Computing with Spatial Trajectories

Edited By

Yu Zheng
Microsoft Research Asia, Beijing, China

Xiaofang Zhou
The University of Queensland, Brisbane, Australia

July, 2011

Foreword

With the rapid development of wireless communication and mobile computing technologies and global positioning and navigational systems, spatial trajectory data has been mounting up, calling for systematic research and development of new computing technologies for storage, preprocessing, retrieving, and mining of trajectory data and exploring its broad applications. Thus, *computing with spatial trajectories* becomes an increasingly important research theme. Although there are many books on spatial databases, mobile computing, and data mining, this is a unique book dedicated to computing with spatial trajectory data, with a broad spectrum of coverage and authoritative overview.

Despite of many years of research on algorithms and methods on general database systems and data mining, spatial trajectory computing deserves dedicated study and in-depth treatment because of its unique nature of data semantics, structures, and applications. Such a unique nature calls for in-depth study of many interesting issues, including spatial trajectory data preprocessing, trajectory indexing and query processing, trajectory pattern mining, uncertainty and privacy in trajectory data, location-based social networks, and application of trajectory computing, such as for driving and other activities. This book, "*Computing with Spatial Trajectories*", by Yu Zheng and Xiaofang Zhou, provides a comprehensive coverage on the above topics timely, with conciseness and clear organization. The authors of the book are active researchers on different aspects on computing with spatial trajectories, and have made tangible contributions to the progress of this dynamic research frontier. This ensures that the book is authoritative and reflects the current state of the art. Nevertheless, the book gives a balanced treatment on a wide spectrum of topics, well beyond the authors' own methodologies and research scopes.

Computing with spatial trajectories is still a fairly young and dynamic research field. This book may serve researcher and application developers a comprehensive overview of the general concepts, techniques, and applications on trajectory indexing, search and data mining, and help them explore this exciting field and develop new methods and applications. It may also serve graduate students and other interested readers a general introduction to the state-of-the-art of this promising research theme.

I find the book is enjoyable to read. I hope you like it too.

July, 2011 *Jiawei Han*
University of Illinois at Urbana-Champaign

Preface

A *spatial trajectory* is a trace generated by a moving object in geographical spaces, usually represented by of a series of chronologically ordered points, e.g., $p_1 \rightarrow p_2 \rightarrow \cdots \rightarrow p_n$, where each point consists of a geospatial coordinate set and a timestamp such as $p = (x, y, t)$.

The advances in location positioning and wireless communication technologies have given rise to the prevalence of mobile computing systems and location-based services (LBS), leading to a myriad of spatial trajectories representing the mobility of a variety of moving objects, such as people, vehicles, animals, and natural phenomena, in both indoor and outdoor environments. Below are some examples.

1) Mobility of people: People have been recording their real-world movements in the form of spatial trajectories, passively and actively, for a long time.

- Active recording: Travelers log their travel routes with GPS trajectories for the purpose of memorizing a journey and sharing experiences with friends. Bicyclers and joggers record their trails for sports analysis. In Flickr, a series of geo-tagged photos can formulate a spatial trajectory as each photo has a location tag and a timestamp corresponding to where and when the photo was taken. Likewise, the "check-ins" of a user in Four-square can be regarded as a trajectory, when sorted chronologically.
- Passive recording: A user carrying a mobile phone unintentionally generates many spatial trajectories represented by a sequence of cell tower IDs with corresponding transition times. Meanwhile, transaction records of a credit card also indicate the spatial trajectory of the cardholder, as each transaction contains a timestamp and a merchant ID denoting the location where the transaction occurred.

2) Mobility of vehicles: In recent years, a large number of GPS-equipped vehicles have appeared in our daily life. For instance, many taxis in major cities have been equipped with a GPS sensor, which enables them to report a time-stamped location to a data center with a certain frequency. Such reports formulate a large amount of spatial trajectories that can be used for resource allocation, security management, and traffic analysis

3) Mobility of animals and natural phenomena: Biologists solicit the moving trajectories of animals like migratory birds for research projects. Similarly, climatologists are busy collecting the trajectories of some natural phenomena, such as hurricanes, tornados, and ocean currents. These trajectories provide scientists with rich information about the objects they are studying.

