highly efficient, that is flexible and scalable algorithm for extracting frequent patterns from trajectory data streams. The problem of mining for frequent trajectories has many real time applications such as vehicle traffic management, supply chain management, user movement management, transportation, animal movement and hand-off in cellular networks. Fazzinga B. et al. [14] proposed a probabilistic framework for reducing the inherent uncertainty of trajectory data and also the authors experimentally evaluated the efficiency and effectiveness of the proposed approach.

The Gook-PilRoh et al. [16] proposed an algorithm to represent a trajectory as a sequence of road segments in a road network. They defined a road segment as its starting and ending positions and require that it should not intersect with any other segment. It is aimed to study pattern matching problems for trajectory data over road networks. HaozhouWang et al. [19] proposed algorithms for finding trajectory similarity measures under different circumstances using a common real world taxi cab trajectory dataset.

Huey-Ru and Wu et al. [21] proposed an algorithm called Divclust for finding regional typical moving styles by dividing and then clustering the trajectories. Travelled path of migrating animals are released by GPS services. Jae-Gil Lee et al. [22] proposed a new partition-and-group framework for clustering trajectories and main advantage of this framework is to discover common sub-trajectories from a trajectory data. Authors also proposed a trajectory clustering algorithm called TRACLUS based on partition-and-group framework.

Jae-Gil Lee et al. [23] proposed UT-patterns. There are several kinds of synchronous movement patterns: for example, a set of moving objects move together, a set of moving objects chase another set of moving objects with a small time delay [23]. Jae-Gil Lee et al. [24] focused on the methods for classifying trajectories on road networks by analyzing the behavior of trajectories on road networks. Based on the analysis, it is evident that (frequent) sequential patterns are good feature candidates since they preserve this order information.

Jing Yuan et al. [25] proposed a smart driving direction system leveraging the intelligence of experienced drivers and it is used to find out the practically fastest route for a particular user at a given departure time. Jinfeng Ni et al. [26] proposed parametric indexing using polynomial approximations that can improve query performance significantly over current schemes using space indexing and also proposed a new indexing scheme called PA-tree indexing for historical trajectory data using polynomial approximations and PA-tree indexing supports both offline and online.

Jun Pang et al. [27] studied the impact from communities on a user's mobility and predict his locations based on his community information. The main aim is to study the impact from communities on a user's mobility and predict his locations based on his community information. Larusso N.D et al. [30] proposed a novel model for accurately tracking coordinated groups of mobile uncertain objects and proposed algorithm is more efficient approximate interface algorithm for updating the current location of each object upon the arrival of new location observations.

Li Y et al. [31] proposed a solution by using well-established spatio-temporal index called inverted index to trajectory data management and authors designed random index sampling (RIS) algorithm to estimate the answers with a guaranteed error bound. For scalability purpose RIS algorithm is extended to concurrent random index sampling (CRIS) algorithm to process a number of trajectory aggregate queries arriving concurrently with overlapping spatiotemporal query regions. Long C. et al. [34] proposed to preserve the direct information of the trajectory for simplification and this referred to as direction preserving trajectory simplification (DPTS).

Luis OtavioAlvares et al. [35] proposed a reverse engineering approach, where data mining is applied to extract only strong and interesting patterns from trajectory data in order to provide a set of pattern relationships that are not known a priori. The use of data mining techniques to extract patterns from trajectory data has some general advantages and the storage of these patterns will help the user to represent efficiently to write queries over moving object patterns [33]. Some methods group whole trajectories into clusters while others attempt to mine regional patterns that trajectories follow in a sub-segment of their evolutions and many of the existing trajectory clustering methods approach the problem by first defining a similarity functions or measures (DTW, LCSS) for trajectories and then using one of the well-established clustering procedures [36]

MikołajMorzy et al. [37] proposed a data mining approach to the problem of predicting the location of a moving object. They mine the database of moving object locations to discover frequent trajectories and movement rules. MingqiLv et al. [38] proposed route pattern mining framework-which is designed to adapt to the high degree of uncertainty of personal trajectory data, proposed a trajectory abstraction technique-which uses a group-and-partition approach to detect common sub-segments,

Ningnan Zhou et al. [40] proposed a model called a general Multi-Context Trajectory Embedding Model, denoted by MC-TEM and its aim is to provide a general and flexible manner to characterize multiple kinds of contextual information for trajectory data. Oznur Kirmemis Alkan et al. [42] stated that, frequent pattern mining is a fundamental research topic that has been applied to different kinds of databases and it has been studied extensively by data mining researchers. Panagiotakis C et al. [44] proposed a method for trajectory segmentation and sampling based on the representativeness of the trajectories in the moving object databases.

