DATA SOURCING TECHNIQUES FOR PERSONALIZED LOCATION BASED TOURIST SPOT RECOMMENDER SYSTEM

SUNITA TIWARI

AMARNATH & SHASHI KHOSLA SCHOOL OF INFORMATION TECHNOLOGY

INDIAN INSTITUTE OF TECHNOLOGY DELHI

JANUARY 2018

DATA SOURCING TECHNIQUES FOR PERSONALIZED LOCATION BASED TOURIST SPOT RECOMMENDER SYSTEM

by

SUNITA TIWARI

Amarnath and Shashi Khosla School of Information Technology

Submitted

in fulfillment of the requirements of the degree of Doctor of Philosophy

to the

INDIAN INSTITUTE OF TECHNOLOGY DELHI

JANUARY 2018

Certificate

The thesis entitled **"Data Sourcing Techniques for Personalized Location Based Tourist Spot Recommender System"** being submitted by **Ms. Sunita Tiwari** to the Indian Institute of Technology Delhi for award of the degree of **Doctor of Philosophy** is a record of original bonafide research work carried out by her. She has worked under my guidance and supervision, and has fulfilled the requirements for the submission of this thesis, which has attained the standard required for a Ph.D. degree of this institute.

The results presented in this thesis have not been submitted elsewhere for the award of any other degree or diploma.

Dr. Saroj Kaushik
Professor
Department of Computer Science and Engineering
Indian Institute of Technology Delhi

New Delhi

Acknowledgements

I would thank the Almighty for giving me the courage and determination to pursue the higher education in the field of Computer Science and Engineering. I would like to tender my heartfelt thankfulness to my supervisor **Prof. Saroj Kaushik** for her priceless supervision, continuous inspiration and enthusiastic support in every stage of the work. Her deep insight into problems, imaginative ideas, technical guidance and sufficient encouragement was a constant source of inspiration and motivation. Her critical comments on my technical writing, eye for detail and persistent encouragement strongly motivated me to strive hard to achieve the high targets set by her in completion of this thesis. Her religious, humanitarian and positive outlook towards the life has been a great source of inspiration even in the non-academic side of life.

I am extremely grateful to the members of my Student Research Committee (SRC) – Prof. M P Gupta, Prof. Aditeshwar Seth, Prof. Punam Bedi, who have been very helpful by offering suggestions and advice. I thank my Ph.D. colleagues, especially, Priti Jagwani, Shivendra Prasad Tiwari, and Kuntal Dey for their help on reviewing my work and offering useful suggestions. I also extend the sincerest gratitude to the summer interns at IIT Delhi Chhavi and Akash and most importantly my students who contributed in conducting experiments by registering in the proposed system and providing their valuable feedback. My sincere gratitude to my colleagues at work place for providing the unconditional support.

My Grandparents, my parents Mr. O. P Sharma and Mrs. Girija Sharma and my husband Sourabh Tiwari and other family members showed immense patience and provided me great support during the course of my work. Their patience, sacrifice, inspiration and moral support are of special mention. I dedicated my thesis to them.

Sunita Tiwari

Abstract

Research in recommender system has matured in the last two decades. The emphasis in recent research is more focused on Location Based Services (LBS) due to the advancement of wireless communication devices and location acquisition technologies like Global Positioning Systems (GPS). The mobile and handheld devices are the primary mode of communication and information access in this era. The research issues in the field of recommender system for mobile devices are more challenging as compared to traditional web based recommender systems. The techniques applied to the web based recommender system cannot be straight away used for such devices because of several reasons such as dissemination of relevant information in small screen as the size of devices are small, dissemination of information in real time as the users are mobile , etc.

The proposed research work aims to design and develop various techniques required for personalized location based recommender systems. Such systems will increase the usability of mobile device to many folds during their journey. Tourism is one of the largest industries in the world. We restrict our focus to tourism domain in order to facilitate the users with the interesting tourist spots, enriched tour details and so on.

The research work broadly addresses the following challenges in the field of the location based tourist spot recommender systems.