Overall, spatial trajectories have offered us unprecedented information to understand moving objects and locations, calling for systematic research and development of new computing technologies for the processing, retrieving, and mining of trajectory data and exploring its broad applications. Therefore, *computing with spatial trajectories* has become an increasingly important research theme, attracting extensive attention from numerous areas, including computer science, biology, sociology, geography, and climatology.

Although there are many books on spatial databases, mobile computing, and data mining, this is the first book dedicated to computing with spatial trajectory data, with a broad spectrum of coverage and an authoritative overview. Aimed at advanced undergraduates, graduate students, researchers, and professionals, this book covers the major fundamentals and the key advanced topics that shape the field. Each chapter is a tutorial that provides readers with an introduction to one important aspect of computing with spatial trajectories and also contains many valuable references to relevant research papers. This book provides researchers and application developers a comprehensive overview of the general concepts, techniques, and applications of trajectory indexing, search, and data mining, and helps them explore this exciting field and develop new methods and applications. It also offers graduate students and other interested readers a general introduction to the most recent developments in this promising research area.

We chose 17 active researchers in the field of computing with spatial trajectories to contribute chapters to this book in their areas of expertise. These chapters are organized according to the paradigm of "trajectory preprocessing (prior databases) → trajectory indexing and retrieval (in databases) → advanced topics (above databases)," as illustrated in Figure 1.

- The first two chapters of the book introduce the foundation of technology dealing with spatial trajectory data: Trajectory Preprocessing (Chapter 1) and Trajectory Indexing and Retrieval (Chapter 2).
- The second section is comprised of 6 advanced topics: Uncertainty in Spatial Trajectories (Chapter 3), Privacy of Spatial Trajectories (Chapter 4), Trajectory Pattern Mining (Chapter 5), Activity Recognition Based on Spatial Trajectories (Chapter 6), Trajectory Analysis for Driving (Chapter 7), and Location-Based Social Networks (Chapter 8 and 9).

Specifically, the book gradually introduces the concepts and technologies for solving the problems that newcomers will be faced with when exploring this field, starting from the preprocessing and managing of spatial trajectories, then to mining uncertainty, privacy, and patterns of trajectories, and finally ending with some advanced applications based on spatial trajectories including activity recognition,
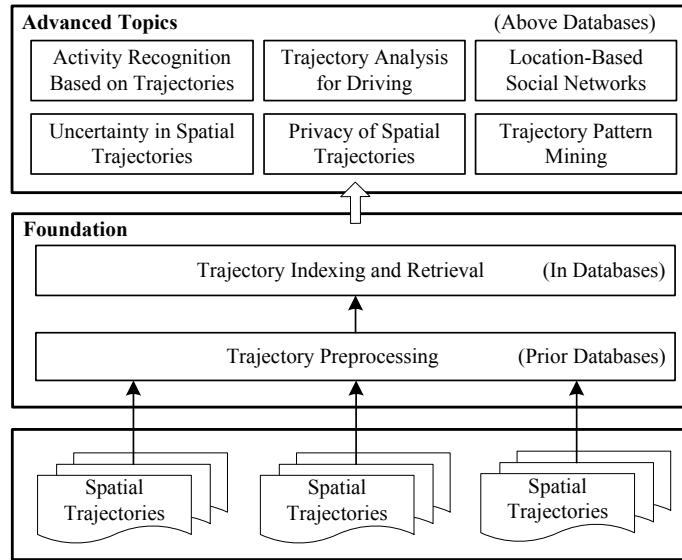
Fig. 1 The framework of the book

driving, and location-based social networks. Below is a brief introduction to the topics that will be covered in each chapter:

