SPATIOTEMPORAL USER AND PLACE MODELLING ON THE GEO-SOCIAL WEB

A thesis submitted in partial fulfillment

of the requirement for the degree of Doctor of Philosophy

Soha Ahmed Ehssan Aly Mohamed

2017

Cardiff University
School of Computer Science & Informatics

Declaration

This work has not been submitted in substance for any other degree or award at this or any other university or place of learning, nor is being submitted concurrently in candidature for any degree or other award.

Signed . . . . (candidate)

Date . . . .

STATEMENT 1

This thesis is being submitted in partial fulfillment of the requirements for the degree of PhD.

Signed . . . . (candidate)

Date . . . .

STATEMENT 2

This thesis is the result of my own independent work/investigation, except where otherwise stated, and the thesis has not been edited by a third party beyond what is permitted by Cardiff University' s Policy on the Use of Third Party Editors by Research Degree Students. Other sources are acknowledged by explicit references. The views expressed are my own.

Signed . . . . (candidate)

Date . . . .

STATEMENT 3

I hereby give consent for my thesis, if accepted, to be available online in the University's Open Access repository and for inter-library loan, and for the title and summary to be made available to outside organisations.

Signed .... (candidate)

Date ....

To my role model, my father.
I hope I made you proud.

Abstract

Users of Location-Based Social Networks (LBSN) are giving away information about their whereabouts, and their interactions in the geographic space. In comparison to other types of personal data, location data are sensitive and can reveal user's daily routines, activities, experiences and interests in the physical world. As a result, the user is facing an information overload that overburdens him to make a satisfied decision on where to go or what to do in a place. Thus, finding the matching places, users and content is one of the key challenges in LSBNs.

This thesis investigates the different dimensions of data collected on LBSNs and proposes a user and place modelling framework. In particular, this thesis proposes a novel approach for the construction of different views of personal user profiles that reflect their interest in geographic places, and how they interact with geographic places. Three novel modelling frameworks are proposed, the static user model, the dynamic user model and the semantic place model. The static user model is a basic model that is used to represent the overall user interactions towards places. On the other hand, the dynamic user model captures the change of the user's preferences over time. The semantic place model identifies user activities in places and models the relationships between places, users, implicit place types, and implicit activities. The proposed models demonstrate how geographic place characteristics as well as implicit user interactions in the physical space can further enrich the user profiles. The enrichment method proposed is a novel method that combines the semantic and the spatial influences into user profiles. Evaluation of the proposed methods is carried out using realistic data sets collected from the Foursquare LBSN. A new Location and content recommendation methods are designed and implemented to enhance existing location recommendation methods and results showed the usefulness of considering place semantics and the time dimension when the proposed user profiles in recommending locations and content.

The thesis considers two further related problems; namely, the construction of dynamic place profiles and computing the similarity between users on LBSN. Dynamic place profiles are representations of geographic places through users' interaction with the places. In comparison to static place models represented in gazetteers and map databases, these place profiles provide a dynamic view of how the places are used by actual people visiting and interacting with places on the LBSN. The different views of personal user profiles constructed within our framework are used for computing the similarity between users on the LBSN. Temporal user similarities on both the semantic and spatial levels are proposed and evaluated. Results of this work show the challenges and potential of the user data collected on LBSN.

Acknowledgements

This thesis would not have been possible without the help and support of ALLAH. I thank ALLAH for giving me strength and patience to complete my thesis.

I am heartily thankful to my supervisor, Alia Abdelmoty for her continuous guidance and support throughout my research. I want to thank her for the fruitful and enjoyable discussions we had. It encouraged me a lot throughout my journey. Also, I would like to thank my second supervisor Chris Jones for his kind advices.

I would like to extend my thanks to my sponsor in Egypt, Helwan University, for the scholarship and the continuous support throughout the years of my study. I am also thankful to my friends and colleagues in the Faculty of Computers and Information for their encouragement and friendship. Also, special thanks and appreciation go to the UK Egyptian Cultural office for their help and support.

I would like to acknowledge the academic and technical support of the School of Computer Science & Informatics at Cardiff University, UK and all its staff members.

My deepest gratitude goes to my wonderful husband, Dr.Ahmed Eltanahy for his unlimited support, encouragement, and sacrifices. He was always keen to prioritise our family over himself. I am also graceful to my children Mohamed and Ali for their unconditional love, care, help and understanding.

I am thankful to my parents who spent endless efforts to make me happy all over past years. I also want to thank my sister and best friend Noha who was always beside me in every walk of my life.

Finally, I want to thank all my friends in Cardiff and in Egypt. (Shahd, Sahar, Fadia, Ghada, Naja, Maha, Dalia, Dr.Mona, Asmaa, Liqaa, Fatma, Dr.Haya, Dr.Shada, Dr.Hana, Dr.Neelam, Shelan, Aseela, Nermeen, Ashwan).

Contents

List of Publications

The work in this thesis has contributed to the following refereed publications:

[1] Mohamed, S. and Abdelmoty, A. 2013. Spatiotemporal Analysis of User- Generated Data on the Social Web. In GIS Research UK GISRUK 2013, University of Liverpool.

[2] Mohamed, S. and Abdelmoty, A. 2014. Dynamic place profiles from geo- folksonomies on the geosocial web. In: Research and Development in Intelligent Systems XXXI. Springer, pp. 239-251.

[3] Mohamed, S. and Abdelmoty, A. 2016. Uncovering user profiles in location-based social networks. Presented at: GEOProcessing 2016 : The Eighth International Conference on Advanced Geographic Information Systems, Applications, and Services, Venice, Italy, 24-28 April 2016. pp. 14-21.

[4] Mohamed, S and Abdelmoty, A. 2016. Computing similarity between users on location-based social networks. International Journal on Advances in Intelligent Systems 9 (3&4) , pp. 542-553.

List of Figures

List of Tables

List of Algorithms

List of Acronyms

API Application United Interface

GPS Global Positioning System

IBCF Item Based Collaborative Filtering

IDF Inverse Document Frequency

JSD Jensen-Shannon divergence

LBSN Location-based Social Networks

LDA Latent Dirichlet Allocation

LRS Location Recommendation Systems

POI Point of Interest

STT Spatio Temporal Topic

TGF Temporal Geo-Folksonomy

UBCF User Based Collaborative Filtering

UTC Coordinated Universal Time

Chapter 1

Introduction

1.1 Background and Motivation

Over recent years, social media has become very popular, and so the topic has attracted many researchers from a variety of fields. Social networks enable users to post their own, user-generated content at any time and from any location. Social media has evolved through the rise of the mobile internet and smart-phones, and it has become an essential part of many people's daily lives. Consequently, with the further emergence and growth of the use of GPS-enabled devices (along with wireless communication technologies) users can associate their geographical location with the content they post on the mainstream social networks (e.g., Twitter and Facebook). In addition, users can interact in relation to their locations with the use of location-based social networks such as Foursquare, Yelp, and Google Place. The emergence of LBSNs doesn't simply imply the addition of location information to existing social network technologies; it also implies changes to the structure of social networks, relating, as it does, users to the places they visit, their associated content and the time they visit the places. Using a LBSN, users check-in at different venues and share their life experiences relative to their various physical locations (past and present). This results in a significant amount of spatiotemporal data being created that, essentially, represents the "4Ws": who, when, where and what.

The content involved in this context includes tags, images, videos, comments, and other posts generated or uploaded by users onto a social media platform, and represents semantic information about users' activities. The temporal and spatial layers of data provide rich contextual information about users' activities. Also, to some extent, the dimension of location bridges the gap between online social networks and the physical world. Using data from location-based social networks, we can enhance situational awareness (what is going on around you) via the assembly of individual perspectives.

The users' movements through physical space, the places that they visit, the people that they interact with and the activities that they carry out all provide important clues about their personalities and personal preferences. Constructing a faithful user profile based on the information implied by the kind of interactions listed above has many potential applications. In particular, such information can be used to provide users with a more personalised search experience. This implicit data can also be used by recommendation services and by personalisation applications to provide users with information relevant to them. Its use can improve the functionality of recommendation services and lead to more targeted adverts for commercial services. The existence of such data can also improve the capabilities of location-based services, providing personalised access to local services.

On the other hand, studying the social-network traces left by numbers of people visiting particular places can enhance the understanding of the places visited and so help to build profiles of places over time. Understanding places and their use by people can help city authorities to understand the social dynamics of those places and so improve service provision and planning.

In addition, developing proficient location recommendation systems (LRSs) based on LBSNs becomes a more achievable task. In LRSs, the task of recommending places of interest to the user is referred to as point-of-interest (POI) recommendation. POI recommender systems have played an important role in the development of LBSNs because they are intended not only to satisfy users' personalised preferences as regards visiting new places, but also to help the LBSNs themselves to increase their revenues by providing users with intelligent location-based services such as location-aware advertisements.

Users of Location-Based Social Networks (LBSN) declare where they go (or check-in) and which places are of interest to them (by tagging or leaving tips). Both these spatial and semantic traces are equally useful in understanding people's relationships with place. Whereas, spatial tracks can be analysed to determine frequency of visits and favourite places, semantic interactions can give clues to the sort of activities they carry out in place and the experiences they share there. Combining both the explicit spatial association to place and the implicit semantics of interaction with place provides a unique opportunity for in depth understanding of both places and users.

The aim of the research presented in this thesis is to provide a user and place modelling framework from location-based social networks. The spatial (where), the semantic (what) dimension, and the temporal dimension (when) of user and place data are used to construct different views of a user profiles. Thus, the proposed approach provides users with the ability to project different views of their profiles, using their direct interactions with the social network or extended with a holistic view of other users' interaction with the network in different regions of geographic space. These profiles are then used as a basis for computing different methods of similarity between users. Studying user similarity from LBSN data is useful; as information available about users, their locations and activities are considered to be sparse. Similarity between users on LBSN is approached in a uniform manner within the proposed framework, thus providing means of computing spatial, semantic or a combined view of user similarity on these networks. Furthermore, user similarities are exploited to predict types of activities and places preferred by a user based on those users with similar preferences.

1.2 Thesis Hypothesis and Research Questions

The main hypothesis addressed in this thesis is:
The proliferation of GPS-enabled devices and their utilisation by users for geo-tagging personal resources, actions, and interactions on the Web is leading to the accumulation of a new type of information concerning individual users and user groups. The accumulation of spatiotemporal (ST) user footprints on the Social Web provides an opportunity for deriving profiles for both places, and users that closely reflect users' interests over space and time. Extracting and making sense of such profiles can enhance both place and content recommendation.

Providing evidence for this hypothesis in this thesis required:

- The design and development of methods for extracting users' spatiotemporal data from their Social Web "footprints".
- The development of methods for constructing a dynamic user profile from the user's spatiotemporal data and for simulating how it may change over time.
- The development of methods for learning place profiles and as result adapting individual users' profiles based on group interaction.
- The development of methods for understanding the effects of interactions in space and time on the constructed user profiles.
- The design and implementation of evaluation methods, using the collected data, to evaluate the quality of the extracted user and place profiles.

In order to verify this hypothesis, a number of important research questions were addressed:

1. Static and Dynamic Location-based User Modelling Framework:

Research Question 1: How can different views of user profiles be constructed from user footprints collected on LSBNs that emphasis the different facets of collected data? (Work of Chapters 4, 6)

2. Spatial and Semantic Enrichment for User Modelling:

Research Question 2: How does the enrichment process impact the quality of personal user profiles? (Work of Chapters 4,6)

3. Extracting Place Semantics from Geo-Folksonomies

Research Question 3: How can implicit semantics of place profiles be used to reflect users experience in geographic places through the activities they carry out in those places? (Work of Chapter 5)

4. Evaluation Methods

Research Question 4: How can we construct a new location recommendation method using different dimension of LBSNs and evaluate it existing methods? (work of Chapter 4, 5, 6)

Research Question 5: How can different user profiles be evaluated using user similarity measures to assess their quality? (Work of Chapter 7)

1.3 Contributions

The primary contribution of this research is the ability to use different views of user profiles to enhance place and content recommendation by combining the spatial, semantic, temporal and social dimensions of data available on the social web.

1. **Static and Dynamic Location-based User Modelling Framework:** We introduce a framework for modelling different levels of user profiles extracted from the heterogeneous user feedback in LBSNs. User-generated traces at venues in LBSNs include both spatial and implicit semantic content. The location traces are treated equally to the semantic traces inferred from their interaction with the place through tagging and tipping. Collective behaviour of users on the network are also used to understand the place characteristics and these in turn are further used in the modelling of user profiles.Both the spatial (where) and the semantic (what) dimensions of user and place data are used to construct different views of a user's profile. A place is considered to be associated with a set of tags or labels that describe its associated place types, as well as summarise the users' annotations in the place. A folksonomy data model and analysis methods are used to represent and manipulate the data to construct user profiles and place profiles.

   The user modelling framework is then adapted and extended to consider the temporal dimension of user interaction on LBSN. The temporal dimension of the data is used to derive dynamic user profiles that project users' activity and association with place over time. References

2. **Spatial and Semantic Enrichment of User Profiles:** The user's direct links to places extracted from the basic user modelling framework are extended to create enriched profiles describing richer views of the place data available on the social network. Two types of enrichment are undertaken: a) Semantic enrichment (based on tag similarity calculations); and b) Spatial enrichment (based on place similarity calculations). The similarity calculations are used to enrich the basic profiles and to build different views of these enriched user profiles. Hence, user profiles can be extended from a basic model that describes user's direct links with a place, to enriched profiles describing richer views of place data on the social network.

3. **A novel semantic place model** To study the place semantics in LBSNs, both

explicit place affordance; the sort of services offered in a place as denoted by its place type or place categories, and implicit place affordance; encapsulated in reference to activities in place annotations, are used in building semantic user profiles.

4. **A new Location Recommendation Method:** The evaluation of the proposed methods is undertaken within a location recommendation problem framework. A location recommendation method is proposed that considers similarities between places and similarities between user profiles (i.e., ratings from similar users on places). Both the temporal and semantic aspects of the data are taken into account.

5. **A novel User Similarity Evaluation Method:** Similarity between users on LBSN is approached in a uniform manner within the proposed framework, thus providing means of computing spatial, semantic or a combined view of user similarity on these networks. Using the profiles developed within the proposed framework, we are able to In particular, different similarity measures are presented based on: similarity of interests, similarity of co-location, similarity of place categories visited and similarity of activities undertaken in a place. "Short-term user similarity" and "Long term user similarity" are used to represent different methods of handling time in the folksonomy.

1.4 Thesis Structure

The rest of the thesis is structured as follows:

Chapter Two provides an overview of the literature related to the research discussed in the thesis. The chapter introduces the notions of user modelling and user similarity measures and reviews related work in both areas in the context of LBSN. Different approaches to location recommendation are also identified and critically evaluated.

Chapter Three A general overview of the approach adopted in this thesis is given here. In addition, a description of the data-set collected and the process of data cleaning, preparation and modelling is also given.

Chapter Four describes the proposed user modelling framework. It begins by introducing the geo-folksonomy model that is at the core of our work. Next, we explain our user modelling strategies by introducing the definitions of basic and enriched profiles. Finally, we describe an experiment used to compare and evaluate different profiles using different evaluation metrics.

Chapter Five introduces the various structures of geo-folksonomies that relate the concepts of place, place category and activity to users. The chapter considers how the folksonomy model is adapted to represent these concepts and how these can then be used to represent different views of user profiles.

Chapter Six describes the time modelling approaches proposed in this work for geo-folksonmies. Two different approaches that represent a snap-shot view of the dataset and a historical view of the dataset are used to develop temporal user and place profiles. Detailed evaluation experiments are carried out to test the quality of the proposed methods.

Chapter Seven explores different views of user similarity within the framework proposed. The similarity methods are systematically introduced and evaluated using realistic samples of datasets.

Chapter Eight highlights the key features of the research, evaluates its achievements in relation to its aims and concludes with an appraisal of the overall research experience and outcomes. Finally an outlook on further research issues and directions is given.

Chapter 2

Background and Related Work

2.1 Overview

This chapter gives an overview of related work in two main relevant research areas: user modelling, and location recommendation. In both areas, general methods are first described and then works that adapt these methods to the location domain are considered and evaluated. Figure 2.1 gives an overview of the topics covered in this chapter.

2.2 User Modelling

User modelling is the process of understanding, learning, and representing information about users [17]. The main goal of user modelling is to develop methods to address user needs, characteristics and preferences. This process can also help users find relevant information, provide feedback, support collaboration between users and predict a user's future behaviour [55].

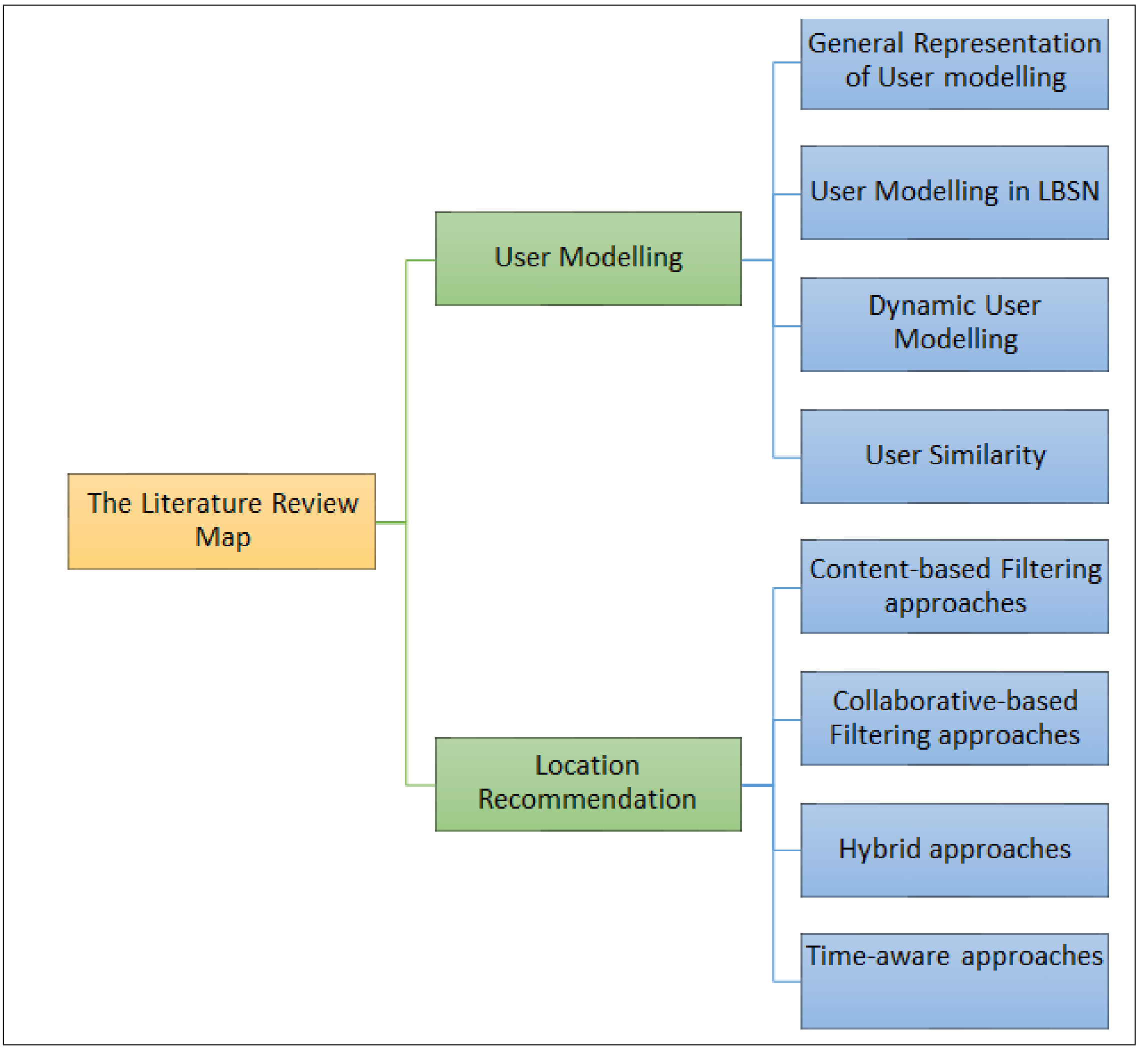


Figure 2.1: Road Map of Literature Review.

2.2.1 General Representation of User Model

User Profile Representation

User models are also known as *user profiles*. There are three fundamental methods that are widely used in user modelling. These are known as: *bag of words, keyword extraction, entity extraction and topic models*. Using the *'bag of words'* method, researchers simply extract words (terms) by removing stop words, punctuation and URLs. This method was widely used to analyse users' micro-blogs [24] [46] [83]. Alternatively, Liu et al. [80] used keyword extraction to mine the interests of users; they combined a translation-based method with a frequency-based method to extract appropriate interest keywords. Moreover, concepts can be represented as named entities that are extracted from textual content [38]. In all the above methods, a user profile is represented as a vector-based profile using a vector of terms and their associated weights. The weights are computed by a certain term weighting scheme such as TF and TF.IDF [99], or by a time-sensitive weighting scheme [29]. Representing the user profile as a vector of concepts has the advantage of growth and adaptive performance when new documents are added at a future date. In addition, it enables a continuous degree of similarity between queries and documents to be computed, allows documents to be ranked according to their possible relevance, and allows partial matching. However, this method may suffer from a noisy and a large number of unrelated words. A *topic model* is a type of statistical model for discovering the abstract "topics" that occur in a collection of documents [1]. Topic modelling is a text-mining tool frequently used for the discovery of hidden semantic structures in a text body [45]. To construct a user profile, Bennett et al. [15] represented users' interests as a set of topics that are extracted from large online individual's history of queries and clicked documents. , however this required expensive manual effort to determine the correct categories for each document. Harvey et al. [45] and Vu et al. [109] applied a latent topic model (i.e., LDA) to determine these topics, which means that the topic space is determined based purely on relevant documents extracted from query logs and hence human involvement is not required to define the topics. Jian el al. [106] used point-wise mutual information (PMI) [84] to measure the quality of the learned topics using LDA, and they argued that although topic modelling has recently been widely used in user modelling, poor performance in terms of of the LDA is expected when documents are too short, even if they are in a very large number.