Pavel Senin [45] evaluated trajectory similarity measures in an objective manner and main aim is to provide a quantitative analysis on the effectiveness of the trajectory similarity measures.. PETTER KIHLSTROM [46] proposed some tips–how various trajectory data mining tasks are defined at an abstract level, what type of information can be extracted from the trajectory data, what are available tools for finding solutions to various problems related to trajectory data, how tools are visualized and used in various trajectory data mining applications etc.

Ranu Sayan et al. [47] proposed TrajTree to formulate a robust distance function called Edit Distance with Projections (EDwP) to match trajectories under inconsistent and variable sampling rates through dynamic interpolation. It is an efficient trajectory retrieval. Extensive experiments on real trajectory databases demonstrate EDwP to be up to 5 times more accurate than the state-of-the-art distance functions.

Ticiana Coelho da Silva et al. [49] proposed an efficient online monitoring framework that manages group patterns updates while new trajectory data are coming scalable methods are used to create trivial pattern groups of events or incidents or gathering with online updates trajectory movements of objects such as people, vehicle and so on.

Vieira M.R. et al. [50] proposed a frame work for efficient processing of flexible pattern queries and the frame work includes an underlying indexing structure and algorithms for query processing using different evaluation strategies. Vijay Kovvali et al. [51] listed the data elements and data formats for the vehicle trajectory and supporting data, and provides a structured approach for future video data collection efforts and then the ultimate objective is to provide a research framework for collecting vehicle trajectory and supporting data for behavioral algorithm and safety research.

Wen-Chih Peng et al. [52] proposed an algorithm to capture the frequent user moving patterns from a set of log data in a mobile environment and it is enhanced with the incremental mining capability and is able to discover new moving patterns efficiently without compromising the quality of results obtained. Traffic data visualization can facilitate understanding of the behavior of moving objects and discovery of traffic, social, geo-spatial and even economic patterns [54]. Types of trajectories are: shipping trajectories, aircraft trajectories, automobile trajectories, train/metro trajectories, pedestrian trajectories and mixed trajectories and so on. Space-Time-Cube (STC) is a widely studied method for data with spatiotemporal attributes in a STC. A 3D trajectory is visualized in a 3D coordinate system [54].

Xikexie et al. [57] proposed a common model to find possible locations of an object when locations are bounded in a closed region (interval) that is imprecise (uncertain) and also proposed a new technique called u-bisector to effectively evaluate trajectory queries over imprecise objects. Yinlai Jiang et al. [58] proposed methods for finding motions of the human body in the fields of medical care, sports, arts, and so on. Yuan N. J et al. [60] proposed techniques to discover functional zones in urban areas using trajectories; typically a city is naturally partitioned into individual regions by major roads - like express ways and ring roads.

Zaiben Chen et al. [61] proposed an algorithm to find the k Best-Connected Trajectories (k-BCT) from a database such that the k-BCT best connect the designated locations geographically. Zhan. L. et al. [62] proposed a general frame work for range search on uncertain trajectories and another paradigm called filtering-and-refinement, where summaries of uncertain trajectories are constructed to facilitate the filtering process and statistical based and partition based filtering techniques are developed to enhance the filtering capabilities.

Zheng K. et al. [64] proposed a new trajectory pattern modeling technique called gathering that can be applied on groups such as celebrations, parades, traffic jams and so on. Usually the size of each group is very large and it is associated with durable and stable areas with high density. ZHENG Y et al. [65] proposed a location-history-based recommended system, which estimates the similarity between users in terms of their movements in geographical spaces. The proposed framework is called hierarchical graph based similarity measure (HGSM).

Zheng K., et al. [66] proposed an algorithm on representing the uncertainty of the objects moving along road networks as time dependent probability distribution functions. Uncertain Trajectories Hierarchy (UTH) indexing structure is proposed for implement efficient algorithms. Zhiwei Lin et. Al. [70] proposed to represent trees as multidimensional sequences (MDSs), and measure their similarity based on this representation.
Zhixian Yan et al. [71] proposed a trajectory ontology framework to capture semantics for trajectory data as well as to support automatic reasoning and this framework consisting of three modules ontology, geometric trajectory ontology, geographic ontology and application domain ontology.