Chapter 1: While spatial trajectories carry rich information that can be used in a variety of applications, we have to deal with a number of issues before using them. Generally, the continuous movement of an object is recorded in an approximate form as discrete samples of location points. A high sampling rate of location points generates accurate trajectories, but will result in a massive amount of data leading to enormous overhead in data storage, communications, and processing. Thus, it is vital to design data reduction techniques that compress the size of a trajectory while maintaining the utility of the trajectory. Meanwhile, a trajectory is usually generated with occasional outliers or some noisy points caused by the poor signal of location positioning systems. For example, when traveling in a "city canyon," the satellite signals to a GPS device might be very poor, thereby generating some location points with a significant offset to the real positions. Sometimes, the offsets are more than a mile, creating noise in the trajectories and reducing the effectiveness of techniques and systems that use such trajectories. As a result, techniques for filtering the noisy points are needed for preprocessing spatial trajectories.

To address these two issues, Chapter 1 first presents data reduction techniques that can run in a batch mode after the data is collected or in an online mode as the data is being collected. The second part of the chapter introduces methods for filtering measurement noise from location trajectories, including mean and median filtering, the Kalman filter, and the particle filter. In short, this chapter provides a newcomer with the fundamentals for preprocessing spatial trajectories.

Chapter 2: The prevalence of various location-based services leads to a myriad of trajectories that create a huge burden of computation for these application systems. It is very time-consuming to find information in a trajectory dataset that is not well organized. For instance, retrieving the trajectories that pass a crossroad is a simple task, but it would make online systems unfeasible if these systems have to scan a large trajectory dataset in a direct way. At the same time, we need to search for particular trajectories satisfying certain criteria, such as spatial and temporal constraints, on many occasions. For example, retrieving the trajectories of tourists passing a given region and within a time span can help with trip planning. The requests from real applications call for effective and efficient trajectory indexing and retrieval technologies. As a consequence, Chapter 2 introduces the types of queries that are usually issued in a trajectory database and the corresponding query processing approaches supported by indexing and retrieval techniques.

Chapter 3: After preprocessing and organizing spatial trajectories with corresponding techniques, we can start using them in a variety of applications. However, positioning devices are inherently imprecise, resulting in some uncertainty with regards to acquired locations of a moving object. For instance, the reading of a GPS sensor usually has a 10 meters or more positioning error. With such a reading, it might not be easy to identify the exact point of interest (like a restaurant or a shopping mall) the moving object visited, especially in a dense urban area. At the same time, objects move continuously while their locations can only be updated at discrete times, leaving the location of a moving object between two updates uncertain. Two reasons that lead to long-interval updates are to save energy consumption and communication bandwidth. When the time interval between two updates exceeds several minutes or hours, the uncertainty in a spatial trajectory will reduce its utility and create new challenges when searching for a moving object.

To handle the above-mentioned uncertainty, Chapter 3 presents a systematic overview of the various issues and solutions related to the notion of uncertainty in the settings of spatial trajectories. The problems related to modeling and representing the uncertainty in Moving Objects Databases (MOD) are introduced. The problems of efficient algorithms for processing various spatio-temporal queries are also discussed. Note that the query processing introduced in Chapter 2 does not consider the uncertainty of a spatial trajectory while this is one of the focuses of Chapter 3.

Chapter 4: Although LBSs provide many valuable applications to mobile users, revealing people's private locations to potentially untrusted LBS service providers poses privacy concerns. There is a trade-off between the quality of services offered by a LBS provider and the privacy of a user's location. The more precisely a user reveals her location, the better the services that can be offered, the less privacy is preserved.

In general, there are two types of LBSs, namely, snapshot and continuous LBSs. For a snapshot LBS, a mobile user only needs to report her current location to a service provider once to get the desired information. In fact, it is not necessary for a user to tell a LBS system her exact location when using such services. For instance, finding hotels close to a user only requires a rough geo-region including the user's

current location. A bunch of literature has discussed a handful of methods protecting a user's snapshot location, so Chapter 4 will not cover them again.