In this work, the vector-based model was employed to represent the user and place profile because of the short text messages (tips) being used. Also, using a vector based model allows the computing a continuous degree of similarity between users and places especially when users are visiting new places.

User Profile Enrichment

Enrichment is the process of enriching user profiles in order to make them more meaningful. One of the problems of user modelling on the social web is the short length of the posts that are generated by users. To improve the understanding of these short messages, Abel et al. [5] created semantic user profiles by linking Twitter posts with related news articles from the web. Furthermore, to better understand the semantic meaning of the posts, [5, 59] enriched the posts with embedded links to enrich their contents. Also, in [6] Abel el al. proposed an extension to user profiles using DBpedia through adding semantics to tweets by extracting entities and enriching them with external resources in order to create facets (e.g. persons, locations, organisations etc.), but their work was only limited to Twitter and they did not consider enriching profiles using internal network structure. Orlani et al. [87] proposed a category-based user profile based on the category information of entities from DBpedia. The results based on a user study demonstrate that the category-based user profiles performed similarly to entity-based ones, but the authors did not evaluate those user modelling strategies in the context of recommendations. Furthermore, the users' interests can be modelled as a network structure of terms and related terms [40]. These profiles are called *semantic network based profiles* where weights are assigned to the terms and their related terms as well as the links between them. Existing work has used WordNet [58] or external knowledge sources such as DBpedia [10]] to ascertain related terms. In InfoWeb [39], user interests are modelled using a personalised information filtering system for online digital documents to create semantic network-based user profiles. Initially, each user profile is made up of a set of concepts that represent a user's interests. As the user continuously interacts with the system, his or her user profile is updated by adding more concepts to the semantic network as well as links between the concepts.

Although extending user profiles from external sources can be useful, other approaches have been proposed to extend the process of building user profiles that use tags that are not directly used by the user using the tag relatedness within the social network [85], [9], [49]. Also, in [31], the place semantics were used to enrich user profiles from the concepts that are semantically related to the tags directly used by each user within a folksonomy dataset, **but the authors did not consider enriching user profile with related spatial preferences.**

2.2.2 User Modelling in LBSNs

In LBSNs, one can differentiate between different types of user profiles depending on their source of information.

1. **User Demographic profiling**: In LBSNs, a user profile may contain demographic data or interests such as description and semantic text and tags. In [88], a user's profile is constructed using information on age, gender, cuisine preferences (from category information), and income data. In [96], Ramaswamy et al. inferred user profiles using low-end devices capable only of handling voice and short text messages (SMS); their user profile approach focused on utilising a user's address book to determine social relations (social affinity). Other research, e.g. [119, 121], focused on improving the accuracy of the location tags and categories by extracting user activity patterns for each location.

2. **User History Profile**: User profiles can be inferred from users' online histories via three main sources: i) user ratings, ii) user interaction patterns, and iii) user search histories. User ratings can be found in many online social networks such as Yelp [1]and YellowPages [2], which allow users to leave explicit ratings for locations to express their opinions about locations. [27, 50, 122] Interaction patterns in LBSNs include user tags or comments made by a user regarding a loc-

[1]https://www.yelp.co.uk/
[2]http://www.yellowpages.com/

ation [41, 115]. User search histories include map browsing histories and spatial searching logs [11, 108, 111].

3. **User Location History Profile** The user location history tracks the places that the user has visited previously in the form of a 'check-in'. We consider the user history location profile to be the most realistic profile to use as it records the actual places they have already visited. Using this profile, one can understand a user' s behaviour and preferences and it can be used for friend recommendation by calculating the similarity between user profiles. For example, when two users co-locate in the same place and at similar visit times, a relationship between them can be assumed as they share the same preferences and interests.

Works on modelling user data in LBSN mainly consider two problems: a) place (or point of interest) recommendation, and b) user similarity calculation. Different types of data are used by different approaches, namely geographic content, social content and textual annotations made by users. Different methods are also used in analysing the data, for example distance estimations for geographic data modelling and topic modelling for annotation data analysis.

In the area of POI recommendation, studies range from generic approaches that use the popularity of places [21] to recommendation methods that are based on a user' s individual preferences [120]. A useful survey on these approaches can be found in [13].

Based on check-in data gathered through Foursquare[3], Noulas et al. [86] exploit factors such as transition between types of places, mobility between venues and spatiotemporal characteristics of user check-in patterns to build a supervised model for predicting a user's next check-in. Ye et al. [120] investigated the geographical influence with a power-law distribution. The hypothesis is that users tend to visit places within short distances of one another. Other works considered other distance distribution models [127]. Gao et al. [37] considered a joint model of geo-social correlations for personalised POI recommendation, where the probability of a user checking in to a new

[3]https://foursquare.com/

POI is described as a function of correlations between user's friends and non-friends close to, and distant from a region of interest. Liu et al. [78] approached the problem of POI recommendations by proposing a geographical probabilistic factor model that combines the modelling of geographical preference user mobility. Geographical influence is captured through the identification of latent regions of activity for all users of the LBSN reflecting activity areas for the entire population and mapping the individual user mobility over those regions. Their model is enhanced by assuming a Poisson distribution for the check-in count which better represents the skewed data (users visiting some places once, while other places hundreds of times). Whilst providing some useful insights into modelling the spatial dimension of the data, the above works do not consider the semantic dimension of the data.

Correlations between geographical distance and social connections were noted in [26, 36]. Techniques of personalised POI recommendation with geographical influence and social connections mainly study these two elements separately, and then combine their output together with a fused model. Social influence is usually modelled through friend-based collaborative filtering [110, 120, 133] with the assumption that a user tend to be friends with other users who are geographically close, or would want to visit similar places to those visited by their friends. Ying et al. [125] proposed to combine the social factor with individual preferences and location popularity within a regression-tree model to recommend POIs. The social factor corresponds to similar users; users with common check-ins to the user in question. In this work, we also use this factor when extending user profiles to represent places of interest within the region of user activity.

More recently, the importance of content information for POI recommendation was recognised. Two types of content can be considered, attributes of places and user-contributed annotations. Place categories are normally used as an indication of user activity, thus a user visiting a French restaurant would be considered as interested in French food, etc. User annotations in the form of tips and comments are analysed collectively to extract general topics in order to characterise places or to extract collective sentiment indications about the place. Examples of works that consider place categories are [53, 77, 82]. In [77, 82], the Latent Dirichlet Allocation (LDA) model was used to represent places as probability distribution over topics collected from tags and categories or comments made in a place and similarly aggregate all tips from places a user has visited to model a user's interest. Aggregation was necessary as terms associated with a single POI are usually short, incomplete and ambiguous. [53], on the other hand, modelled topics from tweets and reviews from Twitter and Yelp, and assumed that the relations between user interests and location is derived from the topic distributions for both users and locations. This work also models users' association to place through the place's relation to tags, but also adds the influence of other users' relationships to the place to the equation. Aiming at improving the effectiveness of location recommendation, Yang et al. [116] proposed a hybrid user POI preference model by combining the preference extracted from check-ins and text-based tips which were processed using sentiment analysis techniques. Sentiment analysis is an interesting type of semantics which is not considered this thesis, but can be incorporated in future work.

2.2.3 Dynamic User Modelling

Based on the hypothesis that the interests of users change over time, several works began to pay attention to dynamic user interests in online social networks [4, 4, 18, 87]. In [4], short-term and long-term user profiles are proposed in the context of news recommendations on Twitter. Short-term user profiles refer to user interests within a short-term period (e.g., the last two weeks), while long-term user profiles consider user interests using all historical user generated content. In contrast, user profiles that describe short-term user interests are usually updated frequently [39, 75]. Gentile et al. proposed an approach to dynamically model user expertise based on information communication exchange such as emails [38]. Alternatively, Ding and Li [29] presented a time weight item-based user profile that calculates the exponential time decay function to compute time weights for different items according to each user and each cluster of items. Xiong et al. [91] modelled time stamped user-item ratings as a 3-D tensor by assuming that each time feature vector depends only on its immediate predecessor; they further argued in [114] that user preferences often exhibit long-term and short-term factors, before proposing a session-based temporal graph model to capture the dynamic preferences of users over time. Koenigstein et al. [61] presented a matrix factorisation model that combines temporal analysis of user ratings and item popularity trends to make music recommendations, while Zhang et al. [129] designed an evolutionary topic pattern mining approach to discover changes in topic structures on a community question answering platform. This approach first extracted question topics via LDA in each time window before determining topic transitions based on cosine similarity; finally, the life cycles of the extracted topics were analysed. Moreover, Rafeh and Bahrehmand [94] proposed an adaptive collaborative filtering algorithm which takes time into account to reflect fluctuations in users' behaviour over time. Liu et al. [79] developed a social temporal collaborative ranking model to recommend movies.

2.2.4 User Similarity

Studying user similarity from LBSN data is useful as information available about users, their locations and activities is considered to be sparse. User similarities can be exploited to predict types of activities and places preferred by a user based on those of users with similar preferences. So far, most works on user similarity have mainly focused on structured data, such as geographic coordinates, or semi-structured data, e.g. tags and place categories. Recently, Lee and Chung [71] presented a method for determining user similarity based on LBSN data. While the authors made use of check-in information, they concentrated on the hierarchy of location categories supplied by Foursquare in conjunction with the frequency of check-ins to determine a measure of similarity. Mckenzie et al. [82] suggested exploring unstructured user-contributed data, namely tips provided by users. A topic modelling approach is used to represent users' interests in places. Venues (places in Foursquare) are described as a mixture of a given number of topics and topic signatures are computed as distributions across venues. User similarity can then be measured by calculating the dissimilarity metric between the topic distributions of users. This method of modelling venues is interesting, but it limits the representation of user profiles as they are based on generated topics derived from collective user place annotations. Thus, individualised association of users with place is somewhat ignored. In contrast with the above approach, the model used in this work does not assume constraints on the number of topics represented by the tags, but instead combines an individual' s association with both tags and places in the creation of user profiles. Social links between users have also been widely utilised to improve the quality of location-based recommender systems since social friends are more likely to share common interests in relation to POIs than strangers. Most current works have derived the similarities between users from social links and put them into traditional memory-based or model-based collaborative filtering techniques. For example, some investigations [32,110,123,124,127] seamlessly integrated the similarities of users into the user-based collaborative filtering techniques, while others [25,116,131] employed the user similarities as the regularisation terms or weights of latent factor models.

Thus, in this work we propose a new method that exploits the social affinity between users by aggregating the check-in frequency and tagging frequency from the historical check-in data of all users. The social affinity in this research is defined as the similarity between users and not by the social relation of users. We compare different types of user similarity based on the temporal dynamics and semantics of user profiles. Similarity between users is computed using different views of user profiles; using their direct interactions with the social network or extended with a holistic view of other users' interaction with the network in different regions of geographic space.

2.3 Location Recommendation Approaches

In general, recommendation systems attempt to predict the user rating or preference on a specific item, which is considered an explicit rating. In recent years, recommendation systems have become widely utilised in a variety of areas: some popular applications including movies, music, news, books, research articles, search queries, social tags, locations and products in general. In location recommendation, the explicit rating of locations is not available. However, the check-ins or tagging history of the users is utilised to reflect users' preferences in locations implicitly. Thus, with the availability of location-based social network data, existing recommendation methods can be applied to location recommendation by treating locations as items. The main approaches of the recommendation systems can be classified into the following three groups:

1. ***Content-based filtering approaches***: Recommendation is based on items similar to those that the user preferred in the past;
2. ***Collaborative-based filtering approaches***: Recommendation is based on users with similar taste and preferences in the past;
3. ***Mixed approaches***: These methods combine collaborative and content-based methods.

In the following subsection we will explain these three approaches and how they can be applied to the location recommendation domain.

2.3.1 Content-based Filtering Approaches

In content-based filtering, items are recommended based on a comparison between the content of the items and past user profile ratings of items. The recommendation in these systems is based on the item itself rather than the preferences of other users. To select such items, content-based filtering measures item-to-item similarity by analysing the content of textual information of the items. This textual information includes the keywords representing the user's characteristics (age, gender, location, etc.) and item's characteristics (product, price, appearance, etc.). For example, to recommend movie m to user u, the content-based recommendation system will obtain data on the previously rated movies by user u and then the movies with the highest similarity to the user preferences are recommended. Different techniques are used to measure the user's preferences and the candidate items' characteristics, including a cosine similarity measure, Bayesian classifier, and decision trees [7], and are often used to recommend items containing textual information, such as books, websites, and news.

Content-based filtering was successfully used for location recommendation in [88] and [96]. The recommendation was performed by matching user preferences that are implicitly inferred from users' profiles, with context features extracted from locations trajectories, such as tags and categories to make recommendations. Moreover, the authors utilised the spatiotemporal contexts to provide a richer user profile and thus make high-quality recommendations.

The advantages of using these methods for location recommendation are: a) if a user profile is constructed accurately to reflect the user preferences, the new user and location problems (cold start) can be avoided, and b) content-based filtering is a user independent method, it doesn't rely on other users preferences to carry out recommendations.

The drawbacks of using content-based filtering in LBSNs is the limited degree of novelty in recommending new locations as the recommendation is dependent on the locations in the user profile [81]. In addition, it is difficult to extract and analyse the features of multimedia content such as audio, image, and video in order to measure the user's and item's contents. Another problem with content analysis is that, if two different items are represented by the same set of features, they are not distinguishable. Since each item is represented by the most important keyword, content-based systems cannot distinguish between a good item and a bad one, if they use the same terms [102].

2.3.2 Collaborative-based Filtering Approaches

Collaborative filtering makes recommendations by measuring similarities of users' preferences [101]. As it analyses patterns of favourable items without analysing any of the content properties of items, it has been possible to discover these items without examining their content properties. Different collaborative filtering systems have been developed in the literature. The methods introduced in [102] were considered the first system to use collaborative filtering algorithms to automate prediction and recommendation. It is clear that collaborative systems do not have some of the limitations of content-based systems, especially when content information is unavailable. However, contrary to content-based approaches, collaborative filtering methods rely mainly on rating information, a user typically needs to provide a sufficient number of ratings before the system can return accurate recommendations. So, when a user is new, the recommender system faces the "*cold start problem*" [74]. Also, collaborative systems depend mainly on user's preferences to make recommendations. Therefore, for a new item that has not been rated, the recommendation system would not be able to process it, which is known as the "*new item*" problem. Moreover, sparsity is another major issue with collaborative filtering due to a lack of information about users and items. However, collaborative filtering problems can be overcome, if both content-based filtering and collaborative filtering are combined [89]. Also, adding context to a user profile can overcome this issue, that is two users can be considered similar not only if they rated the same item similarly, but also if they belong to the same geographic area.

Approaches for collaborative recommendations can be classified into two categories [7, 117]: *Memory (Heuristics)-based approaches*, and *Model-based approaches*

Memory-based Approaches

Memory-based algorithms essentially measure the similarity in previous ratings for the same item by different users. The aggregated ratings from those similar users can then be used as a prediction for the target user's rating. The recommended ratings are calculated from all known ratings by means of a particular mathematical expression. For instance, in the user-based CF approach [47] the collection of ratings is utilized to determine similarity between the users' preferences, and then the ratings of the most similar users (the neighbourhood) to the target user are used to estimate unknown ratings of the target user.

In LBSNs, collaborative filtering can be applied for location recommendation, using the fact that similar users tend to visit similar locations, that is a user is more likely to visit a location that is preferred by similar users. This can be achieved by inferring similarity between users or locations from, 1) implicit ratings resulting from their activity in the place (check-ins or tipping), or 2) explicit online ratings history of locations (such as restaurants and hotels) directly provided by users. There are two traditional models for CF approaches 1) user-based collaborative filtering, where recommendations are based on user similarity [47] [132], and 2) item-based collaborative filtering that use similarity between locations (items) to make recommendations [60].

Many online services, e.g. Yelp and Yellowpage, allow users to explicitly express their preferences for locations using ratings. Based on those ratings, several location recommendation systems are constructed to provide personalised location recommendation based on direct user ratings of locations [27,28,50,122]. Also, other systems infer user similarity from implicit ratings by using the frequency of visits (check-ins or tips) at locations [103,105].

Model-based approaches

Model-based approaches learn a predictive model through the collection of previous ratings, using approximation to predict the target user's ratings. This requires a prior learning process where the model is built, but thereafter the model directly generates rating predictions, which leads to a faster response at recommendation time. Matrix factorisation is a widely used example of model-based collaborative filtering [63]. Matrix factorisation models user-item interactions in a latent factor space, where latent factors are used to efficiently predict unknown ratings by taking into account the user and item biases that are likely to be caused by rating deviations. For example, some users may consistently give higher ratings than other users, and some items may receive higher ratings than other items.

In location recommendation, the user ratings are represented by a matrix (user-location) which reflects the user's visit to the corresponding location. The user and location latent factors can be computed using a user-location matrix by applying the matrix factorisation techniques [62, 64]. Cheng el al. [25] used a multi centre Gaussian model, together with matrix factorisation to investigate the geographical and social influence for location recommendation. In [54], matrix factorisation was used as a technique to associate each category of place with a latent vector and extrapolated the relevance score of a user to a POI based on the latent vectors of the categories of the POI. Gao el al. [35] studied the content information on LBSNs for POI recommendation by investigating different types of content information on LBSNs regarding sentiment indications, user interests, and POI properties with a low-rank matrix factorisation method for POI recommendation.

2.3.3 Hybrid Approaches

Content-based filtering and collaborative filtering can be combined to overcome the drawback of each method and to provide better recommendations. In general recommendation systems, hybrid approaches were introduced in [89]. However, upto our knowledge, hybrid approaches were not used In location recommendation problems before. Collaborative filtering recommender systems suffer from data sparsity and cold start problems. Content-based recommender systems try to avoid these problems, but only provide generic recommendations that ignore user's personal preferences. All the above issues are prominent in location recommendation problems and hence it is worth investigating the effectiveness of a hybrid approach in this domain. Hybrid recommender systems have been classified under seven main categories [19]: (i) weighted: each item gets a number of partial scores from different simple recommendation technique. The score reflects the value of this item with respect to each recommendation technique. The total item score results from the combination of the partial scores. (ii) switching: the system selects from among simple recommender techniques based on the evaluation of the recommendation situation. [43]; (iii) mixed: the output of two or more recommendation techniques is given to the user and the best items among the lists can then be selected by the user; (iv) feature combination: features of one source are fed into an algorithm that is designed to carry out data processing of a various sources; (v) feature augmentation: a recommendation technique is applied to extract a number of features, which are then used as input to another recommendation technique; (vi) cascade: a recommendation technique is applied to enhance the predictions made by another recommendation technique; (vii) meta-level: the recommendation technique model is used as input to another. In this thesis we adopt a hybrid (cascade) approach in the location recommendation component of our framework.

2.3.4 Time-Aware Location Recommendation

User behaviour in LBSNs is often influenced by time. For example, a user is more likely to check-in to a restaurant than going to the library during lunch time. This kind of temporal pattern can facilitate the identification of user habits and interests. Taking this context into account can potentially improve the recommendation results, as well as allow us to predict recommendation at different time points.

Different ways of representing time are as follows.

- **Continuous variable**: where the exact time stamp, e.g '2012-10-09 03:02:30' is used.
- **Categorical values(Discrete Time slots)**: where time periods or seasons, e.g mornings, evenings, weekends, weekdays, Christmas, Easter, are used.
- **Time units**: where common time units of Day, week, or a year. e.g Monday, Tuesday, March, April, 2012, 2013..etc.
- **Time Windows (consecutive hours)**: where time units are clustered into windows relevant to the application in question, e.g 3 hour window.

Time-aware recommendation approaches have been classified as follows [20]:

1. **Continuous time-aware models** The time information is represented as a continuous variable with time-stamps attached to the user-item ratings. The recommendations can then be made for a time different from the input time (e.g where can I go tomorrow?). A common approach in this method is to use different weights to ratings according to their 'age' with respect to the target time, so that recent ratings will have more influence than old data [29, 48, 63].
2. **Categorical time-aware models** The time information is represented as categorical values (e.g weekends, week days) so that the recommendation can be made according to the time context by selecting relevant data, (e.g. where can I go on the weekends) [8, 12, 70, 97].
3. **Time adaptive models**. In these models, data is dynamically adjusted relative to the time dimension. These models are different from categorical and continuous time-aware models because particular time context is not targeted in the recommendation [57, 72, 73].

As discussed above, the time dimension has been widely used for traditional recommendation systems. In LBSNs, the check-in behaviours mostly follow regular patterns. For example, users visit a bar at night rather than in the morning, and tend to go to university in the morning rather than at night. The regular periodic patterns are transformed into discrete time slots to provide location recommendations for users. Some works [34, 53, 126, 131] approached the problem by converting continuous time into discrete time slots and managed the temporal effect separately for each time slot using collaborative filtering techniques. Yuan et al. [126] proposed a time aware POI recommendation algorithm that extends user-based collaborative filtering by including the time factor when calculating the similarity between two users. They only considered check-ins at a certain time slot rather than all time slots to make their recommendations. A model-based POI recommendation algorithm was proposed earlier in [34] based on matrix factorisation with temporal influence. They investigated the temporal cyclic patterns of check-ins in based on two temporal criteria: non-uniformeness and consecutiveness. In the case of non-uniformness criteria, user check-in activities would vary through the day. While in the case of the consecutiveness criteria, a user tends to have more similar check-in preferences in consecutive hours. They introduced a temporal state $T \in [1, T]$ to represent the hour of the day, where T=24 is the total number of temporal states. They also defined other temporal categories like weekends and weekdays by changing the temporal state into T=7 and thus aggregating over weekly patterns. Ye et al. [119] proposed a method to extract location feature based on temporal distribution of user' check-ins. They used explicit patterns such as the total number of check-ins, the total number of visitors and the distributions of check-in times over a week and over 24 hour interval. Moreover, they used implicit relatedness that captures correlations between locations from check-in behaviour based on a moving window and computed a similarity measure based on a probabilistic total variation distance to compare individual feature types via their temporal bands.