Firstly, it addresses the problem of automatic discovery of new popular tourist spots in a geographical region and semantic annotation of point of interests. Location semantics are gathered by crawling the web in a focused way using domain ontology. Our experiment shows that 20-23% of the locations retrieved from the hierarchical graph based approach presented in (Zheng Y. a.-Y., 2009) and relational algebra based approach presented in (Khetarpaul, 2011) are not interesting location from tourist point of view.

Secondly, the problem of automatic discovery of user preferences and personalized location based recommendation is discussed. Past travel histories and genetic algorithm based approach is exploited for the personalization problem of the tourist spot recommender system. The proposed approach is compared with Matrix Factorization based approach presented in (Berjani, 2011) and the proposed

approach shows an improvement in average RMSE of (approx.) 4.63%. The recommendation accuracy is 94% (approx.).

The third problem is to propose, the semantic user similarity measure for recommender system. The proposed approach for user similarity is based on semantic stay trajectories using Earth Mover's Distance (EMD). The proposed approach performs better than the popular user similarity measures such as Pearson's correlation, Jacard and Dice. It shows an average percentage improvement of 10.7% in RMSE and 5.73% in MAE as compared to the above mentioned approaches.

Finally, the problem of Information enrichment of recommender system is addressed. This technique used the concept of location based crowdsourcing with the aim of improving the quality of the recommendation. Here, fuzzy inference system is used to convert the contextual information obtained from the crowd to an appropriate rank of the recommended spot. The experimental evaluation demonstrates the better satisfaction level of users with enriched information.

सार

अनुशंसात्मक प्रणाली (Recommender System) में शोध पिछले दो दशकों में परिपक्व हो गया है। हाल ही में शोध में बेतार संचार उपकरणों और स्थान अधिग्रहण तकनीकों जैसे वैश्विक स्थिति प्रणाली(GPS) की उन्नति के कारण स्थान आधारित सेवाओं(LBS) पर अधिक ध्यान केंद्रित किया गया है। इस युग में मोबाइल और हाथ में धारण करने योग्य उपकरण संचार और सुचना प्राप्ति के प्राथमिक तरीके है । मोबाइल/चल उपकरणों के क्षेत्र में अनुसंधान के मुद्दे पारम्परिक वेब आधारित अनुशंसात्मक तकनीक की तुलना में अधिक चुनौतीपूर्ण हैं। इन कारणों में छोटे/लघु स्क्रीन वाले उपकरण और गतिशील उपयोगकर्ता के पास वास्तविक समय में जानकारी को प्रसार कर पाना सम्मिलित है।

प्रस्तावित अनुसंधान कार्य का उद्देश्य व्यक्तिगत स्थान आधारित अनुशंसात्मक प्रणाली हेंतु विभिन्न तकनीकों का प्रारूपण और विकसित करना है। इस प्रकार की प्रणाली मोबाइल उपकरण की प्रयोज्यता को उसकी विकास यात्रा के दौरान कई गुणा तक बड़ा देगी । पर्यटन दुनिया के सबसे वृहत्तम उद्योगों में से एक है। हम पर्यटन के क्षेत्र में उपयोगकर्ता को दिलचस्प पर्यटन स्थलों की जानकारी देने और पर्यटन से सबंधित विवरण से उन्हें समृद्ध करने के लिए शोध को केंद्रित करते हैं।

शोध कार्य मुख्य तौर पर स्थान आधारित पर्यटन स्थल अनुशंसात्मक प्रणाली के क्षेत्र में अधोलिखित चुनौतियों को सम्बोधित करता है।

सबसे पहले, यह एक भौगोलिक क्षेत्र में नए लोकप्रिय पर्यटन स्थलों की खोज और पर्यटन स्थलों की अर्थिक टिप्पणियों की स्वतः खोज की समस्या को संबोधित करता है। स्थानीय शब्दार्थों को domain ontology का उपयोग करते हुए एक केंद्रित पद्धति से वेब को क्रॉल करके इकट्ठा किया जाता है। हमारा प्रयोग बताता है कि (Zheng Y. a.-Y., 2009) में प्रस्तुत पदानुक्रमित ग्राफ आधारित दृष्टिकोण से और (Khetarpaul, 2011) में प्रस्तुत रिलेशनल बीजगणित पद्धति से प्राप्त 20 से 23% स्थान पर्यटन के लिए दिलचस्प नहीं है।