3. OPPORTINUTIES AND CHALLENGES IN TRAJECTORY DATA MINING

Trajectory patterns are very useful in learning interactions between moving objects. In the trajectory patterns literature, many types of trajectory patterns are available. Many trajectory data mining methods are available for only specific type of trajectory patterns; trajectory data mining is very difficult unless hidden patterns are identified clearly and correctly in the selected trajectory training data sets. There is a need to discover initial patterns and their granularity levels in order to use those patterns in many trajectory data mining applications efficiently and effectively. Patterns must be structured by using an appropriate pattern modeling data structure. There is a need to construct efficient frameworks for easy discovery of various patterns from many real time trajectory datasets. Satellite, mobile, sensor, wireless and video technologies allow to collect and capture object movements for constructing very large trajectory training data sets. Examples for trajectory training datasets are – movement data of people, vehicles, animals, ships navigation and flight paths. Existing trajectory data patterns are classified into:

- Flock patterns, Convey patterns
- Swam patterns, Moving patterns
- Time relaxed trajectory joins

- Hot motion paths and, Sub-trajectory clusters

Trajectory data mining is very useful for effective decision making in many real time applications and day by day very large scale trajectory datasets are being generated by GPS-enabled devices. Many trajectory data researches have not given considerable amount of importance of semantic representation and interpretation of the trajectory data. Raw trajectory data sets must be interpreted to satisfy the semantic meaning of real life entities such as people, vehicles, animals, roads, flights, network channels and so on. Semantic representation of trajectory data improves readability and understandability of trajectory data [48].

There are multifold opportunities with the trajectory data mining when it is applied with current data trends available. Trajectory data patterns can be classified to get interesting groups and the group behaviors which will show the way to solve real problem challenges. Trajectory clustering is the current buzzword in the 21st century data world.

The opportunities are seamless but some major ones include:

- Elicitation of peoples/objects movement pattern to study and predict their trend.
- Finding significant communities in social and business networks.
- Understanding and analyzing the correlations between users and their locations to predict their future plans.
- Grouping users, clustering locations.
- Uncovering movement patterns that provide means to profitable decisions in business situation.
- Location recommendations for a particular interest.
- Outlier trajectory studies and methods.

Many of the means depends on similarity measures, metrics and the consequential results. To measure the similarity of the features of trajectory segments, the features must be normalized. For example, cosine similarities are most widely used vector similarity measure. State-of-the-art positioning techniques such as GPS, mobile devices, smart phones are very convenient for finding the position of people, vehicles, flights, animals, ships and so on at any time. Trajectory databases are actual movements of users executing their tasks in many real time situations. Trajectory data mining is rapidly increasing and as a result many trajectory data sharing websites such as every trail running free are created.

Many research people have been tried to find user communities from trajectory data collected from movement location details. Movement based communities [53] are common in many areas and useful in many applications. Some of the real time applications that are based on movement based user communities are

1. Friend communication
2. Trajectory ranking
3. Community-based traffic sharing services

There is a large ground to go with the trajectory data mining and its application. Hither to the relevant research is going around limited methods and applications. Object movements, sensor outputs, community patterns are given most importance. The real world is a source of ample of trajectory related data and applications. The availability of trajectory data mining methods and there evolution providing the opportunity to go with many data sets available in the real scenario and the application of the results will provide the way to prosper.

Similarity measures between users must be computed using trajectories of respective users before finding movement based user communities. Users' trajectory datasets are collection of sequence of locations associated with timestamps. Users' movement behavior details are captured in the hot regions. Frequent sequential patterns are extracted from sequential patterns.

Extra memory storage is needed for storing frequent sequential patterns as their sizes are very large. In order to compare movement behaviors among the users, the similarity measures must be derived from frequent sequential patterns. Transition probabilities and user movement sequences are collectively needed in order to derive effective movement based similarity measures.

There exist many efficient advanced data structures for accurate representation of user movement profiles of trajectories. Trajectory profiles are very fundamental frequent movement sequences of user mobility. First user movement behaviors are represented in the form of trajectories and then these trajectories are represented using efficient advanced data structures that are very convenient for both representation and processing.