Location recommender systems can also recommend locations based on the current time using temporal characteristics of user check-in behaviour (as in the case of continuous and categorical time-aware traditional recommender systems). In [26], Chao et al. developed a model of human mobility that predicts future individual movements based on the fact that human tend to travel at regular times of the day. They proposed a coherent model that combines three properties: periodic temporal change, geographic influence and social network structure. Rahimi and Wang [95] proposed a novel recommendation algorithm, the Probabilistic Category Recommender, which uses the temporal probability distribution to recommend the category of a place location that would be interesting for the user based on their historical behaviour. They first find pairs of check-ins to the same category of location from the same user, and then plot the frequency of check-in pairs based on the time interval (one-hour) of those check-ins to further predict the probability of future check-ins in an hourly manner. A recent study in [128] develops a continuous temporal model based on the kernel density estimation method to build a continuous time probability density of a user visiting a new location.

Most techniques used in location recommendation suffer from time information loss and may not correlate temporal influences at different time slots because of time discretization. In this thesis, a continuous time-aware location recommendation method is developed in which the temporal aspect is used as a factor that decays the weight of ratings over time. This method can also consider the various types of the temporal pattern (weekly, monthly, yearly) which provides a potential to predict locations in the future.

Also, not much work have considered the rich place semantics implicit in data on LBSN. In this work, we consider other properties, beyond absolute place location in the problem of place recommendation. In particular, we explore place types and activities associated with places.

2.4 Literature Gap Analysis

So far, previous works have studied data produced from LBSN from the point of view of enhancing the services provided by these networks, namely, for point of interest (POI) recommendations. There, the question of concern is to find places of interest to a user based on their history of visits to other places and their general interaction with the social network. Most works relied mainly on the spatial dimension of user data [76], with some works more recently exploring the relevance of the social and content data dimensions on these networks [36]. However, data dimensions are normally treated separately, or their outputs are combined in fused models. Previous works attempting a similar approach used matrix factorisation techniques to handle the multiple data dimensions, but did not consider the use of the range of content data as used in this thesis.

As mentioned earlier , several attempts have been made to enrich user profiles from either external sources [5,5,59], or internal network [85], [9], [49], but their work deals with semantic enrichment only (i.e enriching the user profile with concepts), so there is a gap in enriching user profiles from other views (like spatial, activity, or place type enrichment), or from combined views to integrate multiple dimensions. Thus, there is a need to model the influence of other users relations in the place with the user's association to place. In addition, most of the above works considered only a static view of user profile enrichment, so there is a need to explore an adaptive approach represent the temporal dynamics of user profiles on these networks.

Furthermore, most works on user similarity mainly focused on structured, e.g., geographic coordinates, or semi-structured, e.g., tags and place categories, data [71] [71]. For example, Mckenzie et al. [82] suggested exploring unstructured user-contributed data, namely tips provided by users using a topic modelling approach. However, this method of modelling venues is interesting, but it limits the representation of user profiles as they are based on generated topics derived from collective user place annotations.

Several attempts have been made to enhance the POI recommendation in LSBNs using either content-based collaborative filtering approaches [88] [96], or Collaborative-based filtering approaches [32, 110, 123, 124, 127] . But each approach has its drawbacks [13]. For example, collaborative filtering suffer from data sparseness and cold start problem, but content based filtering overcome these problems. So, to improve the effectiveness of location recommendation systems, estimations of user preferences and user similarity has to be accurate. One solution to achieve this is to integrate and hybridize different types of recommendation methodologies to overcome limitations.

Finally, one of the the main drawbacks of the majority of dynamic user models related to LBSNs is that they rely on a static snapshot of attributes which do not reflect the change in users interests and behaviour over time. So, there is a need to include the temporal dimension more homogeneously in the presentation of user profiles and compare this against the snapshot treatment of time in the literature.

2.5 Discussion

The work presented in this thesis targets dynamic user and place modelling on the geo-social web and its evaluation using recommendation methods. From the literature, there are some gaps that needs to addressed:

The need for Dynamic Spatio-semantic User Profiles:
There is a need to consider, homogeneously, the spatial, semantic and temporal dimensions when considering user profiles on LBSN. Each dimension provides distinct opportunities to explore user interactions with place.

The need for Dynamic Spatio-semantic Place Profiles:
In a similar way to user modelling, LBSN provide a rich opportunity to draw up profiles of geographic places based on actual user interactions and experiences with

places. In comparison to static place models represented in gazetteers and map databases, these place profiles have the potential to provide a dynamic view of how places are used people visiting and interacting with them on LBSN.

The need of deriving a joint method for location recommendation This is a need to design a joint POI recommendation method that can overcome the problems or recommendation systems. This can be done by incorporating factors (e.g., semantic, temporal) into traditional collaborative filtering model which can give better decisions.

The need for Dynamic Spatio-semantic User Similarity Methods:
The proposed approach and framework will allow the exploration of different method of checking similarity between users on LBSN. Studying user similarity from LBSN data is useful, as the information available about users, their locations and activities is generally sparse. User similarities can be exploited to predict types of activities and places preferred by a user based on those of users with similar preferences.

Chapter 3

General Framework

3.1 Introduction

This chapter provides an overview of the proposed framework for the user and place modelling research design. The framework is based on a geo-folksonomy model that links users, places and tags together. A temporal dimension is also included in order to construct, in effect, a temporal geo-folksonomy model. In addition, some semantics related to place categories and activities are captured that provide more understanding of users and places. The Research methodology is presented in 3.2. The proposed framework is presented in Section 3.3. The data preparation process is explained in Section 3.4, the geo-folksonomy model is presented in Section 3.5 and the database design is described in Section 3.6. Finally, a summary of the chapter is provided in Section 3.7.

3.2 Research Methodology

In this research, to verify the hypothesis, we applied the Data Science Research Methodology (DSRM) introduced by Peffers el al. [90] as depicted in figure 3.1. Each step is described and related to this PhD thesis chapters as follows.

1. Problem Identification and Motivation: This phase involves a critical and deep

learning of the user and place modelling strategies and its research related areas. The first step of this phase involves the identification of gaps in related literature as presented in chapter two. In the second step, the research hypothesis statement and the research questions is identified as presented in Chapter one. The third step requires the choice of the data source and the development tools that will be used to test the hypotheses. The last step involves doing a time plan for the research by dividing the main problem into tasks and identifying the required milestones.

2. Objectives of the solution: In this step, the problem definition in the previous step is used in order to propose the objectives of the solution. In this research, different levels of user profiles are extracted from the heterogeneous user feedback in LBSNs. User-generated traces at venues in LBSNs include spatial, temporal and implicit semantic content. Collective behaviour of users on the network are also used to understand the place characteristics and these in turn are further used in the modelling of user profiles. The qualitative spatio-temporal model framework has been proposed in an attempt to investigate the effect of using different dimensions in user and place modelling on place recommendation. Chapter one presented a qualitative objective of the solution.

3. Design and development: This step aims to design a solution of the problem and develop it. This step has been explained in chapter three, four, five and Six. The entire design of the general framework and the data models are explained in chapter three. The design of static and dynamic user modelling has been introduced in chapter four and six. Finally, the design of semantic place model is explained in chapter four.

4. Demonstration: This step involves using the developed framework in a suitable context. In this thesis, different experiments are been carried out in chapters four, five, six using samples of realistic data sets for a representative number of users with different levels of usage of the LBSN to demonstrate the effectiveness of

the proposed framework. Users' interaction on LBSNs can be regarded as user feedback on geographic places they visited and interacted with. Thus, User's visits to places are recorded into a suitable database along with their comments and tags.

5. Evaluation: This step involves assessing the effectiveness of the method proposed compared to other methods. In this research, the evaluation experiments aim to measure the impact of using the full range of content captured on LBSN when building user profiles in comparison to using only partial views based on the check-in information. Two evaluation methods are used, the top-N recommendation used in chapters four, five and six, and user similarity evaluation proposed in chapter 7. The effectiveness of user profiles are measured using recall, precision and F1. We compare the results of the top-N recommendation using basic recommendation methods: Item-based Collaborative Filtering and User-based collaborative Filtering approaches and show the effectiveness improvement.

6. Communication: In this final step, researchers publish their contributions to the audience to get their feedback and stamp the importance of the problem and its novelty. This thesis resulted in four publications, three conference papers and two journal papers. The publications are listed in the list of publications section.

3.3 General Framework

This thesis introduces a new framework for user and place modelling in LSBNs. The proposed framework is able to provide different user and place modelling strategies including interest extraction, content enrichment and the identification of the temporal dynamics of user interests. Different user modelling strategies are then evaluated in the context of recommender systems using standard evaluation metrics such as 'precision'

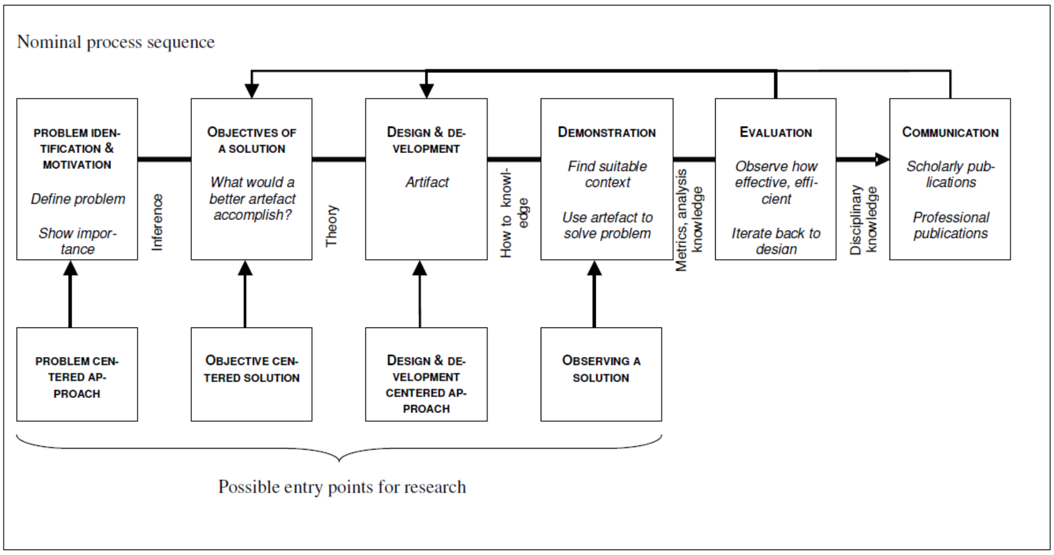


Figure 3.1: Design Science Research Methadology (source: Peffers el al. [90])

and ‘recall’. An outline of the framework is shown in Figure 3.2 and the different stages are described in more detail in Chapters 4, 7, 5, and 6. Two modelling strategies are proposed, static and dynamic modelling. The static model doesn’t take the dynamics of the users interests into account. Thus, the static model provides an overall picture of user behaviour in places and the shift in human preferences can not be learned. Dynamic modelling, on the contrary, combines the spatial, semantic, and temporal aspects to study the change of user behaviour over time. It provides a more up to date picture of user behaviour. Changes in interests influences the learning process and it takes into account current user interests and preferences. The first stage is the static user modelling from geo-folksonomies which involves six stages: a) data collection, b) geo- folksonomy building, c) profile creation, d) similarity measurement, e) profile enrichment, and f) evaluation using location and tag recommendation. The second stage is the user similarity calculation and its evaluation. This stage involves calculating different similarity measures from the various profiles we have and then evaluating the result with a suitable information retrieval evaluation measure, as discussed in Chapter 7. The third stage involves extracting semantics from the place, mainly place type and place activities, by changing the structure of the geo-folksonomy to relate to place

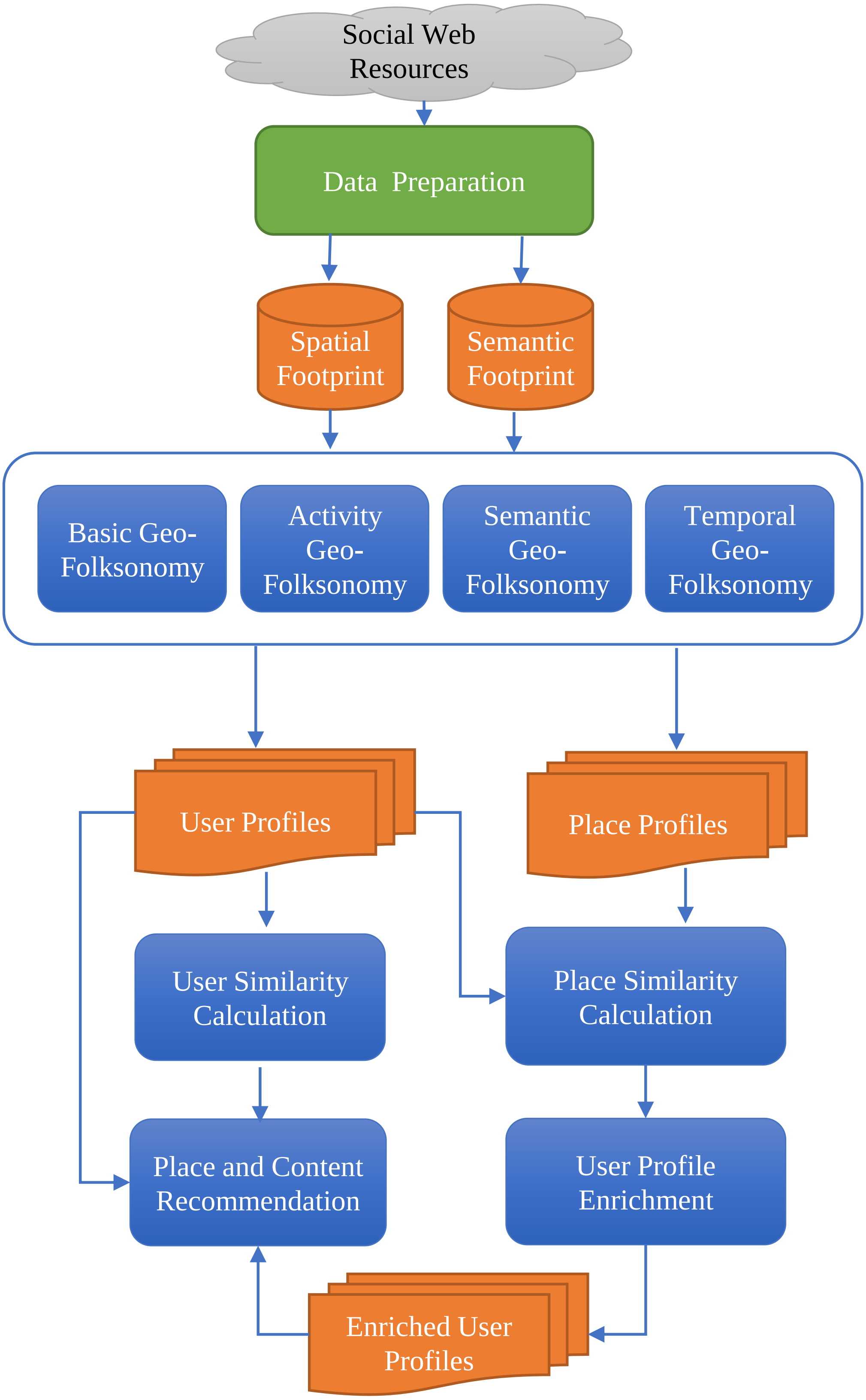


Figure 3.2: The General Framework

activity, place types, users and tags alternatively. The last stage is the temporal user modelling that is discussed in detail in Chapter 6. The design of the temporal user modelling follows the same design as the static user modelling, but with added methods for including temporal dynamics.

3.4 Data Preparation

To build the temporal geo folksonomy model, a step-by step data preparation is required in order to extract the relevant, necessary entities. Figure 3.3 shows the data preparation process; this consists of: a) data collection, b) data pre-processing, and c) time interval partition. These steps will be discussed in the following subsections.

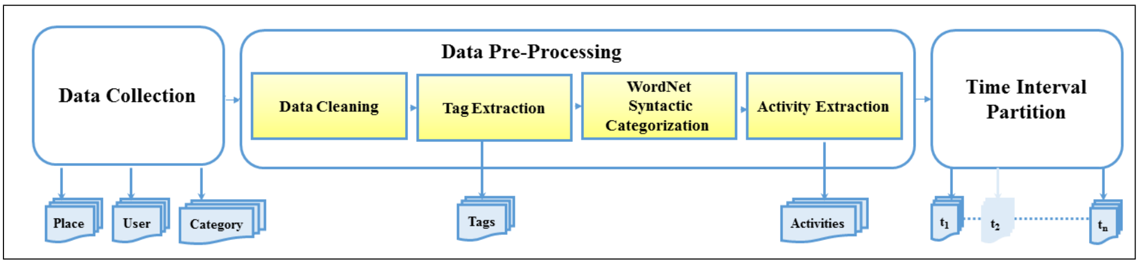


Figure 3.3: Data Preparation Process

3.4.1 Data Collection

The data collection source used for the experiments is Foursquare. Users of Foursquare check-in at different venues, share their check-ins and/or add tips to venues that contain suggestions for things to do, see, or eat at the location. In Foursquare, the location (venue) is the main element that connects users and user-generated content such as tips and tags - which are features of the location. This work is concerned about two types of activities: check-ins and tipping, as shown in Figure 3.4. Data about venues, tips and tipping users can be streamed publicly via Foursquare, but due to privacy issues, user check-in streams can only be accessed with the permission of the users. Therefore, it is difficult to collect user check-ins from Foursquare because it is hard to obtain permissions from a large number of users. Fortunately, users from Foursquare tend to push their check-in activity to Twitter as a Tweet. Tweets are public and can easily be collected. Check-ins, on Twitter, are identified by a shortened URL linked in the tweet. By resolving such a URL, we can obtain the full check-in activity in Foursquare. Figure 3.5 shows a shortened check-in URL and its resolved URL on the Foursquare web application.

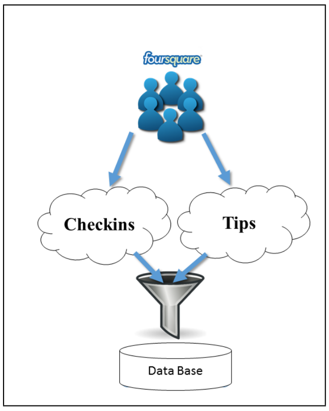


Figure 3.4: Data Collection

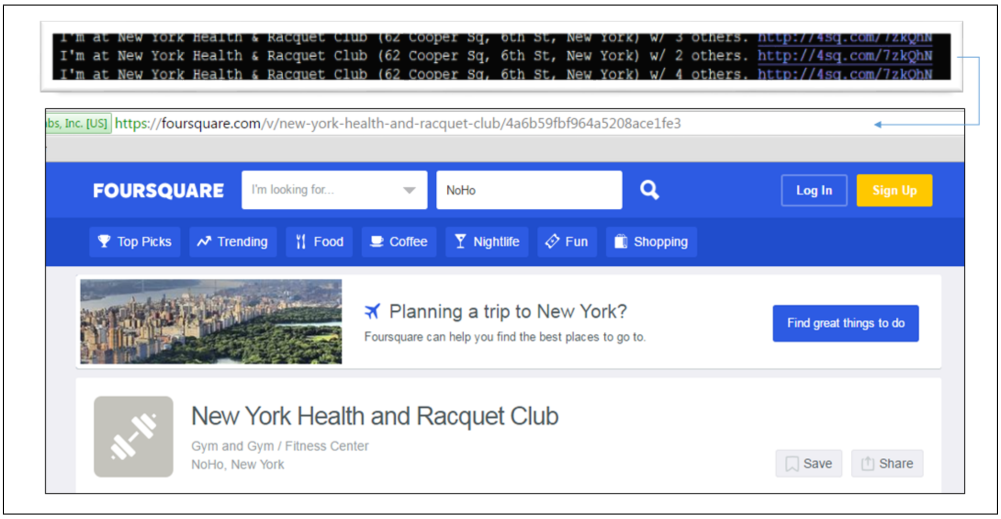


Figure 3.5: Check-in shortened URL and Actual URL

In this work, the following data-sets were collected:

-A long-term (about 10 months) check-in data set relating to New York City was collected from Foursquare, from 12 April 2012 to 16 February 2013 by [1]. The typical format of a check-in activity data item is as follows:

- A unique user ID identified by Foursquare that represents a user account.
- A unique venue ID identified by Foursquare that represents a location point.
- A UTC timestamp which represents the actual time of the check-in (e.g., Tue Apr 03 18:11:04 +0000 2012). We convert the UTC time to the java time format (e.g., 2012-04-3 18:11:04)
- A unique Category ID identified by Foursquare that represents the category of each place.

Each venue ID in the above data-set is then used to collect the following data, using the Foursquare API [1].

- Tip data: The tip- activity is collected from each venue. A typical tip activity format is as follows:

- A unique user ID identified by Foursquare that represents a user account
- A unique venue ID identified by Foursquare that represents a location point.
- A unique tip ID identified by Foursquare that represents a tip.
- This is a short-text that represents the opinion of a user about a place
  (e.g. Their Mac and Cheese is the best I've ever had! YUM!!).

- Venue Details: details about each venue are collected. The details are as follows:

- A unique venue ID identified by Foursquare that represents a location point.

[1] https://developer.Foursquare.com/

- The venue name - identified by Foursquare as referring to a venue ID. (e.g., Island Salad)
- The longitude and latitude of the venue.
- The tags attached to the venue.
- The user count that represents the number of users who have visited the place.
- The check-in count that represents the number of check-ins associated with the place.
- The venue categories for each place.

- User Data: Data about the users who have left tips in each venue as recorded in the check-in data-set. The format of the user data-set is as follows:

- A unique user ID that is identified by Foursquare that represents a user account
- The gender of the user.

- Category data: data about each place category. In Foursquare, venues are organised into a three-level hierarchical category classification. It The format of the category data-set is as follows:

- A unique category ID identified by Foursquare that represents the category of each place.
- A category name identified by Foursquare.(one of 437 categories)
- A parent category name that represents one of the 9 root categories (i.e., Arts & Entertainment, College & University, Food, Great Outdoors, Nightlife Spot, Professional & Other Places, Residence, Shop & Service, Travel & Transport).