On the other hand, a mobile user has to report her location to a service provider in a periodic or on-demand manner to obtain continuous LBS (for example, getting real-time traffic conditions around a user or searching for the nearest gas stations while driving). Protecting user location privacy for a continuous LBS is more challenging than a snapshot LBS, because adversaries may use the spatial and temporal correlations between the samples of a user's trajectory to infer the user's location information with a higher degree of certainty. In short, releasing original spatial trajectories to the public or a third party could pose serious privacy concerns. As a result, privacy protection in a continuous LBS and trajectory data publication has increasingly drawn attention from the research community and industry. Under the circumstances, Chapter 4 describes state-of-the-art privacy-preserving techniques for a continuous LBS and trajectory publication.

Chapter 5: 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. It can also be segments of different trajectories holding similar properties (e.g., formulating a cluster according to spatial and temporal constraints), or a set of full trajectories that satisfy the same conditions. These patterns can benefit a broad range of application areas and services, including transportation, biology, sports, and social services. For example, finding clusters of trajectories, in which trajectories have similar spatial shapes, can help detect the popular driving paths of users or migrating routes of birds. Additionally, identifying a group of people moving together may contribute to the exploration of social relationships, enable friend recommendations, or allow for taxi sharing. For these reasons, Chapter 5 provides an overview of trajectory patterns introduced in existing literature, divides these patterns into categories according to different taxonomies, and reviews indexing structures and algorithms for trajectory pattern mining.

Chapter 6: After the preprocessing, managing, and pattern mining of spatial trajectories, people might wonder what kind of advanced applications can be enabled based on these trajectories. Activity recognition is one of the primary applications that can be used with trajectories. Intuitively, spatial trajectories generated by people imply user behavior and activities, offering new insights into the high-level goals and objectives of users based on low-level sensor readings. First, the activities of an individual can be used as a context to trigger services that satisfy the individual's unspoken needs. For example, if it is known that a user is driving, her mobile phone can automatically show the traffic conditions on roads around the user and disable the phone's entertainment functions (for her safety) that would distract a driver's attention. If we know a user is in a meeting or watching a movie, the mobile phone of a user can be switched into quiet mode. Second, the activities of multiple users enable us to mine collective social knowledge that contributes to the analysis of social networks and transportation. With the activity information of multiple users, we can estimate the similarity between two different users more precisely, thereby, offering better services for community discovery and friend and location recommendations.

While providing richer information beyond snapshot location data, spatial trajectories also demand more advanced techniques for trajectory-based activity recognition. To help newcomers address this issue, Chapter 6 overviews the existing research into trajectory-based activity recognition and classifies them into categories according to the number of users involved in the training and inference stage.

Chapter 7: The trajectories of vehicles have a potentially strong connection to transportation, as driving is one of the most central aspects of our lives. Rich knowledge can be learned from these trajectories, such as information about road networks, traffic conditions, and driver behavior, contributing to different aspects of the driving experience. For example, creating a road map from GPS trajectories can be a less expensive way to make up-to-date road maps than traditional methods. Meanwhile, effective route recommendations can be enabled based on the trajectories of one or more experienced drivers.

Chapter 7 describes how a driver benefits from the analysis of spatial trajectories, following the paradigm of "creating road map from GPS trajectories → mapping a single trajectory to road networks → mining effective driving routes from drivers' trajectories → personalizing driving routes for a particular driver based on the preferences learned from her trajectories."

Chapter 8 and 9: The advances in location positioning and wireless communication technologies have led to a myriad of user-generated spatial trajectories, which imply rich information about user behavior, interests, and preferences. Recently, people have started sharing their trajectory data via online social networking services for a variety of purposes, fostering a number of trajectory-centric LBSNs (location-based social networks). For example, users can record travel routes with GPS trajectories to share travel experiences in an online community (e.g., GeoLife), or log jogging and bicycle trails for sports analysis and experience sharing. In addition, the "check-ins" of an individual in Foursquare and the photo trips of a user in Flickr can be regarded as spatial trajectories. These trajectory-centric LBSNs enable us to understand users and locations respectively, and explore the relationship between them.

On the one hand, we can understand an individual and the similarity between two different users with user-generated trajectories, thereby providing a user with personalized services and enabling friend recommendation and community discovery. On the other hand, we are able to understand a location and the correlation between two different locations based upon the information from users, thereby offering users better travel recommendations.

