AIN SHAMS UNIVERSITY
Faculty of Computer
& Information Sciences
Information Systems Department

Semantically Enhanced Location-based Social Networks

Thesis submitted to the Department of Information Systems Faculty of Computer and Information Sciences Ain Shams University, Egypt
In fulfillment of the requirements for the Master of Science degree, MSc, of Computer and Information Sciences

By

Basma Hassan AlBanna

Teaching Assistant, Information Systems Department,
Faculty of Computer and Information Sciences, Ain Shams University

Under the Supervision of

Dr. Ibrahim Fathy Moawad

Associate Professor, Information Systems Department,
Faculty of Computer and Information Sciences, Ain Shams University

Dr. Sherin Mohamed Moussa

Associate Professor, Information Systems Department,
Faculty of Computer and Information Sciences, Ain Shams University

Dr. Mahmoud Attia Sakr

Assistant Professor, Information Systems Department,
Faculty of Computer and Information Sciences, Ain Shams University

December 2016

Table of Contents

List of Figures

List of Tables

List of Abbreviations

ADT: Abstract Data Type

CF: Collaborative Filtering

DCG: Discounted Cumulative Gain

GIS: Geographic Information Systems

GPS: Global Positioning System

HITS: Hypertext Induced Topic Search

HMM: Hidden Markov Model

IALBR: Interest Aware Location-based Recommender

ICFF: Implicit-feedback Content-aware Collaborative Filtering

IDCG: Ideal Discounted Cumulative Gain

LBSN: Location Based Social Network

MADS: Model Analysis Decision Support

MEA: Mean Absolute Error

MGM: Multi-center Gaussian model

NDCG: Normalized Discounted Cumulative Gain

POI: Points of Interest

SMoT: Stop and Moves of Trajectories

SN: Social Network

STT: Spatio-temporal Trajectory

Abstract

Trajectory data analysis has recently become an active research area. This is due to the large availability of mobile tracking sensors, such as GPS-enabled smart phones. However, those GPS trackers only provide raw trajectories (x, y, t), ignoring information about the geographical locations, transportation mode, etc. This information can contribute in producing significant knowledge about movements, which transforms raw trajectories into semantic trajectories. Therefore, research lately has focused on semantic trajectories; their representation, construction, and applications. Furthermore, advances in location acquisition and mobile technologies also led to the addition of the location dimension to Social Networks (SNs) and to the emergence of a newer class called Location-based Social Networks (LBSNs). One of the key applications of semantic trajectories is location-based recommendation, which is a main function of LBSNs.

This research investigates the current studies on semantic trajectories so far. We propose a new classification schema for the research efforts in semantic trajectory construction and applications. The proposed classification schema includes three main classes: semantic trajectory modeling, computation, and applications. Additionally we proposed a methodology to semantically enhance LBSNs through extracting SN Geo-tagged media annotations and using them as location semantics. This enabled

us to introduce an Interest Aware Location-based Recommender System (IALBR) which combines the advantages of both LBSNs and SNs, in order to provide interest aware location-based recommendations. This recommender system is proposed as an extension to LBSNs. It is novel in: 1) utilizing the Geo-content in both LBSNs and SNs, 2) ranking the recommendations based on a novel scoring method that maps to the user interests. It also works for passive users who are not active content contributors to the LBSN. This feature is critical to increase the number of LBSN users. Moreover, it helps in reducing the cold start problem, which is a common problem facing the new users of recommender systems who get random unsatisfying recommendations. This is due to the lack of user interests awareness, which is reliant on user history in most of the recommenders.

We evaluated the IALBR system with a large-scale real dataset collected from foursquare in respect of precision, recall & f-measure. We also compared the results with yelp, as a ground truth system, using metrics like the Normalized Discounted Cumulative Gain and the Mean Absolute Error. In comparison to the baseline (i.e.Foursquare), the IALBR recommended on average 3 times more venues with a precision of 80% and achieved an F-measure of 0.87 (at N=15). In comparison to Yelp ratings, the IALBR scored an MAE of 1.3 and an NDCG of 0.9.

List of Publications

1. AlBanna, B. H., Moawad, I. F., Moussa, S. M., Sakr, M. A. (2015). Semantic Trajectories: A Survey from Modeling to Application. In *Information Fusion and Geographic Information Systems (IFGIS'2015)* (pp. 59-76). Springer International Publishing.

2. AlBanna, B., Sakr, M., Moussa, S., Moawad, I. (2016). Interest Aware Location-Based Recommender System Using Geo-Tagged Social Media. *ISPRS International Journal of Geo-Information*, 5(12), 245.

Acknowledgment

Thanks merciful GOD for all your gifts in my whole life. I would like to thank my supervisors Dr. Ibrahim Fathy, Dr. Sherin Moussa and Dr. Mahmoud Attia who really supported and guided me greatly during my study and research. Their comments and suggestions helped me not only in this work, but also in my career. It was my pleasure and honor working with them.

I am also grateful to all my colleagues for their continuous support and help. Finally, with all my appreciation and love that no words can express, I dedicate this master's to my lovely family and my friends who always provided me with love, prayers, blessings, advice and care. They are behind every success in my life.

1 INTRODUCTION

1.1 Background

Around 80% of all the available data have either an explicit or an implicit Geographical reference [1]. Explicit references are the actual Geometries e.g., city boundaries, lakes, whereas implicit references are textual references to Geographical objects e.g., street names, city names, etc. There are objects that change their spatial reference with time, or so-called spatiotemporal objects. With the advancement of the current GPS technologies, large-scale capture of motion of those moving spatiotemporal objects became attainable. Typical examples of moving objects include cars and persons equipped with a GPS device, or animals wearing a transmitter whose signals are captured by satellites [2]. Understanding why and how people and animals move, which places they visit and for which purposes, what are their activities, and which resources they use, is of great importance for decision making in a variety of applications. Examples of those applications are location based recommenders, road traffic monitoring, mobile health and animal data ecology which all call for methods enabling rich and expressive representation of moving object activities.

1.1.1 Semantic Trajectories

Semantic trajectories is a growing trend that has recently emerged in Geographic information science and spatiotemporal knowledge discovery. It is mainly concerned with understanding the motion of the moving object with respect to the application of interest. Moving objects generate movement tracks over periods of time. A movement track is the motion history of the moving object, containing the spatial values that change with time. Trajectories are the segments of the object's movement track that are of interest for a given application [3]. A raw trajectory is a trajectory extracted from a raw movement track containing no contextual information that reveals motion semantics. It is a sequence of spatiotemporal observations (x, y, t) using Geodetic coordinates. It doesn't contain background contextual information (e.g., transportation means and Geographical objects) that can contribute significant semantic knowledge about movements. Semantics refer to contextual information available about the moving object and the Geographical objects it comes across as it moves, apart from its mere position data. Semantic is contained both in the Geometric properties of the spatiotemporal stream (e.g., when the user stops/moves) as well as in the Geography on which the trajectory passes (e.g., shops, roads). An example of a semantically enriched trajectory could be the following:

(Begin, home, 9am)→ (move, road, 9am-10am, on-bus)→(stop, office, 10am-5pm, work)→(move, road, 5pm - 5:30 pm, on-metro)→(stop, market, 5:30pm-6pm, shopping) →(move, road, 6pm - 6:20 pm, walking)→(End, home, 6:20pm) [4]

Semantic trajectories are able to preserve motion attributes i.e., place, activity, transportation mean, etc. Adding semantics enhances the analysis of data and facilitates the discovery of semantically implicit patterns and behaviors. Semantic enrichment of trajectories happens through embed-