Tables 3.1 and 3.2 show the statistics of the collected data. The data is geographically related to New York City, in the United States. The reason for collecting the data for this city in particular was the availability of the check-in data-set.

Table 3.1: Check-in Data-set

Total number of checkins	227,428
Number of places	38,333
Number of users	1,083
Average number of checkin/user	210
Average number of places/user	84

Table 3.2: Tip Data-set

Total number of users	167,786
Number of places	28,878
Number of tips	604,924
Average number of tips/user	3.6
Average number of places/user	3.3

3.4.2 Data Pre-Processing

Data Cleaning

This pre-processing involves the following four steps:

1. Removal of all numeric characters.
2. Removal of all non-numeric and special characters (ex. %, * ...etc.).
3. Stop word removal.
4. Filtering of any tag less than 3 characters in length.

Tag Extraction

After pre-processing, tips are tokenized into words (tags) which are then stored in the database. Duplicate tags are removed to avoid redundancy and any misunderstanding of the relationships between places and users which might otherwise ensue.

Word-Net Syntactic Categorisation

In this step, the Word-Net lexicographer is used to categorise tags. The Word-Net lexicographer is a lexical database for the English language. There are 44 lexicographer files that can be used to classify a word into a suitable category. Table 3.3 shows the different word-net lexicographer files [2]. As the figure shows, each noun, verb, and adjective may be classified into multiple categories. For example, a verb may be an emotion, a body, a creation etc., and a noun can be an action, a state, or a time, etc.

Activity Extraction

Noun.act class and the verb.competition class and the *.ing verbs were used to identify activity words. The noun.act class identifies the nouns that denote acts and actions; the verb.competition class contains verbs for fighting, athletic activities and so on; and the *.ing verbs, nouns or adjectives identify an activity carried out.

3.4.3 Time Interval Partition

Users' interests change as time goes by, which reveals that users may be interested in different places at different periods of time. Therefore, users' dynamically changing interests can be expressed at different time intervals. As explained in the data collection section, time is stored as a time-stamp attached to each user activity (check-in or tip). So, time-stamps needs to be partitioned to the time intervals which are termed in this thesis 'time-slices' (for example $t_1, t_2, ...t_n$). Each time-slice denotes a temporal user-tag, user-place or place-tag at a t_i time interval. Figure 3.6 shows an algorithm for calculating the number of time slices between two input dates. As the algorithm shows, the time-slice can be identified as an hour, day, week, month or year.

[2]https://wordnet.princeton.edu/man/lexnames.5WN.html

Table 3.3: A Subset of WordNet Lexicographer Files

Name	*Contents*
noun.act	nouns denoting acts or actions
noun.animal	nouns denoting animals
noun.artifact	nouns denoting man-made objects
noun.attribute	nouns denoting attributes of people and objects
noun.body	nouns denoting body parts
noun.cognition	nouns denoting cognitive processes and contents
noun.communication	nouns denoting communicative processes and contents
noun.event	nouns denoting natural events
noun.feeling	nouns denoting feelings and emotions
verb.competition	verbs of fighting, athletic activities
verb.consumption	verbs of eating and drinking
verb.contact	verbs of touching, hitting, tying, digging
verb.creation	verbs of sewing, baking, painting, performing
verb.emotion	verbs of feeling
verb.motion	verbs of walking, flying, swimming
verb.social	verbs of political and social activities and events
verb.weather	verbs of raining, snowing, thawing, thundering

3.5 Geo-Folksonomy Model

Foursquare holds a large number of crowd-sourced venues (> 65 million places) from a user population estimated recently to around 55 million users. As the application

```
public int timeintervals(String timestamps, String dateto, String timeinterval)
    DateTime start = new DateTime(timestamps.getTime());
    DateTime end= new DateTime(dateto);
    switch(timeinterval) {
        case "weeks":  timeslice = Weeks.weeksBetween(start, end).getWeeks();
        case "months": timeslice=Months.monthsBetween(start, end).getMonths();
        case "days":   timeslice=Days.daysBetween(start, end).getDays();
        case "years":  timeslice=Years.yearsBetween(start, end).getYears();
        case "hours":  timeslice=Hours.hoursBetween(start, end).getHours();
        }
    return(timeslice);
}
```

Figure 3.6: Time Interval Partitioning

defines it, a venue is a user-contributed "physical location, such as a place of business or personal residence." Foursquare allows users to check in to a specific venue, sharing their location with friends, as well as other online social networks such as Facebook or Twitter. Built with a gami fication strategy, users are rewarded for checking in to locations with badges, in-game points, and discounts from advertisers. This game-play encourages users to revisit the application, compete against their friends and contribute check-ins, photos and tips. Tips consist of user input on a specific venue, normally describing a recommendation, experience or activity performed in the place.

In this work, we use a folksonomy data model to represent user-place relationships and derive tag assignments from users' actions of check-ins and annotation of venues. In particular tags are assigned to venues in our data model in two scenarios as follows.

1. A user's check-in results in the assignment of place categories associated with the place as tags annotated by this user. Thus, a check-in by user u in place r with the categories (represented as keywords) x, y and z, will be considered as an assertion of the form $(u, r, (x, y, z))$. This in turn will be transformed to a set of triples $\{(u, r, x), (u, r, y), (u, r, z)\}$ in the folksonomy.

2. A user's tip in the place also results in the assignment of place categories as tags, in addition to the set of keywords extracted from the tip. Thus, in the above example, a tip by u in r with the keywords $(t_1, \cdots, t_n)$, will be considered as an assertion of the form $(u, r, (x, y, z, t_1, \cdots, t_n))$, and is in turn transformed to individual triples between the user, place and tags in the folksonomy.

The process of extracting keywords from tips is done by tokenizing the tip into a set of words (terms) on white space and punctuation. Then we remove all words with non-latin characters and stop words. The output is a set of single words (term vector). Furthermore, we use WordNet syntactic category and logical groupings for classifying the extracted terms [3]. For example, WordNet 'noun.act' category is used to filter action verbs and nouns to describe a user- or place- associated activity (ex. swimming, buying or eating).

The data capturing process results in the creation of a *geo-folksonomy*, which can be defined as a quadruple $\mathbb{F} := (U, T, R, Y)$, where U, T, R are finite sets of instances of users, tags and places respectively, and Y defines a relation, the tag assignment, between these sets, that is, $Y \subseteq U \times T \times R$, [2, 51].

A geo-folksonomy can be transformed into a tripartite undirected graph, which is denoted as folksonomy graph $\mathbb{G}_{\mathbb{F}}$. A geo-Folksonomy Graph $\mathbb{G}_{\mathbb{F}} = (V_{\mathbb{F}}, E_{\mathbb{F}})$ is an undirected weighted tripartite graph that models a given folksonomy $\mathbb{F}$, where: $V_{\mathbb{F}} = U \cup T \cup R$ is the set of nodes, $E_{\mathbb{F}} = \{\{u, t\}, \{t, r\}, \{u, r\} | (u, t, r) \in Y\}\}$ is the set of edges, and a weight w is associated with each edge $e \in E_{\mathbb{F}}$.

The weight associated with an edge $\{u, t\}, \{t, r\}$ and $\{u, r\}$ corresponds to the co-occurrence frequency of the corresponding nodes within the set of tag assignments Y. For example, $w(t, r) = |\{u \in U : (u, t, r) \in Y\}|$ corresponds to the number of users that assigned tag t to place r.

In this research we use a database model to represent the tripartite graph. The nodes are represented as entities, and each table relates the entities to represent the interactions between the entities.

[3] https://wordnet.princeton.edu/man/lexnames.5WN.html

3.6 Database Design

The database engine used in this research is SQ-Lite. This was selected because it supports the date and time functions that are used to extract temporal dynamics of the model. It is also server-less, so the database reads and writes are accessed directly from the database files on disk without an intermediary server process. In addition, this database engine is compatible with Raven [4], a university supercomputer service that was used to run our project on.

The database instance created in this research was designed to support the storing and searching of the collected folksonomy data-set as well as the output of the folksonomy co-occurrence analysis methods implemented. The data model of the database is shown in Figure 3.7. The three distinct components of the geo-folksonomy are modelled, with

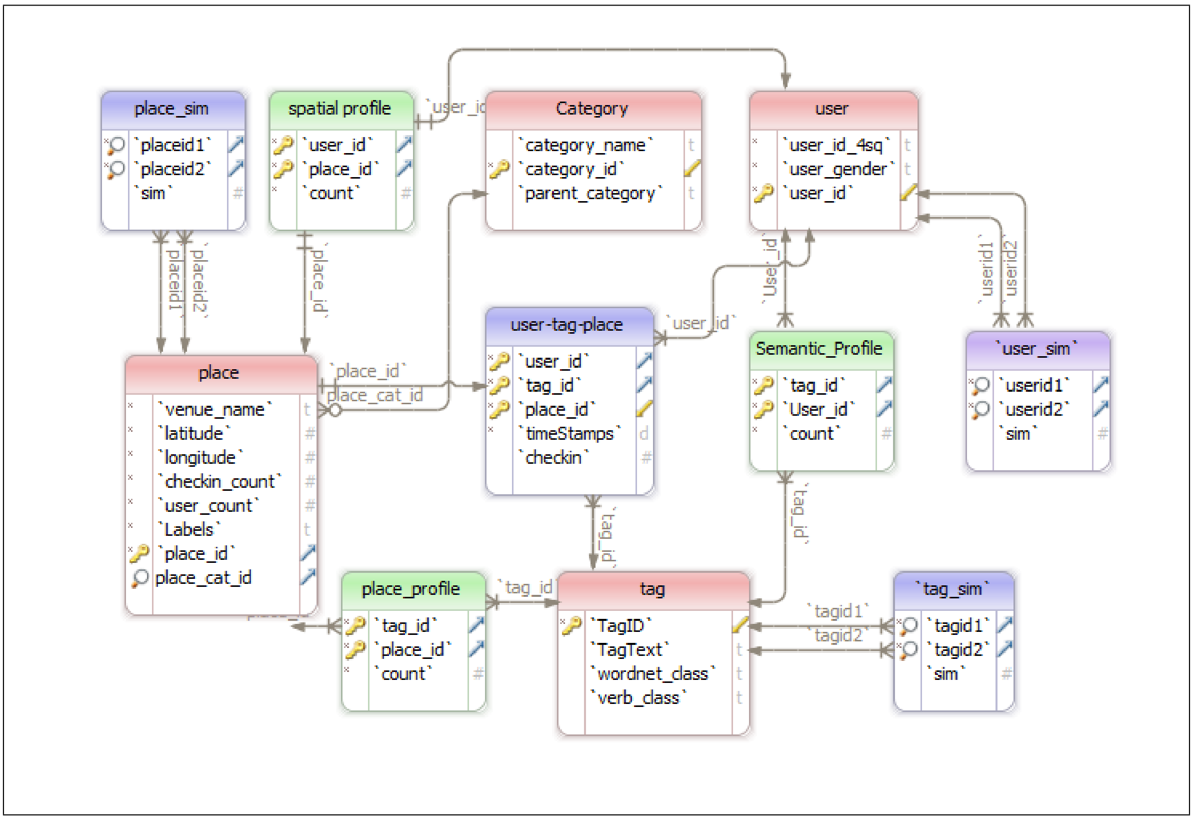


Figure 3.7: The Geo-Folksnomy Database Represenation

[4] http://www.cardiff.ac.uk/arcca/services/equipment/User-Guide/user-guide.html

the place table representing folksonomy tags extracted from tips, and the user table - representing folksonomy users. The user place tag relates the three tables - user, place and tag - together. There is a many to many relationships between users, places, and tags which is then broken to extract the spatial information that represents the relationship between user and place; the semantic information which represents the relationship between user and tag; and the place profile which represents the relationship between place and tag. The database also contains several tables for storing the output of the folksonomy analysis such as tags similarity, place similarity and user similarity.

This database is instantiated multiple times in the database system - for different time models, as discussed in Chapter 6. Also, it is instantiated to represent the places semantics which are represented by the relationship between place category, place activity and, the user, as discussed in Chapter 7. The database schema of place semantics is shown in Figure 3.10. As the figure shows, a ternary relationship is created between user, place-category, and tag which is a many to many relationship. Relevant profiles, such as user-based category profiles and tag-based category profiles are extracted. The similarities between categories, tags and users are then re-calculated using the new structure. Activities are also instantiated from a tag data-set which represents a subset of tags.

Figures 3.8, 3.9 are examples of the queries that can be applied to the database:

```
SELECT User_Profile.user_id, User_Profile.tag_id, dataset_Tag.TagText, count
FROM User_Profile, dataset_Tag
WHERE User_Profile.tag_id=dataset_Tag.TagID
ORDER BY count desc
```

Figure 3.8: A Query for Displaying all Tags Sorted by Tag Frequency, from High to Low.

```
SELECT AVG(places_count) FROM
(SELECT user_id, COUNT (DISTINCT lace_id) as places_count
FROM dataset_UTPT
GROUP BY user_id)
```

Figure 3.9: A Query for Calculating the Average Number of Places for all Users in the Data-set.

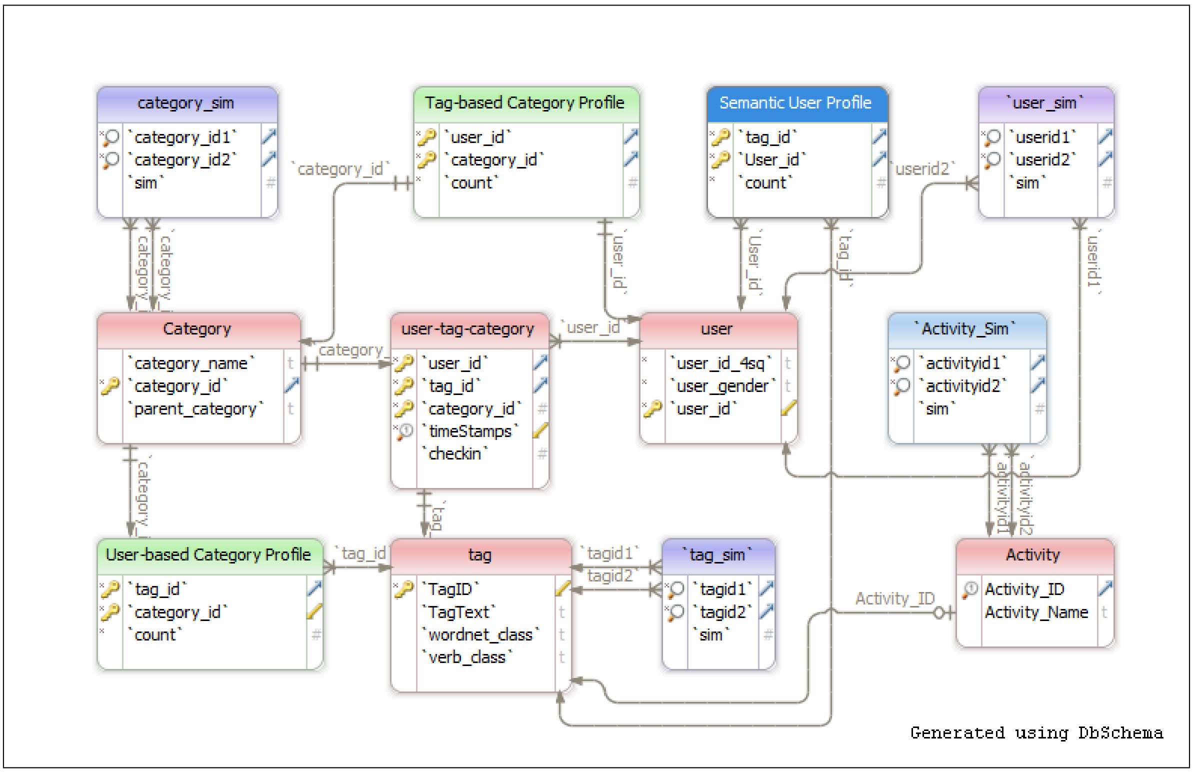


Figure 3.10: The Place Semantics Database Schema

3.7 Summary

A user and place modelling framework is proposed in this chapter. The framework is a pipeline that shows a map of how methods are explained in each chapter. The framework is divided in four sub-frameworks: static user modelling, user similarity calculation and evaluation, semantics extraction, and dynamic user modelling. The common processes involved in the four frameworks are then explained. The first process is the data preparation. This involves: a) data collection, b) data pre-processing, and c) time interval partition. Then the relevant entities are stored in a database which is described in the database design section. The proposed framework will be discussed in the next four chapters.

Chapter 4

User Modelling in Geo-Folksonomies

4.1 Introduction

Our work in this thesis focuses on Location-Based Social Networks (LBSN) that collect information on users' interests in physical places in the real world. By "switching on" location on devices, we are giving away information on our whereabouts, our daily routines, activities, experiences and interests. Thus, in comparison to other personal information, location data are possibly the most crucial type of data of relevance to privacy, as it pulls together our virtual and physical existences and thus raises critical questions about privacy in both worlds. This chapter introduces methods for constructing user profiles that considers the different dimensions of the data captured from users on LBSN. These profiles, when made transparent to users of the network, should empower their sense of awareness and control of their data as discussed in [42].

In this work, both semantic and spatial interactions of users are used to project distinct and complementary views of personalised user profiles. Thus, user's annotations on places they visit are compiled in semantic profiles, while collective user annotations on places are used to create specific profiles for places that encapsulate user's experiences in the place. Place profiles, in turn, are used to construct personalised user profiles. In comparison to previous works in the area of recommendations, LBSN data are treated as folksonomies of users, places and tags. User annotations in the form of tips, their interaction with places, in the form of check-ins, as well as general place properties, namely, place categories and tags, are analysed concurrently to extract relations between the three elements of the folksonomy.

Simple co-occurrence methods and similarity measures are used to compute direct and enriched user profiles.

Thus the proposed approach provides users with the ability to project different views of their profiles, using their direct interactions with the social network or extended with a holistic view of other users' interaction with the network in different regions of geographic space. Previous works attempting a similar approach used matrix factorisation techniques to handle the multiple data dimensions, but did not consider the use of the range of content data as used in this work. Sample realistic data from Foursquare are used to demonstrate the approach and evaluation results show its potential value. In particular it is shown that enriched user profiles offer potentially more accurate views, than direct profiles, of user's spatial as well as semantic preferences. Hence, these should be considered when designing tools for enabling user awareness on these networks. This chapter will address the following research questions:

- How can different views of user profiles be constructed from user footprints collected on LSBNs that emphasis the different facets of collected data?
- How does the enrichment process impact the quality of personal user profiles?
- How can we construct a new location recommendation method using different dimension of LBSNs and evaluate it existing methods?

The rest of the chapter is organised as follows. A folksonomy background literature is presented in Section 4.2. Different types of user profiles are defined in Section 4.3. In Section 4.4, the experiment used to evaluate the approach is described and its results presented and discussed. The chapter concludes in Section 4.5.

4.2 Background

4.2.1 Folksonomy definition

There has been a huge amount of research progress regarding folksonomy analysis and mining. The word “Folksonomy” is a concatenation of two words “folks” and “taxonomy” [107]. “A folksonomy is the result of personal free tagging of information and objects for one’s own retrieval. The tagging is carried out in a social environment (shared and open to others). The act of tagging is performed by the person consuming the information” [107]. Folksonomy is also known as collaborative tagging, social classification, social indexing, and social tagging. Folksonomies allow their users to manage bookmarks online and to annotate them with keywords known as tags. [112]

4.2.2 Folksonomy Structure

Formally, a folksonomy is a tuple F=$< U, T, R, A >$where U, T, R represent users, tags and resources respectively. The relationship ‘A’ relates U, T and R. Consequently, the folksonomy can be represented as a tripartite graph [44]. The vertices of the graph are the users, tags and resources. The vertices of the graph are the users, tags, and resources. Alternatively, the folksonomy graph can be represented as a three-dimensional adjacency matrix. However, to simplify the manipulation, the tripartite graph can be decomposed to three bipartite graphs: tag-user, tag-resource, and user-resource. [14].

4.2.3 Folksonomy Analysis Methods

Folksonomy analysis methods look at the relationships between users, resources and tags in the social tagging systems in order to generate recommendations. These methods can be classified into two main methods: reduction methods, and non-reduction methods. Reduction methods include co-occurrence [68], and collaborative filtering approaches [56], [100]. The reduction analysis methods reduce the three-dimensional folksonomy data into three two-dimensional projections. Non-reduction folksonomy methods include hypergraphs [130], [134], tensors [104], [92]. These methods use the three dimensions to analyse the data.

The main disadvantage of using non-reduction methods is their sparseness, which leads to difficulty processing data with normal machine memory sizes. The act of reducing the data into two dimensions produces denser data-sets. Alternatively, the main disadvantage of reduction methods is that the hidden relationships between the dimensions will be ignored, and thus these methods can miss some important information. In the proposed method, the co-occurrence approach is employed to analyse the three-dimensional data. Although is used a reduction approach, an attempt was made to capture any relation between the three dimensions using combined similarity measures.

A lot of work was carried out in addressing the two folksonomy analysis methods. The work presented in [67] proposed a novel method for including content data into the widely recognised FolkRank tag recommendation algorithm, enabling it to recommend tags for new untagged documents based on their textual content. The results showed that including content information (words and their frequency rather than documents) in the recommendation process gives a significant improvement over content-unaware recommendation in full tagging datasets.

Furthermore, contextual user modelling in folksonomy was addressed by [3]. In this work, semantically meaningful contextual information was deduced from tagging systems, and a ranking algorithm that exploits this contextual information was designed. The idea is to not only use tags alone, but to also use context like the spatial information, categories of the tags, and urls to facilitate further understanding of user behaviour.

There are three measures of tag relatedness as stated in [23]: Co-occurrence, Cosine Similarity and Folk Rank.

In the co-occurrence measure, the tag-tag co-occurrence graph is defined as a weighted undirected graph whose set of vertices is the set T of tags, and two tags, t_1 and t_2, are connected by an edge if there is a least one post. The weight of the edge is given by the number of posts that contain both t_1 and $t1$ [22].