It is evaluated from many research papers that there are several vital challenges along the process arriving at profitable knowledge about moving objects. An inherent uncertainty is observed while studying the location between two sampled positions. To reduce this type of uncertainty and prepare raw data for mining, data often needs to be processed in some way. The quality of pre-processing depends on sampling rate and accuracy properties of raw data as well as the requirements formulated by the specific mining method. A major challenge is to define relevant knowledge and effective methods for extracting this from the data.

There is a need to improve trajectory data mining techniques in many ways. Trajectory data of moving objects must be collected with state-of-the-art techniques and collected data must be modified so that noise, errors and outliers are removed and reduced to the maximum extent. Good privacy preserving techniques are required to provide better security measures to the trajectory databases. Sometimes it may be necessary to combine trajectory data with other data mining methods for obtaining better results.

4. CONCLUSION AND FUTURE WORK

In this paper, the detailed review on trajectory data mining is presented. The current trends and opportunities with trajectory data mining in general and with trajectory data clustering in particular are addressed. The opportunities for the research towards community movement clustering are identified. The high amount of scope and need is there in future to analyze the current data trends to go for better decisions and knowledge of movement behaviors of moving objects obtained by analyzing and processing historical trajectories. It is the known fact that trajectory data mining provides plenty of convenience to the public. Trajectory data mining is useful for both government and private organizations in many ways to reduce cost of supervision and management. The velocity, volume and variety of big data trends are encouraging clustered based analysis. In future, there is a scope to investigate efficient similarity measuring techniques between trajectories of moving objects with respect to the scalability.

The future research in data mining will be contributed with trajectory data trend analysis related to money market movements, traffic analysis, township planning, job search, career planning adviser systems and many more to be added .Therefore the role and requirement of the proposed research in the future is a fine mean to researchers to select the area with plenty of opportunities and challenges.

5. REFERENCES

[1] Aiden Nibali and Zhen He “Trajic: An Effective Compression System for Trajectory Data”. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 27, NO. 11, NOVEMBER 2015.

[2] Ali Shahbazi, Student Member, IEEE and James Miller, Member, IEEE “Extended Subtree: A New Similarity Function for Tree Structured Data” IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 4, APRIL 2014.

[3] Assawer Zekri & Jalel Akaichi, Computer science department. BESTMOD Laboratory,High Institute of Management, University of Tunis, “An ETL for Integrating Trajectory Data”, International Conference on Automation, Control, Engineering and Computer Science (ACECS'14) Proceedings – Copyright IPCO-2014, pp.138-147,ISSN 2356-5608.

[4] Banerjee P., S. Ranu, and S. Raghavan, “Inferring uncertain trajectories from partial observations,'' in Proc. IEEE Int. Conf. Data Mining (ICDM),Shenzhen, China, Dec. 2014, pp. 30_39.

[5] Cao X., G. Cong, and C.S. Jensen, “Mining Significant Semantic Locations from GPS Data,” Proc. VLDB Endowment, vol. 3, no. 1, pp. 1009-1020, Sept. 2010.

[6] Chen Z., H. T. Shen, X. Zhou, Y. Zheng, and X. Xie, “Searching trajectories by locations: An efficiency study,'' in Proc. ACM SIGMOD Int. Conf. Manage. Data (SIGMOD), Indianapolis, IN, USA, Jun. 2010, pp. 255_266.

[7] Chiang M., Y.-H. Lin, W.-C. Peng, and P. S. Yu, “Inferring distant-time location in low-sampling-rate trajectories,'' in Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), Chicago, IL, USA, Aug. 2013, pp. 1454_1457.

[8] CHRISTINE PARENT, University of Lausanne, STEFANO SPACCAPIETRA, Swiss Federal Institute of Technology (EPFL),CHIARA RENSO, ISTI-CNR,GENNADY ANDRIENKO, Fraunhofer Institute IAIS “Semantic Trajectories Modeling and Analysis” Publications Dept., ACM, Inc., 2,Penn Plaza, New York, NY 11201-0701, USA.

[9] Chunyang Ma, Hua Lu, Member, IEEE, Lidan Shou, and Gang Chen “KSQ:Top-k Similarity Query on Uncertain Trajectories”. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 9, SEPTEMBER 2013.

[10] Cynthia Sung, Dan Feldman, Daniela Rus, “Trajectory Clustering for Motion Prediction “.

