Information Systems Department
Faculty of Computer & Information Sciences
Ain Shams University

Enhancing Tracking Techniques in Social Networks

Thesis Submitted in Partial Fulfillment of the Requirements for the Degree of PhD in Computer and Information Sciences
to
Department of Information Systems
Faculty of Computer and Information Sciences
Ain Shams University

by
Ola AlSayed Mostafa Omar AlSenosy
Assistant Lecturer
Information Systems Department
Faculty of Computer and Information Sciences
Ain Shams University, Cairo, Egypt

Under Supervision of
Prof. Dr. Mohamed Hashem Abdalaziz
Professor, Information Systems Department
Faculty of Computer and Information Sciences
Ain Shams University, Cairo, Egypt

Prof. Dr. Hossam Al-Deen Mostafa Faheem
Professor, Computer Systems Department
Faculty of Computer and Information Sciences
Ain Shams University, Cairo, Egypt

Prof. Dr. Nagwa Lotfy Badr
Professor, Information Systems Department
Faculty of Computer and Information Sciences
Ain Shams University, Cairo, Egypt

2017

Abstract

Understanding business behaviors requires acquiring huge amounts of data from diverse field studies. The additive growing use of mobile devices in social media, especially in recent years, provides large amounts of data transactions that can help in understanding business behaviors replacing the data acquired by the exhaustive field studies. Location Based Social Networks (LBSN) are considered as a solution providing such data used in urban analysis for economic reasons.

Towards more insight for business behavior in this dissertation, a suggestion of global perspective exploiting data collected from LBSNs is introduced in order to predict business behavior according to the business geographical location. Moreover, business behavior prediction in LBSNs is studied in this research for big data application. Prediction of customers' presence rates for business venues is introduced to be implemented using machine learning techniques. Machine learning techniques are investigated for both static and dynamic business predictions in LBSNs. Spatial regression models are thoroughly presented as static machine learning techniques. A comparative study is attained in this dissertation for suitability to model the relationships in LBSNs in order to be used for prediction. Geographically Weighted Regression (GWR) model proved to be the appropriate model in handling the sparse geographical distribution imposed by the LBSNs data. A proposed enhancement over the GWR model is introduced through a distributed training process that is integrated into a partitioned-GWR architecture. The proposed architecture includes a three blocks processes that are designed to deal with LBSNs data heterogeneity pursuing more enhanced predictions for business behavior.

For dynamic business predictions, spatial interpolation techniques are unprecedentedly proposed to be used in LBSNs. Enhancements over the spatial interpolation techniques are proposed in this study. A Local Filtered (LF) spatial interpolation is proposed to handle the data instabilities occurrences in LBSNs, to increase business prediction accuracy. As a second enhancement, a Similarity Embedded (SE) spatial interpolation is introduced to consider the diversity of features provided by the LBSNs' data in the prediction process. The proposed SE spatial interpolation suggested a hybrid feature similarity and distance

measurements involvement in the prediction process. Moreover, a design of a Filtered Similarity Embedded (FSE) spatial interpolation is proposed. The FSE spatial interpolation pursue additive accurate prediction results through a fusion of the two previously proposed spatial interpolation enhancements; LF spatial interpolation and SE spatial interpolation.

Since big data analytics are one of the most important topics of social media related research nowadays, an Iterative Nearest Neighbors First (INNF) search method is designed for timely efficient implementation of the proposed interpolation techniques over big datasets. The proposed INNF search method design the timely efficient solution using the geo indexing property provided by Not Only Structured Query Language (NOSQL) big databases.

For assessing both prediction accuracy and implementation time efficiency of the previously proposed techniques, extensive experiments are implemented over data extracted from Foursquare. Data is collected and analyzed for observation about venues registered in Foursquare residing in Texas State in the United States of America. The experiments results show noticeable improvements for prediction accuracies of the spatial interpolation enhancements proposed in this study. The LF spatial interpolation increases the prediction accuracies with 61% and 54% compared to the classical K Nearest Neighbors (KNN) and Inverse Distance Weighting (IDW) spatial interpolation techniques, respectively. While the SE spatial interpolation increases the prediction accuracies with 44% and 37% compared to the KNN and IDW spatial interpolation techniques, respectively. Moreover, the FSE spatial interpolation increases the prediction accuracies with 67% and 63% compared to the KNN and IDW spatial interpolation techniques, respectively.

The proposed INNF search method shows remarkable improvements in experiments. The INNF search method reduces the implementation runtime of the SE spatial interpolation in terms of milliseconds compared to the tens of seconds classical implementations in the experiments attained over the Foursquare dataset. Moreover the INNF search method shows steady runtimes for successive implementations over increased sized several synthetic datasets.

List of Publications

1. Ola Al Sonosy , Sherine Rady , Nagwa Lotfy Badr , Mohammed Hashem, “Exploiting location based social networks in business predictions”, Proceedings of the 11th International Conference on Innovations in Information Technology (IIT '15), Cairo, Egypt, p.40-50, Nov 2015, DOI:10.1109/INNOVATIONS.2015.7381512.
2. Ola Al Sonosy , Sherine Rady , Nagwa Lotfy Badr , Mohammed Hashem, "Machine Learning Techniques for Mining Location-based Social Networks for Business Predictions", The 10th International Conference on Informatics and Systems (INFOS ’16), Cairo, Egypt, May 2016, DOI: 10.1145/2908446.2908475.
3. Ola Al Sonosy , Sherine Rady , Nagwa Lotfy Badr , Mohammed Hashem, “Business Behavior Predictions using Location Based Social Networks in Smart Cities”, accepted for publication in ebook of Information Innovation Technology in Smart Cities, Springer Singapore, 2017, ISBN 978-981-10-1741-4.
4. Ola Al Sonosy , Sherine Rady , Nagwa Lotfy Badr , Mohammed Hashem, “Towards Efficient Business Behavior Prediction using Location Based Social Networks”, submitted to Wires Data Mining and Knowledge Discovery Journal. Impact factor: 1.759.
5. Ola Al Sonosy , Sherine Rady , Nagwa Lotfy Badr , Mohammed Hashem, “A Study of Spatial Machine Learning for Business Behavior Prediction in Location Based Social Networks”, 11thIEEE International Conference on Computer Engineering and Systems (ICCES 2016), Cairo, Egypt, p. 266 - 272, December 2016, DOI: 10.1109/ICCES.2016.7822012.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