In cosine similarity, the measure of tag relatedness is computed by using the cosine similarity of tag-tag co-occurrence distributions. Two tags are considered related when they occur in a similar context, and not when they occur together. [93]

The FolkRank method is derived from the PageRank algorithm, which reflects the idea that a web page is important if there are many pages linking to it, more so if those pages are important themselves. [16] The same principle is employed for FolkRank, a resource which is tagged with important tags by important users becomes important itself . The same holds for tags and users. [52]

For the purpose of this work, the cosine similarity measure is chosen to measure the tag-tag, place-place, and user-user similarity.

4.3 User Modeling Strategies

We propose an approach to modelling users in LBSN that represents a user's spatial, semantic and combined spatio-semantic association with place. A spatial user profile represents the user's interest in places, while a tag-based profile describes his association with concepts associated with places in the folksonomy model. A spatio-semantic profile describes the user specific interest in certain concepts associated with places in his profile. A user profile is built in stages. Starting with a basic profile that utilise direct check-in and annotation histories, a user profile is then extended by computing the relationship between places and concepts derived from collective behaviour of other users in the dataset. A basic profile represents actual interactions with places, while the extended profile describe "recommended" associations given overall interactions between users, places and concepts in the dataset.

Figure 4.1 depicts the overall process of user profile creation. The process starts with data collection of check-ins and tip data from Foursquare, that are then processed to extract users, places and tags and their associated properties. The modelling stage includes the definition of relationships between the three entities and the application of folksonomy co-occurrence methods to extract the different types of profiles. Place and tag similarity calculations are used to further extend the basic profiles to build different views of enriched user profiles.

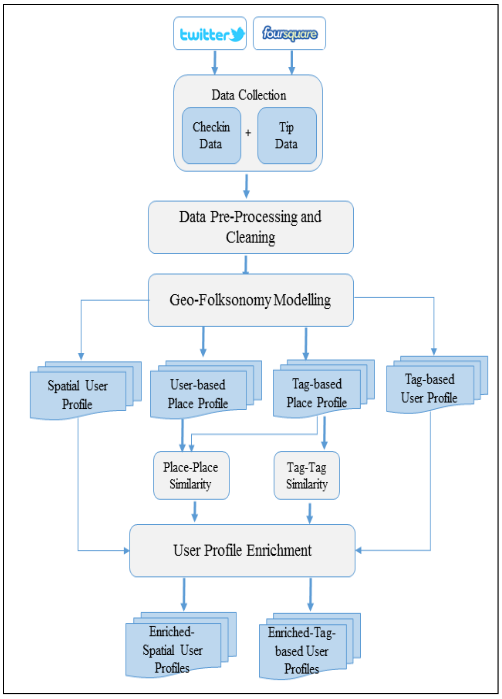


Figure 4.1: The Framework of the User Modlling Framework

We are able to model such interactions separately in the extended profile by controlling the similarity function used to create the profile. For example we can focus on modelling the types of places visited by the user or take into account visit behaviour of other users whose profiles overlap with the user, as discussed below.

4.3.1 Basic User Profiles

Definition 1. ***Spatial User Profile*** *A spatial user profile $P_R(u)$ of a user u is deduced from the set of places that u visited or annotated directly.*

$$\begin{aligned} P_R(u) =& \{(r, w(u,r)) | (u,t,r) \in Y, \\ & w(u,r) = |\{t \in T : (u,t,r) \in Y\}|\} \end{aligned}$$

$w(u,r)$ is the number of tag assignments, where user u assigned some tag t to place r through the action of checking-in or annotation. Hence, the weight assigned to a place simply corresponds to the frequency of the user reference to the place either by checking in or by leaving a tip.

We further normalise the weights so that the sum of the weights assigned to the places in the spatial profile is equal to 1. We use $\overline{P_R}$ to explicitly refer to the spatial profile where the sum of all weights is equal to 1, with
$\overline{w}(u,r) = \frac{|\{t \in T : (u,t,r) \in Y\}|}{\sum_{i=1}^{n} \sum_{j=1}^{m} |\{t_i \in T : (u,t_i,r_j) \in Y\}|}$, where n and m are the total number of tags and resources, respectively. More simply, $\overline{w}(u,r) = \frac{N(u,r)}{N_T(u)}$, where $N(u,r)$ is the number of tags used by u for resource r, while $N_T(u)$ is the total number of tags used by u for all places.

Correspondingly, we define the tag-based profile of a user; $P_T(u)$ as follows.

Definition 2. ***Semantic User Profile*** *A semantic user profile $P_T(u)$ of a user u is deduced from the set of tag assignments linked with u.*

$$\begin{aligned} P_T(u) =& \{(t, w(u,t)) | (u,t,r) \in Y, \\ & w(u,t) = |\{r \in R : (u,t,r) \in Y\}|\} \end{aligned}$$

$w(u,t)$ is the number of tag assignments where user u assigned tag t to some place through the action of checking-in or annotation.

$\overline{P_T}$ refers to the semantic profile where the sum of all weights is equal to 1, with $\overline{w}(u,t) = \frac{N(u,t)}{N_R(u)}$, where $N(u,t)$ is the number of resources annotated by u with t and $N_R(u)$ is the total number of resources annotated by u.

Furthermore, we define a spatio-semantic profile of a user $P_{RT}(u)$, that is a personalised association between user, place and tag.

Definition 3. ***Spatio-Semantic User Profile*** *Let $\mathbb{F}_u = (T_u, R_u, I_u)$ of a given user $u \in U$ be the restriction of $\mathbb{F}$ to u, such that, T_u and R_u are finite sets of tags and places respectively, that are referenced from tag assignments performed by u, and I_u defines a relation between these sets: $I_u := \{(t,r) \in T_u \times R_u | (u,t,r) \in Y\}$.*

A spatio-semantic user profile $P_{RT}(u)$ of a user u is deduced from the set of tag assignments made for place r by u.

$$P_{RT}(u) = \{([r,t], w_u([r,t])) | (t,r) \in I_u,$$
$$w_u([r,t]) = |\{t \in T_u : (t,r) \in I_u\}|\}$$

where $w([r,t])$ is how often user u assigned tag t to place r.

$\overline{P_{RT}}$ is the spatio-semantic profile where the sum of all weights is equal to 1, with $w_u([r,t]) = \frac{N(u,[r,t])}{N_{RT}(u)}$, where $N(u,[r,t])$ is the number of times u annotate r with t, and $N_{RT}(u)$ is the total number of tags assigned by u for r. (Note that tag assignment by users for a place comes from both the explicit action of annotation as well as implicit action of checking-in as represented in the geo-folksonomy model).

4.3.2 Basic Place and Tag Profiles

Let $P_T(r)$ and $P_U(r)$ be the tag-based place profile and user-based place profile for place r (defined in a similar manner to user profiles above). Conceptually, a tag-based place profile is a description of the place by the tags assigned to it and a user-based place profile is an account of users' visits to the place and they are defined as follows:

Definition 4. ***Tag-based Place Profile*** *A tag-based place profile $P_T(r)$ of a place r is deduced from the set of tag assignments linked with r.*

$$P_T(r) = \{(t, w(r,t)) | (u,t,r) \in Y,$$
$$w(r,t) = |\{r \in R : (u,t,r) \in Y\}|\}$$

$w(r,t)$ is the number of tag assignments where user u assigned tag t to some place through the action of checking-in or annotation.

$\overline{P_T}$ refers to the semantic profile where the sum of all weights is equal to 1, with $\overline{w}(r,t) = \frac{N(r,t)}{N_R(r)}$, where $N(r,t)$ is the number of users that annotated place r with t and $N_R(r)$ is the total number of users who use tags to annotate place p.

Definition 5. ***User-based Place Profile*** *A user-based place profile $P_U(r)$ of a place r is deduced from the set of users that u visited or annotated the place directly. It is the same as the spatial user-profile but with opposite assignments (Place-User instead of User-Place)*

$$P_U(r) = \{(u, w(r,u)) | (u,t,r) \in Y,$$
$$w(r,u) = |\{t \in T : (u,t,r) \in Y\}|\}$$

$w(r,u)$ is the number of tag assignments, where u assigned some tag t to place r through the action of checking-in or annotation. Hence, the weight assigned to a place simply corresponds to the frequency of the user reference to the place either by checking in or by leaving a tip.

So far, the basic user profile provides only a limited view of the user association with places and concepts derived directly from captured data. Basic profiles reduce the dimensionality of the folksonomy space by considering only 2 dimensions at a time; user-place and user-tag, leading to a loss of correlation information between all three elements.

Users profiles can be extended to represent possible latent relationships in the data. Thus a user profile can be used to present places (respectively tags) similar to those in the basic profile, where similarity between places (respectively tags) is measured through the collective actions of other users of check-ins and annotations.

To compute tag-tag similarity, profiles for tag are first defined through the places they are used to annotate. Thus, a *place-based tag profile* ($P_R(t)$) of a tag t is a weighted list of places r that are annotated by t. That is, $w(r, t)$ is determined by the number of users' check-ins and tips that resulted in assigning t to r in the geo-folksonomy. Traditional method for calculating similarity are Cosine Similarity [7], Jaccard measure [65]and Pearson correlation [98]. Cosine similarity measures the angle between two vectors and is useful for calculating similarity between two text documents. The Pearson Coefficient is a more complex approach for calculating the similarity because it generates a "best fit" line between attributes in two vectors. Jaccard measure calculates the similarity by eliminating the zero matching attributes.In this thesis, the similarity between tags is defined as the cosine similarity between their place-based tag profiles as follows.

$$CosSim(t_1, t_2) = \frac{|P_R(t_1) \cap P_R(t_2)|}{\sqrt{|P_R(t_1)|.|P_R(t_2)|}} \tag{4.1}$$

On the other hand, similarity between places is defined by measuring the similarity of their tag-based and user-based profiles. Let $P_T(r)$ and $P_U(r)$ be the tag-based place profile and user-based place profile for place r (defined in a similar manner to user profiles above). Conceptually, a tag-based place profile is a description of the place by the tags assigned to it and a user-based place profile is an account of users' visits to the place.

Cosine similarity between tag-based place profiles ($CosSim_{tag}(r_1, r_2)$) and between user-based place profiles ($CosSim_{user}(r_1, r_2)$) construct a tag-oriented ranking and user-oriented ranking, respectively. These similarity rankings can be aggregated using the so-called Borda method [30] to compute a generalised similarity score between two places.

$$Sim(r_1, r_2) = \gamma * CosSim_{tag}(r_1, r_2) + (1 - \gamma) * CosSim_{user}(r_1, r_2) \tag{4.2}$$

where $0 \leq \gamma \leq 1$ is a parameter that determines the balance of importance given to similarity scores from $P_T(r)$ and $P_U(r)$. Conceptually, similarity between two places is a function of the overlap between their tag assignments only (for $\gamma = 0$), a measure of their common visitors only (for $\gamma = 1$), or both (for γ between 0 and 1).

4.3.3 Enriched User Profiles

We extend the basic user profiles by the information extracted from the computation of tag and place similarity above. The enriched user profiles will therefore present a modified view of how users are associated with places that reflect collective user behaviour on the LBSN.

Definition 6. ***Enriched Spatial User Profile*** *An enriched spatial user profile $\acute{P}_R(u)$ of a user u is an extension of the basic profile by places with the highest degree of similarity to places in $\overline{P_R(u)}$. Let R_u be the set of all places in $\overline{P_R(u)}$ and w_i is the weight associated with place i in the profile.*

$$\acute{P}_R(u) = \{< r_i, w_i > |$$

$$w_i = \left\{ \begin{array}{ll} w_i & , if\, r_i \in R_u \\ w_i * Max(Sim(r_i, r_j)) & , \forall (r_i \in \{R - R_u\} \land r_j \in R_u) \end{array} \right\}$$

We compute the maximum similarity of the K most similar places in the dataset for every place in the basic user profile, and use the highest similarity score as the weight for the new place in the enriched user profile. The process of building the enriched spatial profile is shown in the following algorithm

The algorithm has three inputs: the spatial user profile $P_R(u)$, the number of places to be enriched (K), and the γ value that controls the enrichment combination. The algorithm starts with finding all places in the spatial user profile and then compute the similarity between places using the CosSim function. Then, the top K similar places

Algorithm 4.1: SpatialEnrichment ($P_R(u)$,K,γ)

Input: A spatial Profile $P_R(u)$, K, γ

Output: Enriched Profile $\acute{P}_R(u)$

```
for all places r_i in Spatial-Profile P_R(u) do
    if γ = 1 then
        Compute CosSim_tag(r_1, r_2)
    else
        if γ = 0 then
            Compute CoSim_User(r_1, r_2)
        else
            Compute Sim(r_1, r_2, γ)
        end if
    end if
    Find top K similar places r_j to each r_i in P_R(u)
    for each < r_j, sim > in top similar places do
        w_j = w_i * sim
        add < r_j, w_j > to P_R(u)
    end for
end for
return Ṕ_R(u)
```

are fetched and enriched in the profile after calculating their new weights. Finally, the enriched profile $\acute{P}_R(u)$ is returned as an output.

Definition 7. ***Enriched Tag-based User Profile*** *An enriched tag-based user profile $\acute{P}_T(u)$ of a user u is an extension of the basic profile by tags with the highest degree of similarity to tags in $\overline{P_T(u)}$. Let T_u be the set of all tags in $\overline{P_T(u)}$ and w_i is the weight associated with tag i in the profile.*

$$\acute{P}_T(u) = \{< t_i, w_i > |$$

$$w_i = \begin{cases} w_i & , if t_i \in T_u \\ w_i * Max(Sim(t_i, t_j)) & , \forall (t_i \in \{T - T_u\} \wedge T_j \in T_u) \end{cases} \Bigg\}$$

A similar algorithm to that of enriching place profiles is used for choosing the tags and weights.

Definition 8. ***Enriched Spatio-Semantic User Profile***

An enriched spatio-semantic user profile $\acute{P}_{RT}(u)$ of a user u is an extension of the basic profile by tags and places with the highest degree of similarity to tags in $P_{RT}(u)$. Let T_u be the set of all tags in $\overline{P_T(u)}$, R_u be the set of all places in $\overline{P_R(u)}$ and w_{ij} is the weight associated with tag i and place j in the profile.

$$\acute{P}_{RT}(u) =< [r_i, t_j], w_u(r_i, t_j) > |w_u(r_i, t_j) =$$

$$\begin{cases} w_u(r_i, t_j) & , if\ r_i \in R_u and t_j \in T_u \\ w_u(r_i, t_j) * Max(Sim(r_i, r_k)) & , t_j \in P_T(r_k) \wedge r_k \in \{R - R_u\} \\ 0 & otherwise \end{cases} \Bigg\}$$

The spatio-semantic profile is extended with the most similar places to the user profile and these are assigned a weight computed using the place similarity value for all tags in their place-tag profiles and 0 for tags that are not in their profile. Thus the user simply inherits relationships with all the tags and their associated weights from basic places that are deemed similar to those in his profile.

User Profile Example

Here an example is given of a sample user profile created from the data-set used in this work. ‘user164’ checked in 200 different venues, with associated 82 venue categories. Note that one venue can have more than one venue category. Figure 4.2 shows the top 20 tags in his semantic user profile. Figure 4.3 shows filtered tags from his profile representing human activity (approximately 5% of all tags), as derived by mapping to Wordnet noun.act category. Figure 4.4 and 4.5 show the spatial profile and the

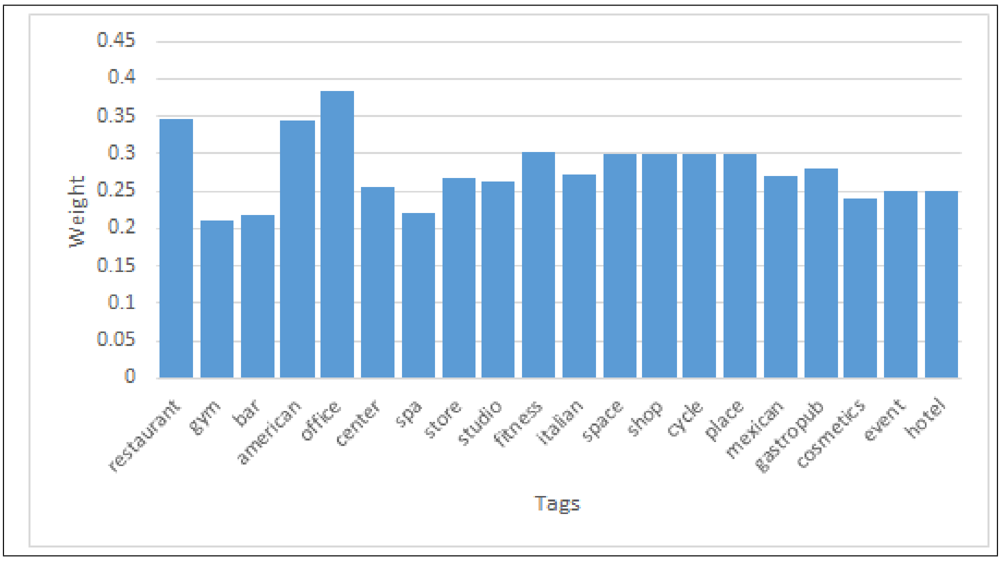


Figure 4.2: Example Semantic User Profile for User 'user164'.

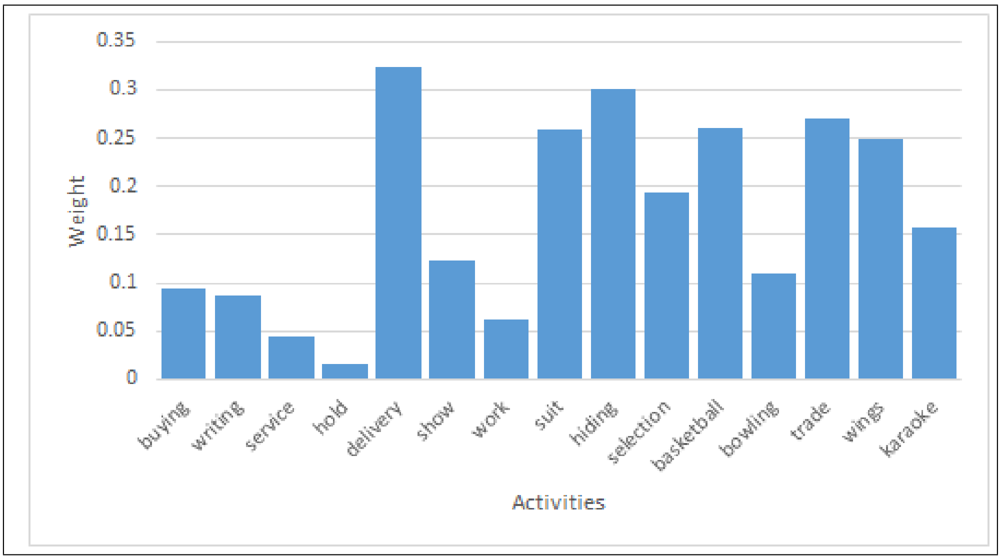


Figure 4.3: Activities in a Semantic User Profile for 'User164'.

enriched spatial profiles for user 'user164', respectively. $\gamma = 0.5$ was used in the place similarity equation of the enriched profile. The size of the dots in the figures represents the weight of the place in the profile.

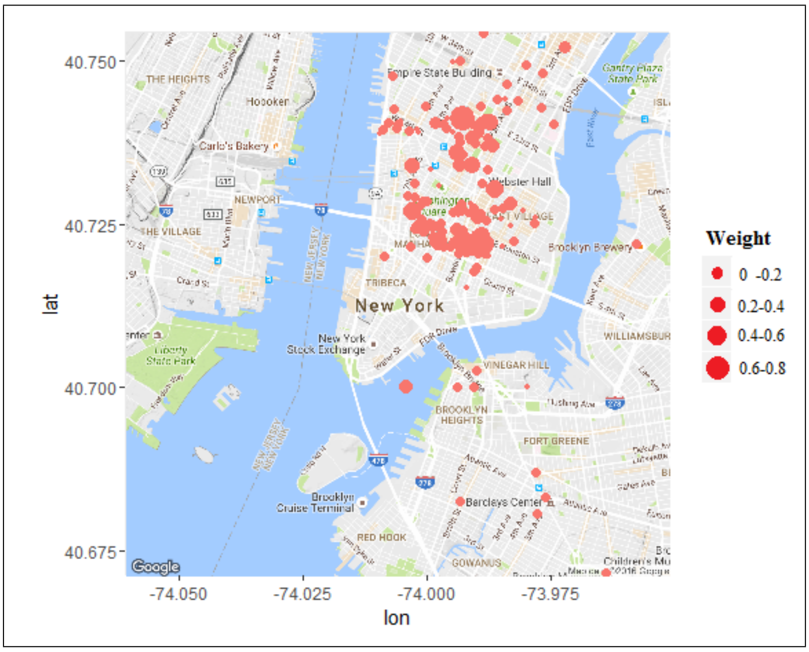


Figure 4.4: Spatial User Profile for User 'user164'.

4.4 Experiments and Results

Experiments in this chapter were carried out using a sample of two hundred users with a high frequency of check-ins, co-location rate and tips. Table 4.1 shows summary statistics of the sample data-set used.