[11] Dai. J, B. Yang, C. Guo, and Z. Ding, ``Personalized route recommendation using big trajectory data,'' in Proc. 31st IEEE Int. Conf. Data Eng. (ICDE), Seoul, South Korea, Apr. 2015, pp. 543_554.

[12] DAVID TANIAR and JOHN GOH, School of Business Systems, Monash University, Clayton, Vic, Australia” On Mining Movement Pattern from Mobile Users”In International Journal of Distributed Sensor Networks, 3: 69–86, 2007,ISSN: 1550-1329 print/1550-1477 online.

[13] Elio Masciari, ICAR-CNR,iRende, , Gao Shi,UCLA,Westwood Los Angeles, USA &Carlo Zaniolo UCLA, Westwood, Los Angeles, USA,” Sequential Pattern Mining from Trajectory Data”.

[14] Fazzinga B., S. Flesca, F. Furfaro, and F. Parisi, ``Cleaning trajectory data of RFID-monitored objects through conditioning under integrity constraints,'' in Proc. 17th Int. Conf. Extending Database Technol. (EDBT), Athens, Greece, Mar. 2014, pp. 379_390.

[15] Gao, J. Tang, X. Hu, and H. Liu, ``Modeling temporal effects of human mobile behavior on location-based social networks,'' in Proc.22nd ACM Int. Conf. Inf. Knowl. Manage. (CIKM), San Francisco, CA, USA, Oct./Nov. 2013, pp. 1673_1678.

[16] Gook-Pil Roh, Jong-Won Roh, Seung-Won Hwang, and Byoung-Kee Yi “Supporting Pattern-Matching Queries over Trajectories on Road Networks” IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 23, NO. 11, NOVEMBER 2011.

[17] Guting, R.H., T. Behr, and J. Xu, ``Efficient k-nearest neighbor search on moving object trajectories,'' VLDB J., vol. 19, no. 5, pp. 687_714, Oct. 2010.

18] Gupta. A. et al., ``A mobility simulation framework of humans with group behavior modeling,'' in Proc. IEEE 13th Int. Conf. Data Mining, Dallas, TX, USA, Dec. 2013, pp. 1067_1072.

[19] Haozhou Wang1 Han Su, Kai Zheng1 Shazia Sadiq, Xiaofang Zhou,School of Information Technology and Electrical Engineering, The University of Queensland, “An Effectiveness Study on Trajectory Similarity Measures” Proceedings of the Twenty-Fourth Australasian Database Conference (ADC 2013), Adelaide, Australia.

[20] Hoyoung Jeung, Hua Lu, Saket Sathe, and Man Lung Yiu, “Managing Evolving Uncertainty in Trajectory Databases,” IEEE Trans. Knowledge and Data Eng., vol. 26, no. 7, pp. 1692–1705, July. 2014.

[21] Huey-Ru Wu, Mi-Yen Yeh, and Ming-Syan Chen, “Profiling Moving Objects by Dividing and Clustering Trajectories Spatiotemporally,” IEEE Trans. Knowledge and Data Eng., vol. 25, no. 11, pp. 2615–2628, Nov. 2013.

[22] Jae-Gil Lee, Jiawei Han, Department of Computer Science,University of Illinois at Urbana-Champaign, jaegil@uiuc.edu, hanj@cs.uiuc.edu & Kyu-Young Whang, Department of Computer Science / AITrc KAIST,kywhang@cs.kaist.ac.kr “Trajectory Clustering: A Partition-and-Group Framework”.

[23] Jae-Gil Lee, Jiawei Han, and Xiaolei Li, “A Unifying Framework of Mining Trajectory Patterns of Various Temporal Tightness,” IEEE Trans.Knowledge and Data Eng., vol. 27, no. 6, pp. 1478–1490,June.2015.

[24] Jae-Gil Lee, Member, IEEE, Jiawei Han, Fellow, IEEE, Xiaolei Li, and Hong Cheng “Mining Discriminative Patterns for Classifying Trajectories on Road Networks” IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING,VOL.23, NO.5, MAY 2011.

[25] Jing Yuan, Student Member, IEEE, Yu Zheng, Senior Member, IEEE, Xing Xie, Senior Member, IEEE, and Guangzhong Sun, Member, IEEE “T-Drive: Enhancing Driving Directions with Taxi Drivers Intelligence”. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 1, JANUARY 2013.