दूसरी बात , उपयोगकर्ता की प्राथमिकताओं की स्वतः खोज और व्यक्तिगत स्थान आधारित अनुशंसा की स्थापना की समस्या पर चर्चा की गई है। विगत यात्रा विवरण और आनुवंशिक विधि (algorithm) पर आधारित तकनीक का उपयोग पर्यटन स्थल अनुशंसा करने वाली प्रणाली की निजीकरण समस्या के समाधान के लिए किया जाता है।

प्रस्तावित पद्धति की तुलना मैट्रिक्स फैक्टोरिजेशन आधारित पद्धति (Berjani, 2011) से की गई है और प्रस्तावित विधि में औसत RMSE (लगभग) 4.63% का सुधार दर्शाता है। अनुशंसात्मक परिशुद्धता लगभग 94% है।

तीसरी बात, अनुशंसात्मक तकनीक हेतु सार्थक उपयोगकर्ता समानता उपाय प्रस्तावित करना है ।

उपयोगकर्ता सादृश्य हेतु प्रस्तावित प्रणाली सिमेंटिक ट्राजेक्टोरिज(semantic trajectories) और Earth Mover's Distance (EMD) तकनीक का उपयोग करती है। प्रस्तावित पद्धति लोकप्रिय उपयोगकर्ता समानता उपायों जैसे Pearson's Correlation Coefficient method, Jacard Method तथा Dice Method आदि से बेहतर प्रदर्शन कर रही है। उपर्युक्त विधियों की तुलना में प्रस्तावित विधि औसत रुट मीन स्क्वायर एरर (RMSE) में 10.7% और औसत मीन स्क्वायर एरर (MAE) में 5.73% की औसत प्रतिशत का सुधार दर्शाती है।

और अततोगत्वा अनुशंसात्मक प्रणाली की सुचना संवर्धन (Information Enrichment) की समस्या को समबोधित किया जाता है । इस तकनीक ने अनुशंसा की गुणवत्ता में सुधार लाने के उद्देश्य से स्थान आधारित क्रोवडसोर्सिंग (crowdsourcing) की अवधारणा का प्रयोग किया। यहां एक स्थल पर उपलब्ध भीड़ से प्राप्त प्रासंगिक पैरामीटर फजी निष्कर्ष प्रणाली(Fuzzy inference system) द्वारा उस पर्यटन स्थल के उपयुक्त स्तर(rank) प्राप्त करने के लिए उपयोग किया जाता है। प्रायोगिक मूल्याकन सूचना समृद्धीकरण (information enrichment) वाले अनुशंसात्मक प्रणाली के उपयोगकर्ताओं के बेहतर संतुष्टि स्तर को दर्शाते है।

Table of Contents

List of Figures

List of Tables

Acronyms

AGPS	Assisted Global Positioning System
API	Application Program Interface
CARS	Context Aware Recommender System
CBRS	Content Based Recommender System
CF	Collaborative Filtering
CFRS	Collaborative Filtering Recommender Systems
DBSCAN	Density Based Spatial Clustering of Applications with Noise
DS	Dataset
EMD	Earth Mover's Distance
GA	Genetic Algorithm
GIS	Geographic Information System
GPS	Global Positioning System
GSM	Global System for Mobile Communication
HGSM	Hierarchical Graph Based Similarity Measurement
HRS	Hybrid Recommender Systems
IES	Information Enrichment System
KBRS	Knowledge Based Recommender Systems
LBRS	Location Based Recommender Systems
LBS	Location Based Services
LBSN	Location Based Social Network
LCARS	location and context aware recommender system

MAE	Mean Absolute Error
MF	Matrix Factorization
PHP	Hypertext Preprocessor
POI	Point of Interest
RFID	Radio-frequency identification
RMSE	Root Mean Squared Error
RS	Recommender systems
SDK	Software Development Kit
TSRS	Tourist Spot Recommender Systems
URL	Uniform Resource Locator
VVIP	Very Very Important Person
WWW	World Wide Web
XML	Extensible Markup Language
PND	Personal Navigation Device