Table 4.1: High Frequent Dataset

Number of Distinct Venues	10,988
Total number of Check-ins	4,212
Total Number of Tips	10,469
Total Number of Tags	13,396
Number of users	200
Total Number categories	459
Total Number of Relationships	165,453
Average places/user	121
Average tag/user	365

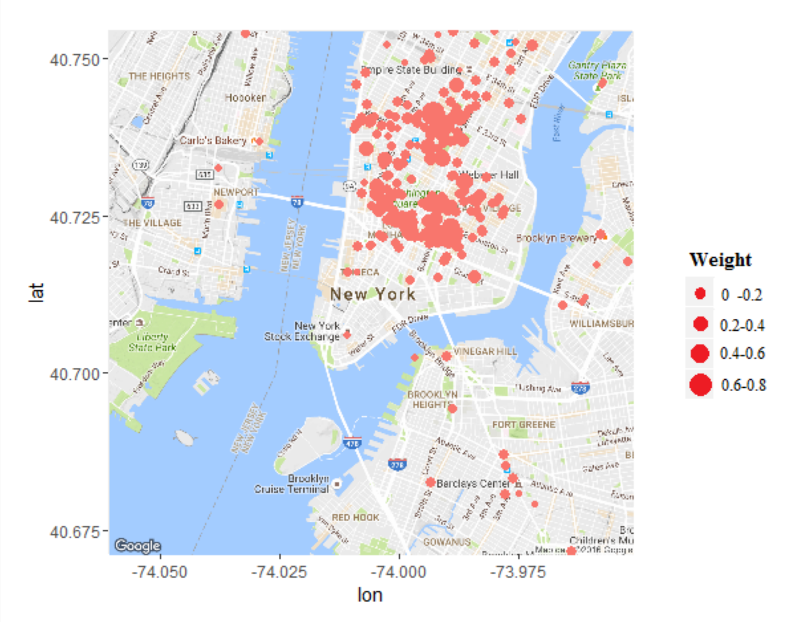


Figure 4.5: Enriched Spatial User Profile for user 'user164' with $\gamma = 0.5$.

4.4.1 Experiment Setup

The evaluation experiment aims to measure the impact of using the full range of content captured on LBSN when building user profiles in comparison to using only partial views based on check-in information. The experiment takes the form of place (and tag) top-N recommendation problem using the different constructed user profiles based on the users profiles cosine similarities and seeks to establish how well the profiles reflect the user spatial and semantic character when using the LBSN. Figure 4.2 show the Spatio semantic Top-K Recommendation algorithm.

The algorithm has two inputs: the number of top places to be recommended (K), and the γ value that controls the enrichment process. The algorithms starts with calling the spatial enrichment function discussed earlier to enrich the profiles with the required places, then the user similarity is computed between each two user profiles in the dataset using the $UserSim$ function. The top k places are recommended by finding the top similar user to each user. The Top-K places r_i, and weights w_i are then returned as the output of the algorithms.

Algorithm 4.2: Spatio-semantic Top-K Location Recommender

Input: γ,K.

Output: Top-K$< r_i, w_i >$

for each u_i **do**
  SpatialEnrichment($P_R(u_i), K, \gamma$)
end for
for all u_i, u_j **do**
  Fetch profiles $P_R(u_i), P_R(u_j)$
  Compute UserSim(u_i, u_j) .
end for
for each U_i **do**
  Fetch most similar user u_j
  Sort $< r_i, w_i >$ of $P_R(u_j)$
  Recommend top K r_i that are not in $P_R(u_i)$
end for
return Top-K$< r_i, w_i >$

We use recall@N, precision@N and F1@N as our success measures, where N is the predefined number of places (or tags) to be recommended. Recall measures the ratio of correct recommendations to the number of true places (or tags) of a test check-in or tip record, whereas precision measures the ratio of correct to false recommendations made. Recall and precision are given by

$$
\begin{aligned}
recall &= \frac{TP}{TP+FN} \\
precision &= \frac{TP}{TP+FP}
\end{aligned}
$$

where TP (true positives) is the number of correct place (or tags) recommended, FP (false positives) is the number of wrong recommendations and FN (false negatives) is the number of true place (or tags) which were not recommended. F1 is a combination of recall and precision and is given by

$$F1 = \frac{2 * precision * recall}{precision + recall}$$

The values of TP, FP, FN are determined by randomly splitting the users into two sets; the training set and the testing set. Multi-fold cross-validation was used to ensure a fair partitioning between test data and training data. Data were split 90% for training and 10 % for testing, and the process was repeated 5 times to create 5 folds and the mean of the performance was reported.

4.4.2 Evaluation of Different Similarity Measures

We used the basic spatial profile to train the user-based collaborative filtering method with different similarity measures. The similarity between users were measured using the cosine, jaccard, and pearson similarity measure. Table 4.2 shows the results of precision and recall when comparing different methods. The results of the three measure are almost the same, but the cosine similarity seems to be the best in terms of the F1-measure because it compares the two profile vectors and calculates the similarity based on the weights in each vector. Thus, we used cosine similarity to calculate similarity between users in our proposed recommendation method.

4.4.3 Evaluation of Spatial Profiles

Results for the enriched user profiles using the proposed top-N recommendation method are presented. Different versions of the enriched spatial profiles, using different place similarity measures were created, a) using $\gamma = 0$ (to represent place-tag similarity only), b) using $\gamma = 1$, (to represent place-user similarity only), and c) using $\gamma = 0.5$ for an aggregated view of both effects. We choose these three values of γ to show compare the results of using each similarity alone against the combined similarity. The

Table 4.2: Comparison between different similarity measures.

Top-N	Pearson Correlation		Jaccard Measure		Cosine Similarity	
	Precision	***Recall***	***Precsion***	***Recall***	***Precision***	***Recall***
1	0.5625	0.090385	0.575	0.096635	0.575	0.096635
2	0.2875	0.096635	0.2875	0.096635	0.30625	0.097381
3	0.195833	0.096706	0.204167	0.097594	0.216667	0.098632
4	0.153125	0.097707	0.15625	0.098054	0.1625	0.098632
5	0.1225	0.097707	0.1275	0.098379	0.1325	0.098703
10	0.07375	0.104646	0.0725	0.10139	0.0725	0.101038
20	0.046875	0.111808	0.03875	0.103073	0.04125	0.107751
30	0.03375	0.114669	0.030417	0.107076	0.031667	0.111437
40	0.028438	0.119753	0.024688	0.109422	0.028125	0.115961
50	0.026	0.12525	0.02225	0.11478	0.0255	0.123466

value 0.5 combines the two similarities equally and thus gives an insight of the benefit of combination. Hence, result sets are shown for the following user profiles. 1. Enriched-Spatial(Tag) 2. Enriched-Spatial(User) 3. Enriched-Spatial(All).

We compare the results of the top-N recommendation using the three different profiles with traditional Item-based Collaborative Filtering (IBCF) [100] and the User-based collaborative Filtering (UCBF) [89] approaches, applied against the basic spatial user profile. The results of the precision, recall and F1 measures for recommending top-1, 2, 3, 4, 5, 10, 20, 30, 40, 50 places are shown in Figures 4.6, 4.7 and 4.8, respectively.

As shown in the figures, results show that the enriched profiles perform consistently better than basic profiles using both UBCF and IBCF, with the best results achieved for the enriched user profiles. We also observer from results that the proposed method of recommendation using tag-based and user-based place similarity gives the best overall results. We conclude that enrichment with the combined similarity has more influence on place recommendation than using tag-based or user-based place similarity, which

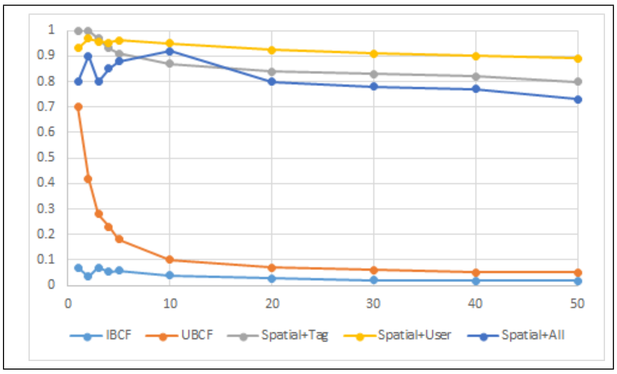


Figure 4.6: Precision Values for the Top-N Place Recommendations.

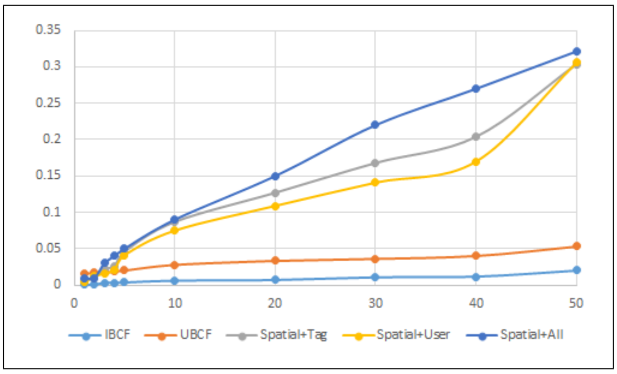


Figure 4.7: Recall Values for the Top-N Place Recommendations

reveals the importance of the spatial and semantic features in location recommendation. The results also demonstrates that the combined place similarity method improves the accuracy of the recommendation compared with the place similarities from users and

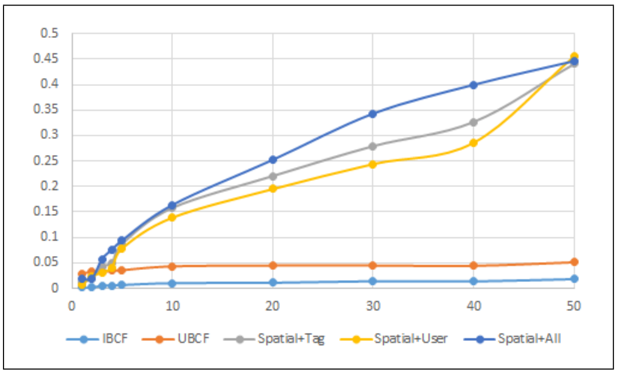


Figure 4.8: F1-Measure Values for the Top-N Place Recommendations

from tags. This shows that the influence of user preference on user location update gets better recommendation.

4.4.4 Evaluation of Semantic profiles

A similar experiment was carried out to evaluate the semantic user profiles. Again, the results were compared to the UBCF and IBCF approaches. Figures 4.9, 4.10 and 4.11 show the results of the top-10, 20, 30, 40, and 50 tag recommendations using the different methods. Results demonstrates the quality of the enriched user profiles, and thus confirm their utility for more accurate representations of user profiles.

The profile enrichment helps in constituting the missing data with inferred weights based on tag-similarity which solves the problem of data sparseness. Hence, the performance of top-k recommendation using the semantic enrichment method that uses the tag enrichment showed better results than the UBCF and IBCF that uses the basic semantic profile.

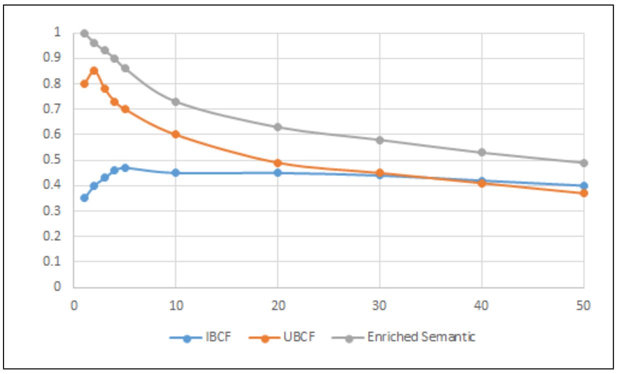


Figure 4.9: Precision Values for the Top-N Tag Recommendations.

We believe that the data used in this experiment is sufficient because it represents a sample of users in one region. Previous experiments were done on the smaller data set of 20 users. The statistics of the data-set is listed in table 4.3. It is noted that the same conclusions is drawn from the small dataset as shown in figures 4.12, 4.13, 4.14. This

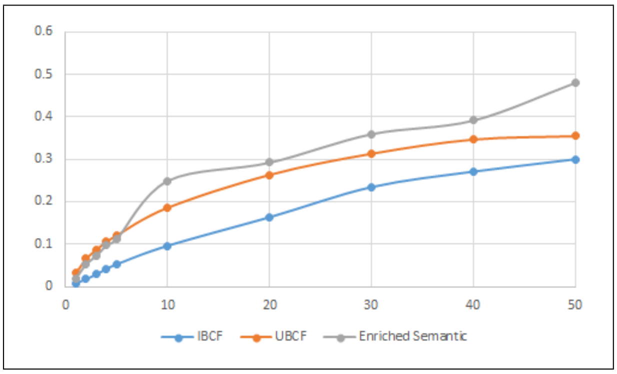


Figure 4.10: Recall Values for the Top-N Tag Recommendations.

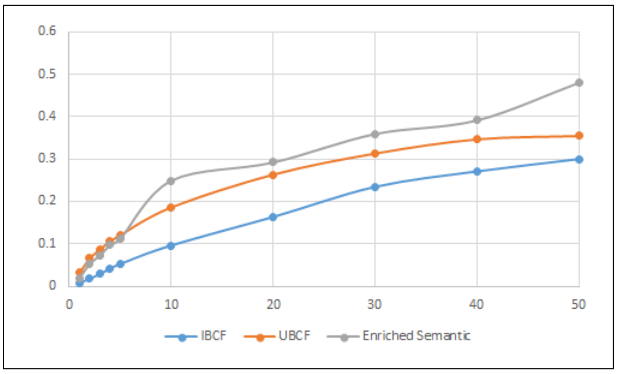


Figure 4.11: F1-measure Values for the Top-N Tag Recommendations.

assures that the enriched spatial combined profile is the best profile for representing the spatial user profile.

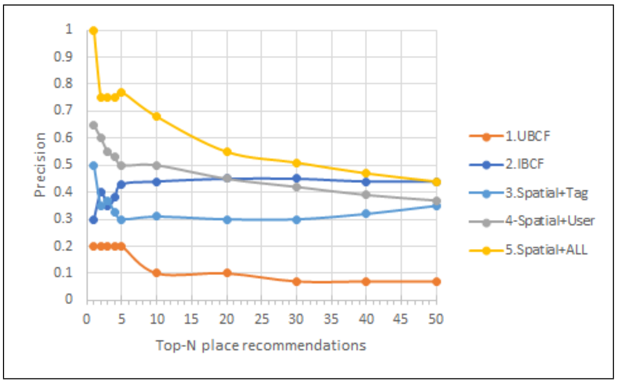


Figure 4.12: Precision values for the top-N place recommendations-Small Data-set..

Table 4.3: Small DATASET

Number of Venues	2,041
Total number of Checkins	4,212
Total Number of Tips	942
Total Number of Tags	3,357
Number of users	20
Total Number categories	317
Total Number of Relationships	17,955
Average Checkins/user	601
Average tag/user	363

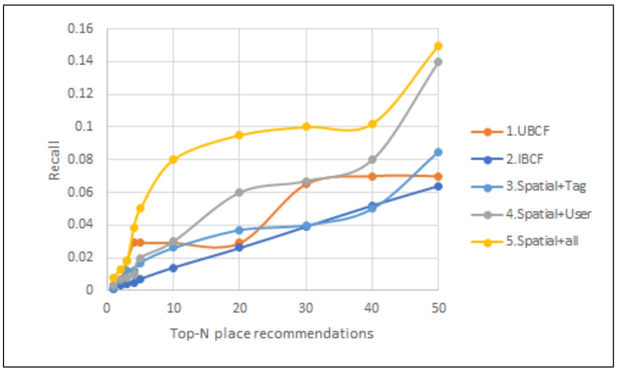


Figure 4.13: Recall values for the top-N place recommendations-Small Dataset

4.5 Summary

This work considers the problem of user profiling on location-based social networks. This chapter was able to answer successfully the first two research questions which are How can different views of user profiles be constructed from user footprints collected on LSBNs that emphasis the different facets of collected data? How does the enrichment process impact the quality of personal user profiles? Both the spatial (where) and the semantic (what) dimensions of user and place data are used to construct different views of a user's profile. A place is considered to be associated with a set of tags or labels that describe its associated place types, as well as summarise the users' annota-

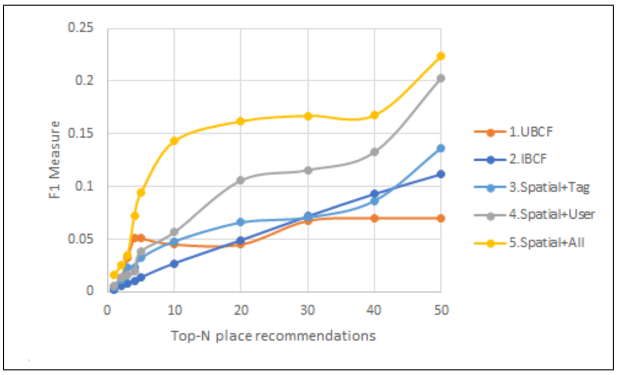


Figure 4.14: F1 measure values for the top-N place recommendations-Small Data-set.

tions in the place. A folksonomy data model and analysis methods are used to represent and manipulate the data to construct user profiles and place profiles. It is shown how user profiles can be extended from a basic model that describes user's direct links with a place, to an enriched profiles describing richer views of place data on the social network. The model is flexible and can be adjusted to focus on the spatial and semantic dimensions separately or in combination. Flexibility means the ability for the solution to adapt to possible or future changes in its requirements. The model proposed can be adopted to include more dimensions using the same framework. Temporal dimension can be adopted as discussed in Chapter 6. Also, the model can be adopted to spatial types and semantic activities as will be explained in Chapter 7. Results demonstrate that the proposed methods produce user profiles that are more representative of user's spatial and semantic preferences. This chapter answered the first two research questions proposed in this thesis. Also, this chapters proposed a location recommendation algorithm that was used to evaluate the proposed profiles against existing methods of recommendation. So, this chapter also contributed in answering research question which states How can we construct a new location recommendation method using different dimension of LBSNs and evaluate it existing methods?. These three questions will also be revisited in chapter six when we explain the dynamic user modelling methods.

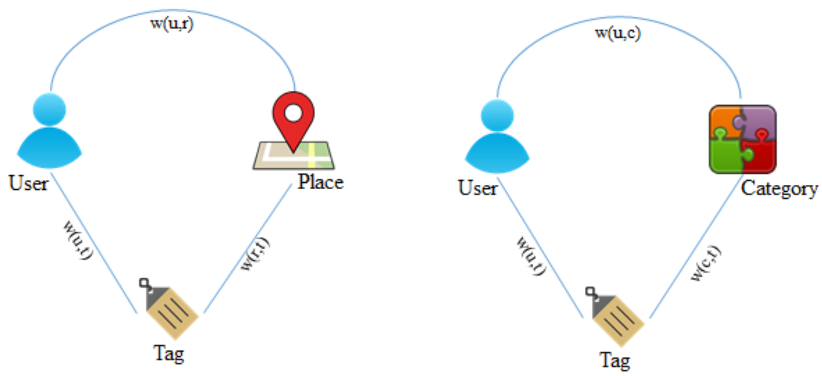
Chapter 5

Place Profiles in Geo-Folksonomies

5.1 Introduction

Foursquare provides a type attached to each place that is known as a place category. These categories build a category hierarchy based on the relationship between categories. There are eight main categories in Foursquare: Arts and Entertainment, College and University, Events, Food, Nightlife Spots, Outdoors and Recreation, Professional and Other Places, and Transport. These categories contain about 525 subcategories. Every user is related to a place and its place categories, and every place is related to its place category. The place can have more than one category assigned to it.

The aim of this chapter is to study each category based on the users' behaviour in places, and thus construct a behavioural place category profile. This chapter will also cover the third research question: 'how can implicit semantics of place profiles be used to reflect users experience in geographic places through the activities they carry out in those places?'. In this research, place semantics is represented by place categories and place activities. Towards this, the first step of constructing the category profile is to replace each place instance of its place category and construct the relationship between it and the user and the tag using the geo-folksonomy model. Thus, relationship Y becomes the relationship between place categories, users and tags. Figure 5.1 shows the different structures of geo-folksonomy. Figure 5.1a shows the geo-folksonomy structure among places, users, and tags, and Figure 5.1b shows the geo-folksonomy



(a) Place, User, Tags (b) Place Category, User, Tags

Figure 5.1: Geo-Folksonomy Structure

structure using place categories, users and tags.

5.2 Definitions

5.2.1 Basic User Profiles

Definition 9. ***Category-based User Profile*** *A category user profile $P_C(u)$ of a user u is deduced from the set of categories of the places that u visited or annotated directly.*

$$P_C(u) = \{(c, w(u,c)) | (u,t,c) \in Y,$$
$$w(u,c) = |\{t \in T : (u,t,c) \in Y\}|\}$$

$w(u,c)$ is the number of tag assignments, where user u assigned some tag t to a place category c through the action of checking-in or annotation. Hence, the weight assigned to a place category simply corresponds to the frequency of the user reference to the place category either by checking in or by leaving a tip.

Definition 10. ***Parent Category User Profile*** *A parent category user profile $P_{C_p}(u)$ of a user u is deduced from the set of parent categories of the places that u visited or* *annotated directly.*

$$P_{C_p}(u) = \{(c, w(u,p)) | (u,t,p) \in Y,$$
$$w(u,p) = |\{t \in T : (u,t,p) \in Y\}|\}$$

$w(u,p)$ is the number of tag assignments, where user u assigned a tag t to a parent place category p through the action of checking-in or annotation. Hence, the weight assigned to a parent place category simply corresponds to the frequency of the user reference to the place category either by checking in or by leaving a tip.

Figure 5.2 shows an example of a spatial user profile and their category-based user profile.

5.2.2 Basic Category Profiles

Let $P_T(c)$ and $P_U(c)$ be the tag-based category profile and user-based category profile for place category c (defined in a similar manner to the user profiles above). Conceptually, a tag-based category profile is a description of the place categories by the tags assigned to it and a user-based category profile is an account of users' visits to the place categories. They are defined as follows:

Definition 11. ***Tag-based Category Profile*** *A tag-based place profile $P_T(c)$ of a place category c is deduced from the set of tag assignments linked with c.*

$$P_T(c) = \{(t, w(c,t)) | (u,t,c) \in Y,$$
$$w(c,t) = |\{c \in C : (u,t,c) \in Y\}|\}$$

$w(c,t)$ is the number of tag assignments where user u assigned tag t to some place categories through the action of checking-in or annotation.