[26] Jinfeng Ni and Chinya V. Ravishankar, Senior Member, IEEE "Indexing Spatio-Temporal Trajectories with Efficient Polynomial Approximations” IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 19, NO. 5, MAY 2007.

[27] Jun Pang University of Luxembourg jun.pang@uni.lu & Yang Zhang University of Luxembourg FSTC, “Location Prediction: Communities Speak Louder than Friends”.

[28] Kai Zheng1 Goce Trajcevski2 Xiaofang Zhou1;3 Peter Scheuermann, “Probabilistic Range Queries for Uncertain Trajectories on Road Networks”.

[29] Kong, L. He, X. Y. Liu, Y. Gu, M.Wu, and X. Liu, ``Privacy-preserving compressive sensing for crowd sensing based trajectory recovery,'' in Proc. 35th IEEE Int. Conf. Distrib. Comput. Syst. (ICDCS), Columbus, OH, USA, Jun./Jul. 2015, pp. 31_40.

[30] Larusso N. D. and A. Singh, ``Efficient tracking and querying for coordinated uncertain mobile objects,'' in Proc. 29th IEEE Int. Conf. Data Eng. (ICDE), Brisbane, QLD, Australia, Apr. 2013, pp. 182_193.

[31] Li et al. Y., ``Sampling big trajectory data,'' in Proc. 24th ACM Int. Conf. Inf. Knowl.Manage. (CIKM), Melbourne, VIC, Australia, Oct 2015 pp.941_950.

[32] Li Q., Y. Zheng, X. Xie, Y. Chen, W. Liu, and W.-Y. Ma, “Mining User Similarity Based on Location History,” Proc. 16th ACM SIGSPATIAL Int’l Conf.Advances in Geographic Information Systems(GIS’08), Nov 2008.

[33] Liu.S, Q. Qu, and S.Wang, ``Rationality analytics from trajectories,'' ACM Trans. Knowl. Discovery Data, vol. 10, no. 1, p. 10, Jul. 2015.

[34] Long C., R. C.-W. Wong, and H. V. Jagadish, ``Trajectory simplification: On minimizing the direction-based error,'' in Proc. PVLDB, 2014, vol. 8. no. 1, pp. 49_60.

[35] Luis Otavio Alvares, Vania Bogorny, Jose Antonio Fernandes de Macedo,Bart Moelans, Stefano Spaccapietra “Dynamic Modeling of Trajectory Patterns using Data Mining and Reverse Engineering”.

[36] Mahdi M. Kalayeh, Stephen Mussmann_, Alla Petrakova_, Niels da Vitoria Lobo and Mubarak Shah “Understanding Trajectory Behavior: A Motion pattern approach” arXiv:1501.00614v1 [cs.CV] 4 Jan 2015.

[37] MikołajMorzy”Mining Frequent Trajectories of Moving Objects for Location Prediction”, Institute of Computing Science Pozna´n University of Technology Piotrowo 2, 60-965 Pozna´n, Poland.

[38] MINGQI LV, YINGLONG LI, ZHENMING YUAN AND QIHUI WANG,College of Computer Science and Technology,Zhejiang University of Technology,Zhejiang, 310023 P.R. China “Route Pattern Mining From Personal Trajectory Data”, JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 31, 147-164 (2015).

[39] Niedermayer.J, A. Zü_e, T. Emrich, M. Renz, N. Mamoulis, L. Chen,and H.-P. Kriegel, ``Probabilistic nearest neighbor queries on uncertain moving object trajectories,'' in Proc. PVLDB, 2013, vol. 7, no. 3,pp. 205_216.

[40] Ningnan Zhou, Wayne Xin Zhao, Xiao Zhang, Ji-Rong Wen, and ShanWang, “General Multi-Context Embedding Model for Mining Human Trajectory Data,” IEEE Trans. Knowledge and Data Eng., vol. 28, no. 8, pp. 1945–1958, August. 2016.

[41] Noulas. A., S. Scellato, N. Lathia, and C. Mascolo, ``Mining user mobility features for next place prediction in location-based services,'' in Proc.12th IEEE Int. Conf. Data Mining (ICDM), City of Brussels, Belgium, Dec. 2012, pp. 1038_1043.

[42] Oznur Kirmemis Alkan and Pinar Karagoz, “CRoM and HuspExt: Improving Efficiency of High Utility Sequential Pattern Extraction,” IEEE Trans.Knowledge and Data Eng., vol. 27, no. 10, pp. 2645–2657, October. 2015.

