On the Similarity of Motions

Goce Trajcevski
Dept. of Electrical Engineering and Computer Science
Northwestern University
Evanston, IL 60208
email: goce@eecs.northwestern.edu

Dimitrios Gunopulos
Dept. of Informatics and Telecommunications
University of Athens
Athens, Greece
email: dg@di.uoa.edu

I. INTRODUCTION
Advances in the miniaturization of the computing devices and networking have proliferated various sensors that, in addition to measuring different values, can self-organize in networks upon deployment in different environments – from laboratories, through wildlife, to roadside. These, along with the technological advances in remote (satellite) imaging and GPS technologies, have enabled generation of large volumes of spatiotemporal datasets which store the (location, time) information pertaining to various mobile entities.

The main motivation for this tutorial is that one particular property of the mobile entities – their similarity, has been of interest in wide range of application domains where some form of location-based awareness is needed.

Research works dealing with similarity of motions in different application domains abound – from similarity of geographic and seismic processes and events, through categorization of the physical properties of similar movement characteristics of objects in everyday-life, similarities of the motions at the level of molecules/particles, to similarities of the time-series for financial data. Similarity of time-varying data is at the heart of the methodologies for indexing, classification, clustering and approximation of time series.

In the recent years, various aspects of the similarity of trajectories of moving point objects have been extensively investigated in Moving Objects Databases and Geographic Information Systems settings.

The quest for effectiveness and efficiency is a common thread to all the application domains and, when it comes to similarity, the semantics of each individual domain dictates one of the most important characteristics which, in turn, has a crucial impact on the methodologies and the algorithms used. That feature is the distance function used for evaluating how (dis)similar are two particular trajectories.

The objective of this tutorial is to give an overview of the different facets of the problem of detecting the similarity among moving objects’ trajectories. Specifically, we will present different techniques for assessing the similarity and identify the implications that different applications context(s) have on the specific approaches. We will also discuss the impact of the similarity-based mining/clustering of spatiotemporal data on applications that have large societal impacts, such as efficient traffic management and disaster remediation.

II. TUTORIAL OUTLINE
Following are the main themes addressed in this tutorial

Motivation: Applications Settings
We motivate the importance of the problem of efficiently detecting the similarity of mobile entities by presenting its relevance in several applications, and we identify what are the key characteristics of the participating entities.

Similarity of Geometric Entities
One of the important insights is that the similarity can be evaluated subject modulo certain transformations – i.e., rotation, translation or scaling, which can be applied to one set for the purpose of assessing its similarity to another. An important benefit of the research works on similarity of geometric shapes was the realization of the importance of the distance function.

Time-Series Data
The problem of efficient evaluation of similarity is central to the management of time series data, and the two important concepts towards that end are the representation methods, and the similarity measures.

Similarity of Moving Objects Trajectories
Unlike the time-series data, the motion of the objects in many application domains occurs in 2 (and higher) dimensions, and a large body of works have have focused on different aspects of the similarity problem in such settings. An important problem for similarity for multidimensional trajectories is the robustness of the underlying techniques, and special settings include motions on road network.

Impact of Similarity Techniques
We overview the significance of the similarity of motion in application domains such as: – clustering of moving objects trajectories and detecting anomalies; – identifying flocks and convoys of trajectories; – mining and warehousing of trajectories data.

*Research Supported by the NSF – CNS 0910952
MDM 2010 Seminars

Seminar 1: On the Similarity of Motions

Goce Trajcevski, Northwestern University
Dimitrios Gunopulos, University of Athens, Greece

Goce Trajcevski: Goce Trajcevski received his B.Sc. degree from the University of Sts. Kiril i Metodij, and his MS and PhD degrees from the University of Illinois at Chicago. His main research interests are in the areas of spatio-temoral data management, routing and data management in wireless sensor networks, and reactive behavior in dynamic systems. He has (co)authored over 40 publications in refereed conferences and journals and received a Best Paper Award at the CoopIS 200 conference. He is currently an Assistant Chairman with the Department of Electrical Engineering and Computer Science at the Northwestern University.