Definition 12. ***User-based Category Profile*** *A user-based place profile $P_U(c)$ of a place category c is deduced from the set of users that u visited or annotated the place*

(a) Spatial User Profile for user900

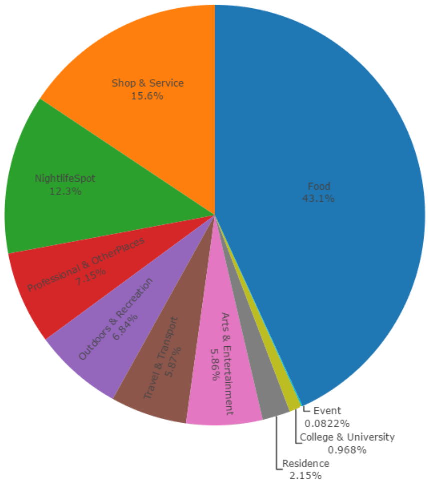
(b) Category-based User Profile for user900

Figure 5.2: An Example of Spatial vs. Category User Profile

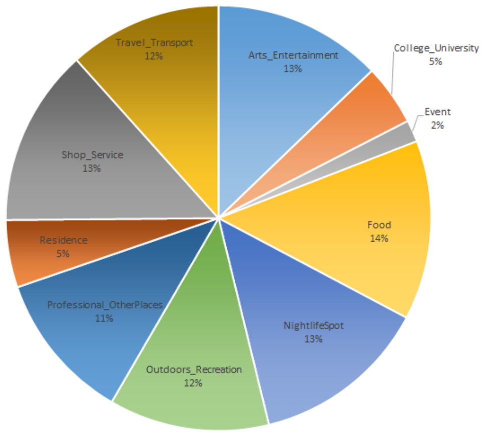
category directly. It represents the relationship between a user and a place category.

$$P_U(c) = \{(u, w(c,u)) | (u,t,c) \in Y,$$
$$w(c,u) = |\{t \in T : (u,t,c) \in Y\}|\}$$

$w(c,u)$ is the number of tag assignments, where u assigned some tag t to place category c through the action of checking-in or annotation. Hence, the weight assigned to a



(a) Place Distribution for Main Categories



(b) User Distribution for Main Categories

Figure 5.3: A Pie-Chart for Place and User Distributions in Categories

place category simply corresponds to the frequency of the user reference to the place cateogry either by checking in or by leaving a tip.

5.3 Dataset

Figure 5.3aAs expected, the category that includes the largest number of places is Food. Shop & Service, and Nightlife Spots have a substantial set of places, while the remaining categories hold less than 10% of the total places. Figure 5.3b shows the distribution of users across the main categories of Foursquare. It is observed from the Figure that the users are distributed equally, except for amongst the event, college and university, and residence categories.

Table 5.1 show each category and the number of its subcategories deducted from the data set.

Table 5.1: Subcategories Distribution for Main Categories

Category Name	**Subcategories count**
Arts & Entertainment	45
College & University	35
Event	7
Food	127
NightlifeSpot	19
Outdoors & Recreation	65
Professional & OtherPlaces	68
Residence	4
Shop & Service	111
Travel & Transport	41

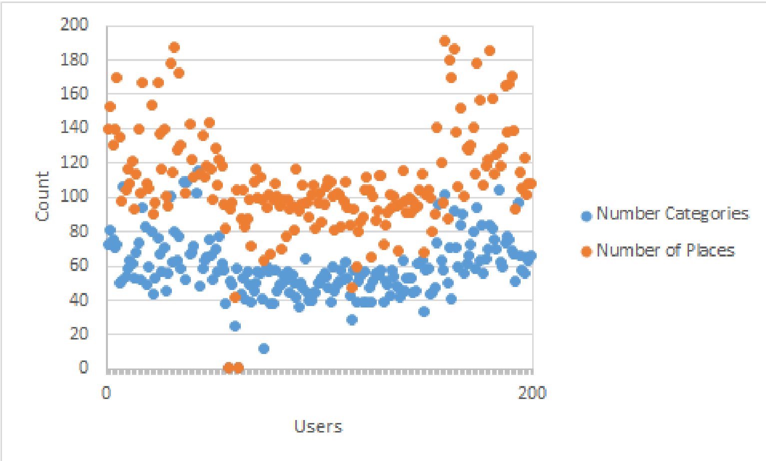
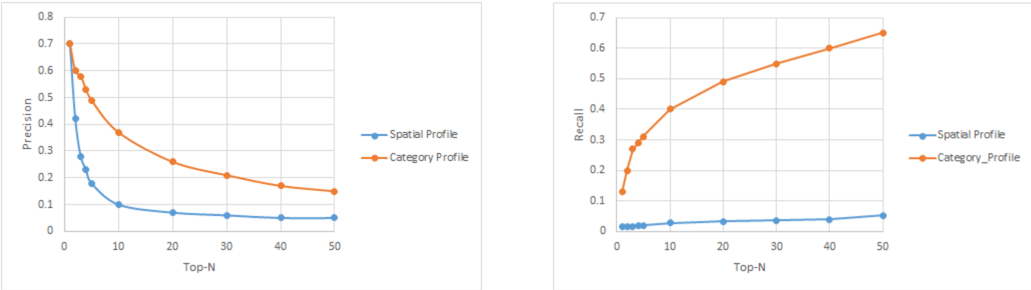


Figure 5.4: The Number of Distinct Categories and Places for Users

5.4 Evaluation of Category-based User Profile

As mentioned in the above definitions, the category-based user profile is a clustered view of the spatial profile. The user is related to the category of a place rather than the places themselves. The category of a place implies the property of the place and what kind of activities attract the user through their history of checking-in. Thus, with



(a) Precision (b) Recall

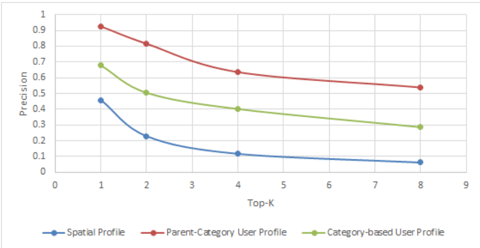
Figure 5.5: Evaluation of Category versus Spatial User Profile

the help of the category information of the place, the specific preference of a user can be obtained. A place grouping is performed to place instances in the spatial profile. Figure 5.4 shows the place and category distribution for each user in the dataset. The average number of categories per user is half the average number of places visited or tipped.

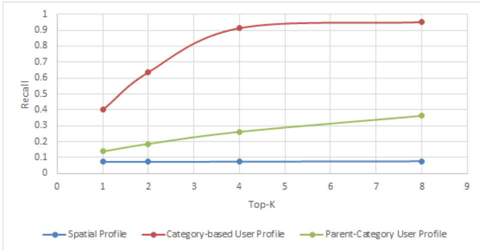
The user model based on place category maps user's preference from places to place categories. As a result, a reduction of the user model is achieved. The traditional high-dimensional user-place rating model is transformed into low-dimensional user-place category rating statistical model.

Using a category-based user profile, relevant categories can be recommended using collaborative based filtering by calculating the similarity between users. This similarity is calculated using the cosine similarity between category-based user profiles. Thus, the similarity deduced represents the common categories between two users. The similarity value is 1 if they are very similar, and 0 represents no relation.

To evaluate the importance of the category recommendations, the place recommendation versus the category recommendation category based user profile and spatial user profiles was compared. Then, the precision and recall were calculated as our evaluation metrics. Figure 5.5a and 5.5b show the precision and recall results for evaluating the spatial and the category profile. The precision and recall using the category-based user profile outperforms the spatial user profile.



(a) Precision



(b) Recall

Figure 5.6: Evaluation of Parent Category, Category and Spatial User Profile

Also, user's interest can be modelled with greater granularity using the parent category user profile. In this profile, the categories in the category-based user profiles are grouped to the nine main categories provided by Foursquare. This makes the profile less spare and reduces its dimension. We compared the spatial, category-based, and spatial user profiles using user-based collaborative filtering. Figure 5.6 shows the precision and recall of Top-N recommendation. As the figure shows, when increasing the granularity of the user model, the parent category user profile shows the best values of precision and recall. Only the top-k values of 1, 2, 4, and 8 were shown because of restriction by the eight main categories, and thus the other profiles with the same k values were compared.

5.5 Category Clustering

As mentioned in the above definitions, there are two types of category profiles: the user-based category profile, and the tag-based category profile. When calculating the category similarity between the profiles, the cosine similarity equation is applied to both profiles. Each similarity matrix has values that range from 0 to 1. In order to evaluate the different category similarities, the category profile is clustered into four classes. Very High, High, Moderate, and Low. The process of clustering is carried out by sorting the weight in the category profile descending, and then calculating the four quantiles. These clusters represent the relationship between the category and its tags or its users. The m-quantiles of the underlying distribution of the $(min; Q_1; ...; Q_m - 1; max)$ observed data values are used, which is known as Symbolic Clustering Based on Quantile Representation.

Table 5.2 shows an example of the category profile. Category similarity 1 is the similarity between user-based category profiles, and category similarity 2 is the similarity between the tag-based category profile. As the table shows, the similarity between categories from different profiles is different. In category similarity 1, the highest similarities are all categories that contains 'school' word , and all categories have the same parent category 'Professional and Other Places'. However, in category similarity 2, the highest similarity to category school reflects the behaviour of the users in the places, which means for example that people who visited the sub-category school also visited the college history building and college classroom. It is important to note that the similarity between the place categories (place types) depends on the users' behaviour. So, we expect that for different regions, the similarity between categories can differ according to the behaviour of the users in that particular region. So, it is important to suggest relevant category types to the user.

Category similarity was also evaluated using a recommendation method that calculates the category cosine similarity and uses it to recommend new tags or users. Figure 5.7a and 5.7b show the evaluation results. It is observed from the figure that the recall values for the user-based category profile have higher values than the tag-based category profile.

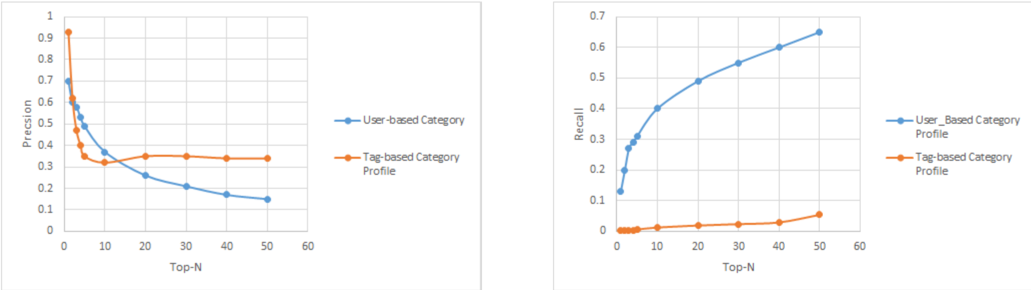
5.6 Place related Human Activities

As people share their experience, emotions, and opinions about places on the geosocial web, it is important to extract place semantics from the tags attached to them. Place semantics help in the discovery of structured knowledge from unstructured data (tags). Discovering semantics from tags can make the tags more useful, and thus will

Table 5.2: Different Clusters of Category Similarities

Cluster	***Rank***	***Category Similarity1***	***Category Similarity 2***
Very High	***1***	College History Building	High School
	2	College Classroom	Elementary School
	3	Board Shop	Nursery School
	4	Luggage Store	Law School
	5	Cemetery	Trade School
High	***1***	Breakfast Spot	Veterinarian
	2	Toy / Game Store	Malaysian Restaurant
	3	Science Museum	City Hall
	4	Irish Pub	Bike Rental / Bike Share
	5	Seafood Restaurant	Mediterranean Restaurant
Moderate	***1***	Pier	Greek Restaurant
	2	Home (private)	Synagogue
	3	Bridge	Piano Bar
	4	Medical Center	Hotel
	5	Belgian Restaurant	Ukrainian Restaurant
Low	***1***	Cuban Restaurant	Post Office
	2	Vietnamese Restaurant	Hookah Bar
	3	Bike Shop	Juice Bar
	4	Club House	Bike Shop
	5	Middle Eastern Restaurant	Vegetarian / Vegan Restaurant

be of great benefits to users interacting with the geo-social web. Extraction of place semantics is important for many applications such as place search, tag recommendation for places, and the inference of place information for untagged places in other geo-social applications like Twitter. An important semantic related to place is discovering the place activities from tags. In other words, it is desirable to construct a model



(a) Precision (b) Recall

Figure 5.7: Category Similarity Evaluation

for activity recognition in places. Towards this, the tips in Foursquare can be mined to extract, categorise, and map people's activities. Thus, we want to answer the question of "what activities are there in nearby places?". By identifying these activities, this chapter will answer the research question of how can implicit semantics of place profiles be used to reflect users experience in geographic places through the activities they carry out in those places?. As activities are extracted from tags, an activity can be associated with a place instance, place type, or a user. Moreover, activities can be modelled using the geo-folksonomy model by replacing the tag instances with activity instances. We define different activity related profiles in the following sub-section.

5.6.1 Activity Related Profiles

Definitions

Definition 13. ***Activity User Profile*** *An activity user profile $P_A(u)$ of a user u is deduced from the set of the places that u visited or annotated directly with activity tags.*

$$P_A(u) = \{(r, w(u,r)) | (u,a,r) \in Y,$$
$$-w(u,c_r) = |\{t \in T : (u,a,r) \in Y\}|\}$$

$w(u,r)$ is the number of tag assignments, where user u assigned some activities a to a place r through the action of checking-in or annotation. Hence, the weight assigned to

a place simply corresponds to the frequency of the user reference to the place either by assigning an activity tag to a place.

Definition 14. ***Activity-based Category Profile*** *An activity-based category profile $P_A(c)$ r is deduced from the set of activity assignments linked with category C.*

$$P_A(c) = \{(a, w(c, a)) | (u, a, c) \in Y,$$
$$w(c, a) = |\{c \in C : (u, a, c) \in Y\}|\}$$

$w(c, a)$ is the number of tag assignments where user u assigned tag a to a place that belongs to category c through the action of checking-in or annotation.

$\overline{P_A}$ refers to the activity profile where the sum of all weights is equal to 1, with $\overline{w}(c, a) = \frac{N(c,a)}{N_R(c)}$, where $N(c, a)$ is the number of users that annotated place category c with a and $N_R(c)$ is the total number of users who use activity tags to annotate place category c.

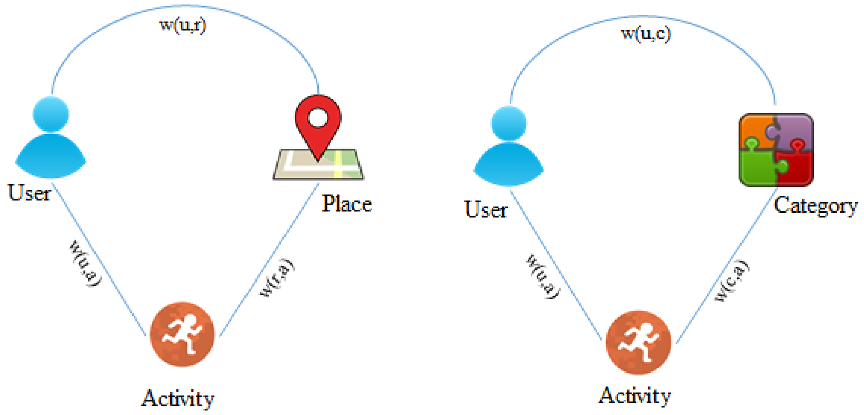
Definition 15. ***Activity-based Place Profile*** *An activity-based place profile $P_A(r)$ r is deduced from the set of activity assignments linked with place r.*

$$P_A(r) = \{(a, w(r, a)) | (u, a, r) \in Y,$$
$$w(c, a) = |\{c \in C : (u, a, r) \in Y\}|\}$$

$w(r, t)$ is the number of tag assignments where user u assigned activity a to some place that belongs to place r through the action of checking-in or annotation.

$\overline{P_A}$ refers to the activity profile where the sum of all weights is equal to 1, with $\overline{w}(r, a) = \frac{N(r,a)}{N_R(c)}$, where $N(r, a)$ is the number of users that annotated place r with a and $N_R(r)$ is the total number of users who use activity tags to annotate place category r.

As the definitions show, an activity performed at a particular location represents some ‘doing’ sense. A user can be related to an activity, a places or a place category. Figure 5.8 the two geo-folksonomy structures using activities. Figure 5.8a shows the relationship between a user, an activity and a place, and Figure 5.8b shows the relationship between a user, an activity and a place category. Figure 5.9 shows an example of



(a) Activity Geo-Folksonomy using place instances

(b) Activity Geo-Folksonomy using place category instances

Figure 5.8: Activity Geo-Folksonomy Structure.

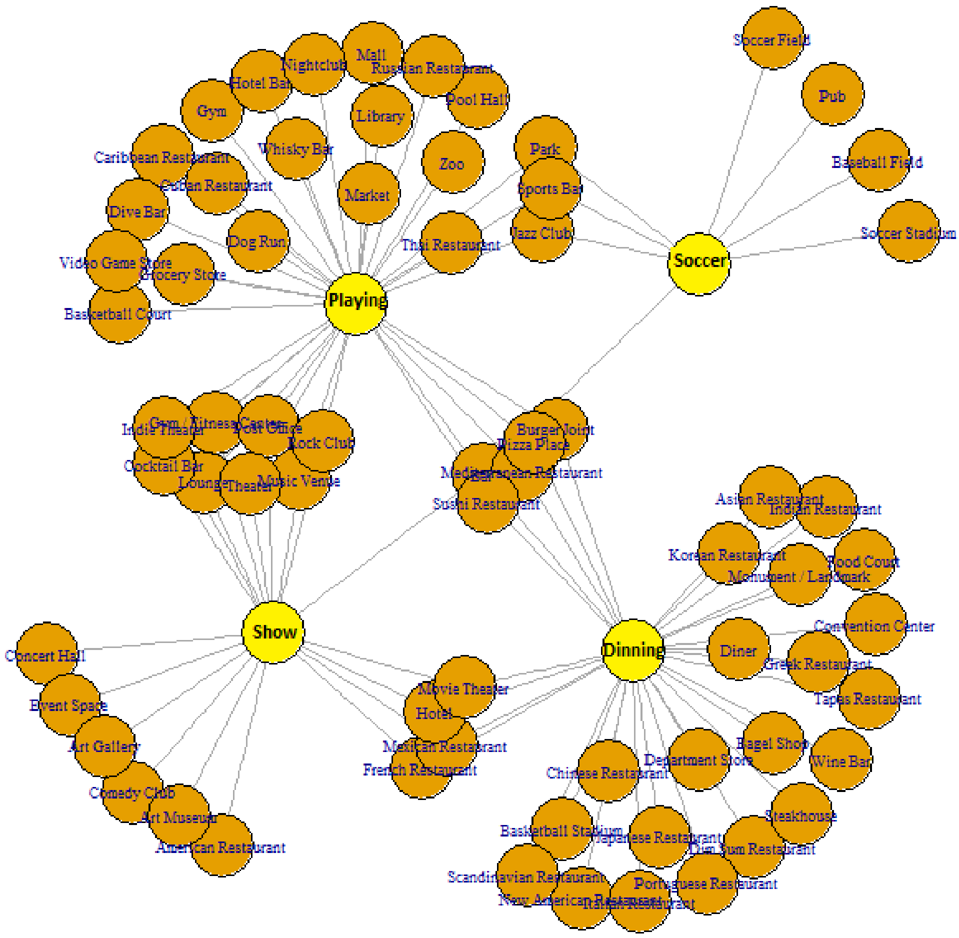


Figure 5.9: An Example of Activities Related to Place Categories.

four activities and their relationship between different categories. As the Figure shows, places categories may contain more than one activity and activities are shared among different categories.

Figure 5.10 shows the MAP, MSE, and MAE values when we compared the activity user profiles with the tag user profile using user-based collaborative filter. Although the semantics profile results is more promising than activity profile results, it is important to recommend location-specific activities to users. Also, using the activity similarity for enriching profiles can enhance the activity recommendation.

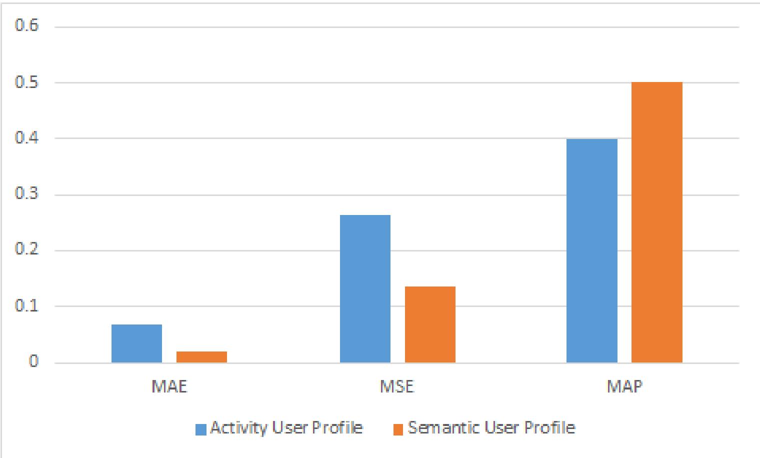


Figure 5.10: Recall values for Tag-based Category User Profile and Activity-based Category User Profile.

5.7 Activity-aware Category Recommendation

In this section a new category recommendation method is proposed. The recommendation method is considered as a hybrid recommendation method. It recommends categories similar to the categories in the category-based user profile (content filtering recommendation) (list1) and then it enriched list1 into the category-based user profile and use the user similarity to recommend categories from high similar user (list2) (collaborative-based filtering). In this algorithm, the similarity between categories is

Algorithm 5.1: CategoricalEnrichment ($P_C(u)$,**K**,γ)

Input: A Category-based User Profile $P_C(u)$, K,γ.

Output: Enriched Category-based User Profile $\acute{P}_C(u)$

for all places C_i in Category-based User Profile $P_C(u)$ **do**
    if $\gamma = 1$ **then**
        Compute $Cat_C osSim_{activity}(r_1, r_2)$
    else
        if $\gamma = 0$ **then**
            Compute $Cat_C osSim_{User}(r_1, r_2)$
        else
            Compute $CatSim(r_1, r_2, \gamma)$
        end if
    end if
    Find top K similar places c_j to each c_i in $P_c(u)$
    for each $< c_j, sim >$ in top similar places **do**
        $w_j = w_i * sim$
        add $< c_j, w_j >$ to $P_c(u)$
    end for
end for
return $\acute{P}_c(u)$

calculated using the activity-based category profile, or the user-based category profile profile or a combination from both.