[43] Pelekis. N, A. Gkoulalas-Divanis, M. Vodas, D. Kopanaki, and Yheodoridis, ``Privacy-aware querying over sensitive trajectory data,'' in Proc. 20th ACM Conf. Inf. Knowl. Manage. (CIKM), Glasgow, Scotland, Oct. 2011, pp. 895_904.

[44] Panagiotakis C., N. Pelekis, I. Kopanakis, E. Ramasso, and Y. Theodoridis,`Segmentation and sampling of moving object trajectories based on representativeness,'' IEEE Trans. Knowl. Data Eng., vol. 24, no. 7, pp.1328_1343, Jul. 2012.

[45] Pavel Senin, Information and Computer Science Department University of Hawaii “Dynamic Time Warping Algorithm Review “ at Manoa Honolulu, USA senin@hawaii.edu December 2008.

[46] PETTER KIHLSTRÖM “Literature Study and Assessment of Trajectory Data Mining Tools”, Degree project in built environment.

[47] Ranu S., D. P, A. D. Telang, P. Deshpande, and S. Raghavan, ``Indexing and matching trajectories under inconsistent sampling rates,'' in Proc. 31st IEEE Int. Conf. Data Eng. (ICDE), Seoul, South Korea, Apr. 2015, pp. 999_1010.

[48] Su H., K. Zheng, K. Zheng, J. Huang, and X. Zhou, ``STMaker: A system to make sense of trajectory data,'' in Proc. PVLDB, 2014, vol. 7. no. 13, pp. 1701_1704.

[49] Ticiana Coelho da Silva, Federal University of Ceará, Brazil ticianalc@ufc.br, Karine Zeitouni, Université de Versailles-St-Quentinrance karine.zeitouni@uvsq.fr, José A. F. de, Macêdo Federal University of Ceará, Brazil jose.macedo@lia.ufc.br Marco A. Casanova, Department of Informatics - PUC-Rio, Brazil casanova@inf.puc-rio.br “On-Line Mobility Pattern Discovering using Trajectory Data” , OPEN PROCEEDINGS, Series ISSN: 2367-2005.

[50] Vieira M. R., P. Bakalov, and V. J. Tsotras, ``Querying trajectories using flexible patterns,'' in Proc. 13th Int. Conf. Extending Database Technol. (EDBT), Lausanne, Switzerland, Mar. 2010, pp. 406_417.

[51] Vijay Kovvali, Ph.D., T.E., Vassili Alexiadis, Ph.D., and Lin Zhang, Ph.D.,P.E., “Video-Based Vehicle Trajectory Data Collection”Cambridge Systematics, Inc., Kovvali, Alexiadis and Zhang 2.

[52] W.-C. Peng and M.-S. Chen, “Developing Data Allocation Schemes by Incremental Mining of User Moving Patterns in a Mobile Computing System,” IEEE Trans. Knowledge and Data Eng., vol. 15, no. 1, pp. 70-85, Jan./Feb. 2003.

[53] Wen-Yuan Zhu, Wen-Chih Peng, Chih-Chieh Hung, Po-Ruey Lei, and Ling-Jyh Chen, “Exploring Sequential Probability Tree for Movement-Based Community Discovery,” IEEE Trans. Knowledge and Data Eng., vol. 26, no. 11, pp. 2717–2730, November. 2014.

[54] Wei Chen, Fangzhou Guo, and Fei-Yue Wang, Fellow, “A Survey of Traffic Data Visualization” IEEE Trans. Knowledge and Data Eng.,

[55] Xiaohui Li,Vaida Ceikute,Christian S. Jensen, Fellow, IEEE, and Kian-Lee Tan “Effective Online Group Discovery in Trajectory Databases” IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 25, NO. 12, DECEMBER 2013.

[56] Xiaoliang Geng, Hiroki Arimura ,Graduate School of Information Sciece and Technology,IST, Hokkaido University,N14, W9, Sapporo 060-0814, Japan & Takeaki Uno, National Institute of Informatics,Japan,” Pattern Mining from Trajectory GPS Data” Proc.3rd IIAI International conference on e-service and knowledge management (IIAI ESKM’12) Fukoaka,September’2012.