Dimitrios Gunopulos: Dimitrios Gunopulos is an Associate Professor in the Department of Informatics and Telecommunications, University of Athens. He got his PhD from Princeton University in 1995. He has held positions at the Max-Planck-Institut for Informatics, the IBM Almaden Research Center, and at the Department of Computer Science and Engineering in the University of California Riverside. His research is in the areas of Data Mining, Knowledge Discovery in Databases, Databases, Sensor Networks, Peer-to-Peer systems, and Algorithms. He has co-authored over a hundred journal and conference papers that have been widely cited and a book. His research has been supported by NSF (including an NSF CAREER award), the DoD, the Institute of Museum and Library Services, the Tobacco Related Disease Research Program, the European Commission, AT&T, and Nokia. He has served as a General co-Chair in the IEEE ICDM 2010 conference, as a PC co-Chair in the ECML/PKDD 2011, IEEE ICDM 2008, ACM SIGKDD 2006, SSDBM 2003, and DMKD 2000 conferences, and as an associate Editor at the IEEE TKDE, IEEE TPDS, and ACM TKDD journals.
Seminar 2: Trajectory Warehousing and Mining:
A State-of-the-Art and Research Directions
Alejandro Vaisman, Universidad de Buenos Aires

Moving object data (MOD) representation and computing have received a fair share of attention over recent years from the database and GIS (Geographic Information Systems) communities. Moving object positions at different time instants are captured by means of electronic devices, resulting in the generation of huge amounts of trajectory data. The analysis of such trajectory data raises opportunities for discovering behavioral patterns that can be exploited in applications like traffic management and service accessibility. Online analytical processing (OLAP) and data mining (DM) techniques have been employed in order to convert this vast amount of raw data into useful knowledge. Different approaches have been followed. One of them consists in applying data warehousing modeling and implementation techniques in order to build so-called trajectory data warehouses from which useful information could be obtained. Data Extraction, Transformation and Loading (ETL) techniques are intensively applied to build these data warehouses. Another use of these large amounts of trajectory data is aimed at automatically (and efficiently) discovering spatio-temporal patterns. Central to this approach is the ability to replace raw trajectory data by sequences of application-dependent stops occurred at so-called places of interest (PoIs), leading to the notion of semantic trajectories. Different techniques and languages have been proposed for sequential pattern analysis of trajectories defined in this way.

In the first part of the tutorial we will address the topic of trajectory data warehouses and its applications. In the second part we will discuss and compare different mining methods for obtaining trajectory patterns, as well as the use of other data mining techniques (like clustering and classification) for MOD analysis. Finally, we will present and propose possible research directions in the field.

Prof. Alejandro Vaisman
Alejandro Vaisman was born in Buenos Aires (UBA), Argentina. He received a BA degree in Civil Engineering, a BA in Computer Science, and a PhD in Computer Science from UBA, and he has been a post-doctoral researcher at the University of Toronto. He is a Professor at UBA since 1994. He was Vice-Dean of the School of Engineering and Information Technology at the University of Belgrano, Argentina in 2001 and 2002, and Vice-Head of the Computer Science Department at UBA. He was a visiting researcher at the University of Toronto, Universidad Politecnica de Madrid, University of Hasselt, and Universidad de Chile. His research interests are in the field of databases, particularly in OLAP, Data Mining, and Geographic Information Systems, XML and the Semantic Web. He has authored and co-authored several scientific papers presented in major database and GIS conferences and journals. Since 2005 he chairs the Masters Program in Data Mining at the Computer Science Department, UBA. In 2008 he worked for the University of Hasselt in the GeoPKDD project. He was the Director of the project “Using OLAP Techniques in Geographical Information Systems”, funded by the Argentinian Scientific Agency.