CLAF	Collaborative Location and Activity Filtering
CSE-Tree	Compressed Start End Tree
FSE	Filtered Similarity Embedded
GIS	Geographic Information Systems
GLS	Generalized Least Squared
GPS	Geographical Positioning Systems
GWR	Geographically Weighted Regression
HGBSM	Hierarchical Graph Based Similarity Measurement
IDF	Inverse Document Frequency
IDW	Inverse Distance Weight
INNF	Iterative Nearest Neighbor First
k-BCT	k Best-Connected Trajectory
KNN	K Nearest Neighbors
LARS	Location Aware Recommender System
LBS	Location Based Services
LBSN	Location Based Social Network
LDA	Latent Dirichlet Allocation
LF	Local Filtered
MTM	Maximal Travel Match
NOSQL	Not Only Structured Query Language
OLS	Ordinary Least Squared
PCLAF	Personalized Collaborative Location and Activity Filtering
POI	Points Of Interests
RAE	Relative Absolute Error

RPCLAF	Ranking-based Personalized Collaborative Location and Activity Filtering
SAR	Spatial Auto Regression
SDM	Spatial Durbin Model
SE	Similarity Embedded
SEM	Spatial Error Model
SErr	Standard Error
SLH	Semantic Location History
SLX	Spatial Lag X
TBHG	Tree Based Hierarchical Graph
WCH	Weighted Category Hierarchy

Chapter 1

Introduction

With the recent improvements of mobile devices manufacturing and location acquisition technologies, a huge number of location-based data transactions has been available for data scientists. These amounts of location related data is forming a very rich soil for data analytics and realizing new location aware services. Various large scale location based data, including user generated data in location based social networks or geographical positioning systems data reported from vehicles or sensors, have led to research challenges and opportunists to provide location based services, intelligent transportation systems, geographic information systems, urban analysis reports, and smart cities services.

Different computer science techniques, like data mining, machine learning, artificial intelligence, and spatial or spatio-temporal databases, are used or combined to address such challenges.

1.1. Motivation

The rapid technological advances of the mobile networks, social media and the fact that mobile devices with Geographical Positioning Systems (GPS) modules are now been used everywhere, motivate social networks users to share information about their real movements as online transactions, by checking in places or venues especially in Location Based Social Networks (LBSNs) such as Foursquare and Facebook Places. LBSNs have been attracting more and more users by providing services that integrate social activities with geographic information. In

LBSNs, a user can discover places of interests around his current location, check-in at these places and also share his/her check-in information and opinions about places with his/her online friends.

Over years of heavy use, LBSNs have accumulated very large amounts of information related to business activities along with their associated geographical information. Unprecedentedly, urban analysis can be used for mining LBSN's originated datasets to infer information that enables more understanding about business behavior. Business behavior understanding can be beneficial to business owners for making business related decisions like opening new business places, initiating advertisements campaigns, improving provided services qualities or even closing business decisions.

1.2. Objectives

The main objective of this study is to introduce a global comprehension of LBSNs' data to serve business owners. As, LBSNs accumulated with large rates through time, a global analysis of these huge amounts of data can give a deep insight about business behavior in geographical terrains. This dissertation includes a suggestion of quantitative analytical study to provide conclusions of the business behavior and patterns for business owners and decision makers. The study aims to exploit the huge amounts of information available about users' check-ins for business venues on the cumulative LBSNs datasets in predicting users' number of check-ins for venues in order to understand the business behavior.

As data originated by social media are accumulated in millions of records, therefore, methods and algorithms that will be used in analyzing LBSNs data has to be assessed for big data efficient implementations.

1.3. Research Contributions

The research study of this dissertation can be summarized in the following contributions:

- Investigation of spatial regression models as static prediction methods to be applied in LBSNs' data. Several presented models are discussed in a scientific comparative study for suitability for LBSNs' data. The study proved the Geographically Weighted Regression (GWR) as the model the suitable for presenting the relationships of data originated by LBSNs. This is presented in chapter 4, sections 4.3.1- 4.3.6.
- Proposing a distributed training GWR model to deal with the data heterogeneity imposed by the LBSNs' diverse features characterizing their entities. The proposed model is integrated into a design of a proposed partitioned-Geographically Weighted Regression architecture to be used for predictions in LBSNs. This is presented in chapter 4, section 4.3.7.
- Unprecedented introduction of Spatial Interpolation techniques to be used for predictions in LBSNs. The spatial interpolation techniques are used as dynamic prediction techniques which fits the changeable nature of features describing entities in LBSNs. This is presented in chapter 4, section 4.4.1.
- A proposal of a Local Filtered spatial interpolation enhancement is introduced in this study. The proposed method provides enhanced predictions in LBSNs by decreasing the negative effect of the local extremes that are detected in the LBSNs' data. This is presented in chapter 4, section 4.4.2.