[57] Xike Xie, Man L. Yiu, Reynold Cheng, and Hua Lu, “Scalable Evaluation of Trajectory Queries over Imprecise Location Data,” IEEE Trans. Knowledge and Data Eng., vol. 26, no. 8, pp. 2029–2044, Aug. 2014.

[58] Yinlai Jiang, Isao Hayashi, and Shuoyu Wang, “Knowledge Acquisition Method Based on Singular Value Decomposition for Human Motion Analysis,” IEEE Trans. Knowledge and Data Eng., vol. 26, no. 12, pp.3038–3050, Dec. 2014.

[59] Yuan. J, Y. Zheng, and X. Xie, ``Discovering regions of different functions in a city using human mobility and POIs,'' in Proc. 18th ACM SIGKDD Int.Conf. Knowl. Discovery Data Mining (KDD), Beijing, China, Aug. 2012, pp.186_194.

[60] Yuan N. J., Y. Zheng, X. Xie, Y. Wang, K. Zheng, and H. Xiong, ``Discovering urban functional zones using latent activity trajectories,'' IEEE Trans. Knowl. Data Eng., vol. 27, no. 3, pp. 712_725, Mar. 2015.

[61] Zaiben Chen, Heng Tao Shen, Xiaofang Zhou, Yu Zheng, Xing Xie,” Searching Trajectories by Locations – An Efficiency Study”,SIGMOD’10, June 6–11, 2010, Indianapolis, Indiana, USA.

[62] Zhan L., Y. Zhang, W. Zhang, X. Wang, and X. Lin, ``Range search on uncertain trajectories,'' in Proc. 24th ACM Int. Conf. Inf. Knowl. Manage.(CIKM), Melbourne, VIC, Australia, Oct. 2015, pp. 921_930.

[63] Zheng. B, N. J. Yuan, K. Zheng, X. Xie, S. W. Sadiq, and X. Zhou, ``Approximate keyword search in semantic trajectory database,'' in Proc.31st IEEE Int. Conf. Data Eng. (ICDE, Seoul, South Korea, Apr. 2015,pp. 975_986.

[64] Zheng K., Y. Zheng, N. J. Yuan, S. Shang, and X. Zhou, ``Online discovery of gathering patterns over trajectories,''IEEE Trans. Knowl.Data Eng.,vol.26, no. 8, pp. 1974_1988, Aug. 2014.

[65] Zheng Y., L. Zhang, Z. Ma, X. Xie, and W.-Y. Ma, “Recommending Friends and Locations Based on Individual Location History,” ACM Trans. Web, vol. 5, no. 1, article 5, Feb. 2011.

[66] Zheng K., G. Trajcevski, X. Zhou, and P. Scheuermann, ``Probabilistic range queries for uncertain trajectories on road networks,'' in Proc. 14th Int. Conf. Extending Database Technol. (EDBT), Uppsala, Sweden, Mar. 2011, pp. 283_294.

[67] Zheng, S. Shang, N. J. Yuan, and Y. Yang, ``Towards efficient search for activity trajectories,'' in Proc. 29th IEEE Int. Conf. Data Eng. (ICDE), Brisbane, QLD, Australia, Apr. 2013, pp. 230_241.

[68] ZHENNI FENG1 AND YANMIN ZHU (Member, IEEE), Department of Computer Science and Engineering, Shanghai Jiao Tong University, Shanghai 200240,China Shanghai Key Laboratory of Scalable Computing and Systems, Shanghai Jiao Tong University, Shanghai 200240, China. “A Survey on Trajectory Data Mining: Techniques and Applications”.

[69] Zheng Y. 2015. “Trajectory Data Mining: An Overview”, ACM Transactions on Intelligent Systems and Technology (TIST) - Survey Paper, Regular Papers and Special Section on Participatory Sensing and Crowd Intelligence, Volume 6 Issue 3, May 2015 Article No. 29.

[70] Zhiwei Lin, Hui Wang, and Sally McClean, “A Multidimensional Sequence Approach to Measuring Tree Similarity,” IEEE Trans. Knowledge and Data Eng., vol. 24, no. 2, pp. 197–208, Feb. 2012.

[71] Zhixian Yan, supervised by Prof. Stefano Spaccapietra, EPFLSwiss,Federal Institute of Technology, & Lausanne, Switzerland, “Towards Semantic Trajectory Data Analysis: A Conceptual and Computational Approach”.