Chapter 8 defines the meaning of location-based social network and discusses the research philosophy behind LBSNs from the perspective of users and locations. Under the circumstances of a trajectory-centric LBSN, this chapter explores two fundamental research points concerned with understanding users in terms of their locations. One is modeling the location history of an individual using the individual's trajectory data. The other is estimating the similarity between two different people according to their location histories. The similarity represents the strength of connection between two users in a location-based social network, and can enable friend recommendations and community discovery. Some possible methods for evaluation

of these applications have been discussed, and a number of publically available datasets have been listed in Chapter 8 as well.

While chapter 8 studies the research philosophy behind a location-based social network from the point of view of users, Chapter 9 gradually explores the research into LBSNs from the perspective of locations. A series of research topics is presented with respect to mining the collective social knowledge from many users'GPS trajectories to facilitate travel. First, the generic travel recommendations provide a user with the most interesting locations, travel sequences and travel experts in a region, as well as an effective itinerary conditioned by a user's start location and available time length. Second, the personalized travel recommendations discover the locations matching an individual's interests, which can be gleaned from the individual's historical data.

We hope you will find this book provides a useful overview of and a practical tutorial on the young and evolving field of computing with spatial trajectories.

July, 2011   *Yu Zheng*,
                                                                      Microsoft Research Asia, Beijing, China

                                                                                               Xiaofang Zhou,
                                                              The University of Queensland, Brisbane, Australia

Acknowledgements

Contents

Part II Advanced Topics

3   Uncertainty in Spatial Trajectories   63
Goce Trajcevski

List of Contributors

Wang-Chien Lee
The Pennsylvania State University, USA, e-mail: `wlee@cse.psu.edu`

John Krumm
Microsoft Research, Redmond, WA, USA, e-mail: `jckrumm@microsoft.com`

Ke Deng
The University of Queensland, Brisbane, Australia,
e-mail: `dengke@itee.uq.edu.au`

Kexin Xie
The University of Queensland, Brisbane, Australia,
e-mail: `kexin@itee.uq.edu.au`

Kevin Zheng
The University of Queensland, Brisbane, Australia,
e-mail: `kevinz@itee.uq.edu.au`

Xiaofang Zhou
The University of Queensland, Brisbane, Australia,
e-mail: `zxf@itee.uq.edu.au`

Goce Trajcevski
Northwestern University, USA, e-mail: `goce@ece.northwestern.edu`

Mohamed F. Mokbel
University of Minnesota, USA, e-mail: `mokbel@cs.umn.edu`

Chi-Yin Chow
City University of Hong Kong, China, e-mail: `chiychow@cityu.edu.hk`

Hoyoung Jeung
École Polytechnique Fédérale de Lausanne, Switzerland,
e-mail: `hoyoung.jeung@epfl.ch`

Man Lung Yiu
Department of Computing, Hong Kong Polytechnic University, Hong Kong,
e-mail: `csmlyiu@comp.polyu.edu.hk`

Christian S. Jensen
Department of Computer Science, Aarhus University, Denmark,
e-mail: `csj@cs.au.dk`

Yin Zhu
Hong Kong University of Science and Technology, China,
e-mail: `yinz@cse.ust.hk`

Vincent Wenchen Zheng
Hong Kong University of Science and Technology, China,
e-mail: `vincentz@cse.ust.hk`

Qiang Yang
Hong Kong University of Science and Technology, China,
e-mail: `qyang@cse.ust.hk`

Yu Zheng
Microsoft Research Asia, China, e-mail: `yuzheng@microsoft.com`

Xing Xie
Microsoft Research Asia, China, e-mail: `xing.xie@microsoft.com`

Editorial Board
Ralf Hartmut Güting
University of Hagen, Germany, e-mail: `rgh@fernuni-hagen.de`

Hans-Peter Kriegel
Institut für Informatik, Ludwig-Maximilians Universität München, Germany,
e-mail: `kriegel@dbs.ifi.lmu.de`

Hanan Samet
University of Maryland, USA, e-mail: `hjs@cs.umd.edu`