Cosine similarity between activity-based place profiles ($CatSim_{activity}(c_1, c_2)$) and between user-based category profiles ($CatSim_{user}(c_1, c_2)$) construct an activity-oriented ranking and tag-oriented ranking, respectively. These similarity rankings can be aggregated using the so-called Borda method [30] to compute a generalised similarity score between two place categories as depicted in equation 5.1. By setting the value of gamma to 0.5, the similarity score between two place categories using its activities is summed up

Algorithm 5.2: Spatio-semantic Top-K Category Recommendation

Input: The combination factor γ, and the number of top categories to recommend K.

Output: Top-K recommended categories$< c_i, w_i >$

```
for each u_i do
    CategoricalEnrichment(P_C(u_i)), γ)
end for
for all u_i, u_j do
    Fetch profiles P_C(u_i), P_C(u_j)
    Compute User-Sim(u_i, u_j) .
end for
for each U_i do
    Fetch most similar user u_j
    Sort < c_i, w_i > of P_C(u_j)
    Recommend top K c_i that are not in P_C(u_i)
end for
return Top-K< c_i, w_i >
```

with the similarity score between the same two place categories using its users. This gives an equal balance between the two similarity measures and thus the result will be a generalised similarity score between two place categories.

$$CatSim(c_1, c_2) = \gamma * Cat_C osSim_{activity}(c_1, c_2) + (1 - \gamma) * Cat_C osSim_{user}(c_1, c_2) \tag{5.1}$$

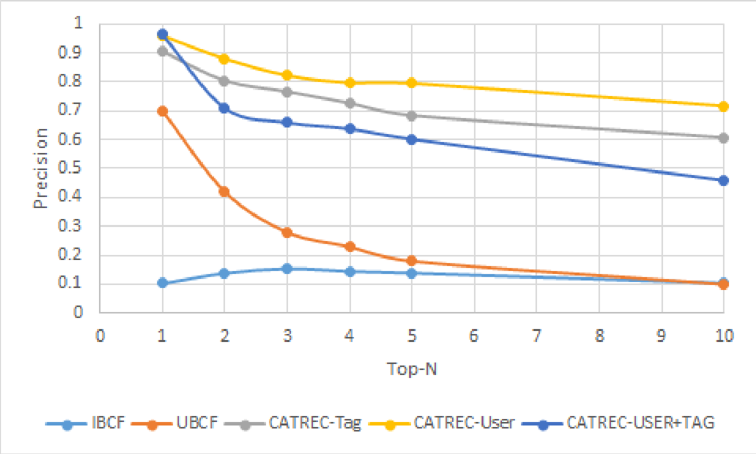
where $0 \leq \gamma \leq 1$ is a parameter that determines the balance of importance given to similarity scores from $P_A(c)$ and $P_U(c)$. Conceptually, similarity between two place categories is a function of the overlap between their activity assignments only (for $\gamma = 0$), a measure of their common visitors only (for $\gamma = 1$), or both (for γ between 0 and 1).

Algorithms 5.1 and 5.2 are used to do the recommendation which are similar to the two proposed methods in Chapter 4.

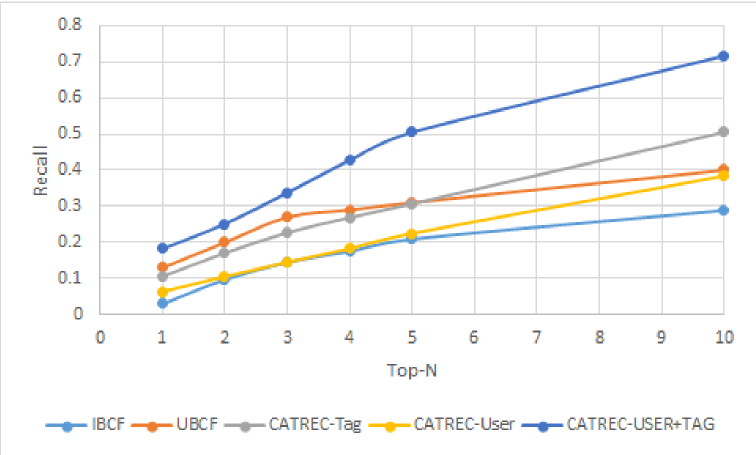
Figures 5.11a, 5.11b, 5.11c show a significant increase in precision and recall when using the combined method to predict categories.

5.8 Summary

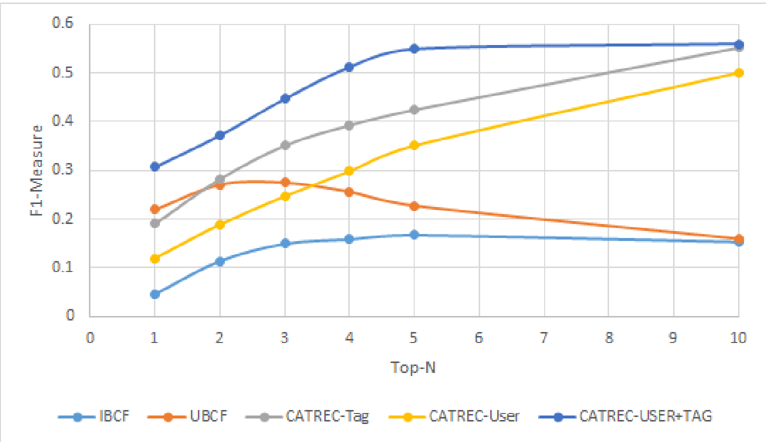
This Chapter propose a place model based for extracting place semantic from LBSNs, namely place categories and place activities. The chapter aimed to answer the question of how can implicit semantics of place profiles be used to reflect users experience in geographic places through the activities they carry out in those places? Towards this, Category-based and activity based user profiles are also proposed and evaluated against basic spatial and semantic profiles conducted in Chapter 4 using collaborative filtering recommendation algorithm. A behavioural category clustering method is also proposed and evaluated. Finally, an activity aware category recommendation is proposed and evaluated against basic recommendation methods. Experiment results shows that the proposed model can effectively deal with sparsity score data, which enhance the quality of recommendation.



(a) Category Recommendation-Precision



(b) Category Recommendation-Recall



(c) Category Recommendation-F1-Measure

Figure 5.11: Category Recommendation

Chapter 6

Dynamic User Modelling

6.1 Introduction

Users' interests are not static, they always change as time goes by and their interests in places also changes. In this chapter, we propose a temporal Geo-folksonomy model to analyse users' behaviours. The model learns users' preferences by extracting keywords from Tips over a period of time, and then the impact of time is considered to deal with interest drifts. Using this model, we are also proposing a new time-aware location recommendation algorithm. This Chapter follows the same steps of Chapter 4, but with temporal extension of the user model. Thus, it aims to answer the following research questions:

- How can different views of user profiles be constructed from user footprints collected on LSBNs that emphasis the different facets of collected data?
- How does the enrichment process impact the quality of personal user profiles?
- How can we construct a new location recommendation method using different dimension of LBSNs and evaluate it existing methods?

6.2 Temporal Geo-Folksonomy (TGF) Model

Different representations of time context information can be used in this model. For example, time may be represented as a continuous variable whose values are the specific times at which places are tagged (e.g. a timestamp like "January , 1st, 2010 at 00:00:00"). Another option is to specify meaningful categorical values with time intervals that characterise people's visit patterns to places. For instance, continuous time units can be defined by days or months, or categorical groups of time units can be defined, e.g. weekends and weekdays, seasons, etc. In this sense, storing the timestamp of tagging is the most flexible option, since it enables the exploitation of diverse representations of the time context. Thus, in general, tag assignments in the folksonomy are associated with a time stamp, that is, an instance of the set of time units chosen for representation. For example, a folksonomy representing users' interactions clustered over different months of the year may contain assertions of the form:
$(u, r1, \{g1, g2\}, January)$
$(u, r1, \{g1, g3\}, January)$
$(u, r2, \{g1, g4\}, January)$
representing user's u checkins and tips in place r1 and r2 in January. Different views of folksonomy can be thus be created using different time units.

The data capturing process results in the creation of a *geo-folksonomy*, which can be defined as a quintuple $\mathbb{F} := (U, G, R, T, Y)$, where U, G, and R are finite sets of instances of users, tags and places respectively, T is a set of time intervals at which the data were captured, and Y defines a relation, the tag assignment, between these sets, that is, $Y \subseteq U \times G \times R \times T$,

A geo-folksonomy can thus be transformed into a finite set of tripartite undirected graphs: $\{\mathbb{G_F}^1, \ldots \mathbb{G_F}^n\}$, where n is the number of time units defined in T. Let $\mathbb{F}^i = (U^i, G^i, R^i, I^i)$ be the restriction of $\mathbb{F}$ to t_i, such that, G^i and R^i are finite sets of tags and places respectively, that are referenced from tag assignments performed by the set of users U^i at t_i, and I^i defines a relation between these sets: $I^i := \{(u, g, r) \in$

$U^i \times G^i \times R^i | (u, g, r) \in Y\}$.

A geo-Folksonomy graph at time slot t_i defined as $\mathbb{G}_{\mathbb{F}}{}^i = (V_{\mathbb{F}^i}, E_{\mathbb{F}^i})$, is an undirected weighted tripartite graph that models a given folksonomy $\mathbb{F}^i$, where: $V_{\mathbb{F}^i} = U^i \cup G^i \cup R^i$ is the set of nodes, $E_{\mathbb{F}^i} = \{\{u, g\}, \{g, r\}, \{u, r\} | (u, g, r) \in I^i\}\}$ is the set of edges, and a weight w is associated with each edge $e \in E_{\mathbb{F}^i}$.

The weight associated with an edge $\{u, g\}, \{g, r\}$ and $\{u, r\}$ corresponds to the co-occurrence frequency of the corresponding nodes within the set of tag assignments I^i. For example, $w(g, r) = |\{u \in U^i : (u, g, r) \in I^i\}|$ corresponds to the number of users that assigned tag g to place r at time slot t_i.

Figure 6.1 depicts the overall process of user profile creation.

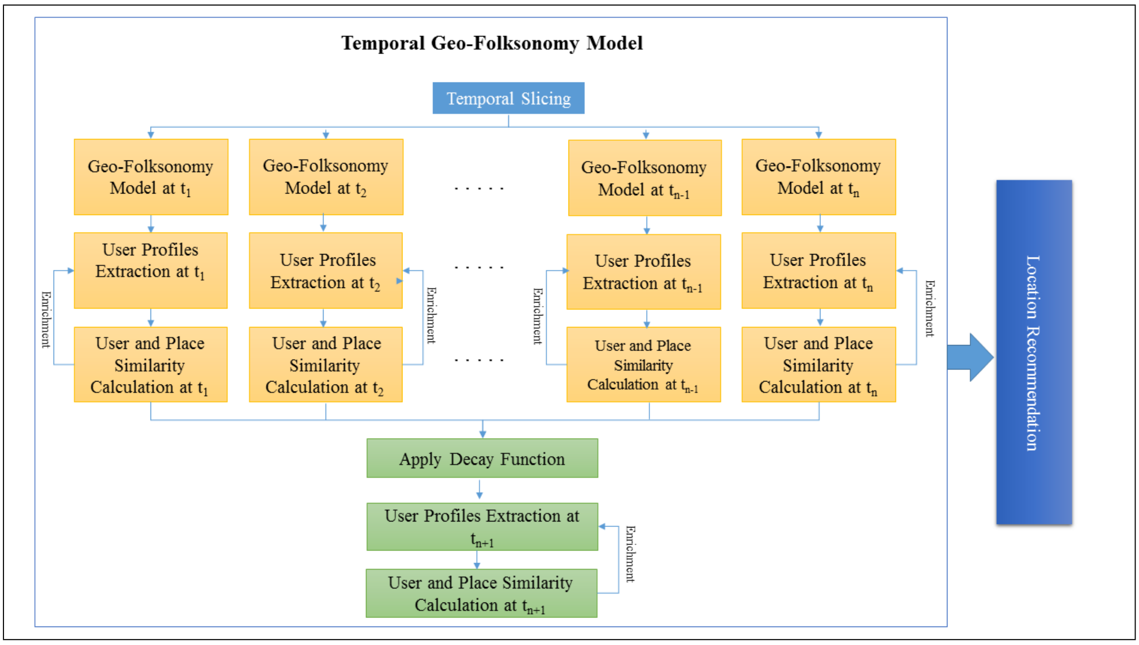


Figure 6.1: The Temporal Geo-Folksonomy Framework

6.3 User Modelling

User modelling is used here to learn more about user preferences and interests, and enhance recommendation and personalization systems. The user interests describes the association of a user to places and concepts in a region of interest.

Thus, an approach to modelling users in LBSN that represents a user's spatial, semantic and combined spatio-semantic association with place is proposed. A spatial user profile represents the user's interest in places, while a semantic profile describes their association with concepts associated with places in the folksonomy model. A spatio-semantic profile describes the user combined interest in specific place-concept associations.

Two different method for modelling time in user profiles are used here, denoted, the *time-slice* approach and the *decay* approach. In the time-slice approach, user profiles are simply computed from the geo-folksonomy temporal graph ${\mathbb{G}_{\mathbb{F}}}^i$ for any time slot of interest t_i, whilst the other folksonmy graphs for $t \neq t_i$ are discarded. For example, in a geo-folksonomy representing data collected over a year and aggregated by months, a user profile in the month of June will be computed only from the folksonomy graph for June, etc. The decay approach, on the other hand, considers the historical interactions in all sub graphs of the folksonomy prior to the time point of interest. Thus, user profiles for the month of June will reflect users' interactions from all the folksonomy graphs from January to June.

In both approaches, a user profile is built in stages. Starting with a basic profile that utilises direct check-in and annotation histories, a user profile is then extended by computing the relationship between places and concepts derived from collective behaviour of other users in the dataset.

A basic profile represents actual interactions with places, while the extended profile describe "recommended" associations given overall interactions between users, places and concepts in the dataset. It is possible to model such interactions separately in the extended profile by controlling the similarity function used to create the profile. For example one can focus on modelling the types of places visited by the user or take into account visit behaviour of other users whose profiles overlap with the user, as discussed below.

The decay approach is intended as a dynamic view of user profiles that will capture the change of interests in places over time. In particular, the degree of associations between users, places and concepts are attenuated by a factor proportional to their age in the folksonomy graph. We use a pre-filtering approach to the application of time function, where the decay factor is applied to the folksonomy graph before the profiles are computed.

The Decay is an exponential function defined as the way in which a quantity naturally decreases over time at a rate proportional to its current value [33]. Tags and posts are considered to have a lifetime. When they are first defined, they are fresh and interesting to users then they decay over time and their interest value decreases. [118]. Using the decay model for modelling the dynamic user and place profiles is as it is important to maintain the freshness of the tags and to monitor the change of the tag behaviour over time. Different time decay functions have been used [69] in the literature. Decay functions can be represented as an exponential [29], power [113], Linear [72], logistic function [29] and as a Base Level Learning (BBL) function [66] In this work, an exponential decay function is used and different values of the decay factor is evaluated using the location recommendation evaluation method. Exponential function is suitable for modelling interests in places visited over time. Users' historical interests may influence his future interests, and more recent interests may have stronger impact on the future preference prediction than earlier interests. The following equation shows how to calculate the weights after applying the exponential decay equation.

$$w(t) = w_{t_0} \cdot e^{-\lambda \Delta t} \tag{6.1}$$

where, $w(t)$ is the weight at time slot t, w_{t_0} is the weight at time slot $t_0 < t$, λ is the decay rate and Δt is the number of time slots between t and t_0.

In what follows, the different types of user profiles are presented, both the time-slice approach and, the decay approach.

Definition 16. ***Basic Spatiotemporal User Profile*** *A spatiotemporal (ST) user profile $P_{R_{t_c}}(u)$ of a user u is deduced from the set of places that u visited or annotated directly.*

Time-slice:

$$
\begin{aligned}
(P_R(u))_{t_c} =& \{(r, w(u,r)_{t_c}) | (u,g,r) \in Y, \\
& w(u,r)_{t_c} = |\{g_{t_c} \in G : (u,g,r) \in Y\}|
\end{aligned}
$$

$w(u,r)_{t_c}$ is the number of tag assignments in the time slot t_c.

Decay:

$$
\begin{aligned}
(P_R(u))_{t_c} =& \{(r, w(u,r)_{t_c}) | (u,g,r) \in Y, \\
w(u,r)_{t_c} =& \sum_{t=t_i}^{t_c} |\{g_{t_i} \in G : (u,g,r) \in Y\}| \cdot df(t_c - t_i)\}
\end{aligned}
$$

$w(u,r)_{t_c}$ is the number of tag assignments accumulated over the timeline of the user data to time slot t_c, attenuated by the decay factor df defined relative to t_c.

Here, tag assignments are defined when user u associates some tag g to place r through the action of checking-in or annotation. Hence, the weight assigned to a place at a specific time point simply corresponds to the frequency of user's tag assignments, decayed over time.

The the weights are further normalised so that the sum of the weights assigned to the places in the spatial profile is equal to 1. $\bar{P}_R$ is used to explicitly refer to the spatial profile where the sum of all weights is equal to 1, with $\bar{w}(u,r)_{t_c} = \frac{|\{g \in G : (u,g,r) \in Y\}|}{\sum_{i=1}^{n} \sum_{j=1}^{m} |\{t_i \in T : (u, g_i, r_j) \in Y\}|}$, where n and m are the total number of tags and resources, respectively. More simply, $\bar{w}(u,r)_{t_c} = \frac{N(u,r)}{N_G(u)}$, where $N(u,r)$ is the number of tags used by u for resource r, while $N_G(u)$ is the total number of tags used by u for all places.

Correspondingly, the semantic profile of a user; $P_T(u)$, is defined as follows.

Definition 17. ***Semantic-Temporal User Profile*** *A semantic temporal (SemT) user profile $P_{G_{t_c}}(u)$ of a user u is deduced from the set of tag assignments linked with u at time point t_c.*

Time-slice:

$$P_G(u)_{t_c} = \{(g, w(u,g)_{t_c)} | (u,g,r) \in Y,$$
$$w(u,g)_{t_c} = |\{r_{t_c} \in R : (u,g,r) \in Y\}|$$

$w(u,g)$ is the number of tag assignments at t_c.

Decay:

$$P_G(u)_{t_c} = \{(g, w(u,g)_{t_c)} | (u,g,r) \in Y,$$
$$w(u,g)_{t_c} = \sum_{t=t_i}^{t_c} |\{r_{t_i} \in R : (u,g,r) \in Y\}| \cdot df(t_c - t_i)\}$$

$w(u,g)_{t_c}$ is the number of tag assignments accumulated over the timeline of the user data to time slot t_c, attenuated by the decay factor defined relative to t_c.

Here, tag assignments are defined when user u assigns tag g to some place r through the action of checking-in or annotation. $\bar{P}_{G_{t_c}}$ refers to the semantic profile where the sum of all weights is equal to 1, with $\bar{w}(u,g)_{t_c} = \frac{N(u,g)}{N_R(u)}$, where $N(u,g)$ is the number of resources annotated by u with g and $N_R(u)$ is the total number of resources annotated by u.

Furthermore, a spatio-semantic temporal profile of a user $P_{RG_{t_c}}(u)$, that describes how a user is associated to both place-tag combined elements, is defined.

Definition 18. ***Spatio-Semantic Temporal User Profile*** *Let $\mathbb{F}_u = (G_u, R_u, I_u)$ of a given user $u \in U$ be the restriction of $\mathbb{F}$ to u, such that, G_u and R_u are finite sets of tags and places respectively, that are referenced from tag assignments performed by u, and I_u defines a relation between these sets: $I_u := \{(g,r) \in G_u \times R_u | (u,g,r) \in Y\}$.*

A spatio-semantic temporal user profile $P_{RGt_c}(u)$ of a user u is deduced from the set of tag assignments made for place r by u, defined as follows. Time-slice:

$$P_{RGt_c}(u) = \{([r,g], w_u([r,g])) | (g,r) \in I_u,$$
$$w_u([r,g])_{t_c} = |\{g_{t_i} \in G_u : (g,r) \in I_u\}|\}$$

where $w([r,g])$ is how often user u assigned the specific tag g to place r.

Decay:

$$P_{RGt_c}(u) = \{([r,g], w_u([r,g])) | (g,r) \in I_u,$$
$$w_u([r,g])_{t_c} = \sum_{t=t_i}^{t_c} |\{g_{t_i} \in G_u : (g,r) \in I_u\}| \cdot df(t_c - t_i)\}$$

where $w([r,g])$ is how often user u assigned the specific tag g to place r accumulated over the timeline of the user data to time slot t_c, attenuated by the decay factor df defined relative to t_c.

$\bar{P_{RG}}$ is the spatio-semantic profile where the sum of all weights is equal to 1, with $w_u([r,g]) = \frac{N(u,[r,g])}{N_{rG}(u)}$, where $N(u,[r,g])$ is the number of times u annotate r with g, and $N_{rG}(u)$ is the total number of tags assigned by u for r. (Note that tag assignment by users for a place comes from both the explicit action of annotation as well as implicit action of checking-in as represented in the geo-folksonomy model).

6.3.1 Enriched User Profiles

The basic user profiles are expanded by the information extracted from the computation of tag and place similarity defined in section 4.3.2. The enriched user profiles will therefore present a modified view of how users are associated with places that reflect collective user behaviour on the LBSN.

Definition 19. ***Enriched Spatiotemporal User Profile***

An enriched ST user profile $\acute{P}_{R_{t_c}}(u)$ of a user u is an extension of the basic profile by places with the highest degree of similarity to places in $\bar{P_{R_{t_c}}}(u)$. Let R_u be the set of all places in $\bar{P_R}(u)$ and w_i is the weight associated with place i in the profile.

$$\acute{P}_{R_{t_c}}(u) = \{< r_i, w_i > |$$
$$w_i = \left\{ \begin{array}{ll} w_i & , if\, r_i \in R_u \\ w_i * Max(Sim(r_i, r_j)) & , \forall (r_i \in \{R - R_u\} \wedge r_j \in R_u) \end{array} \right\}$$

The maximum similarity of the N most similar places in the dataset is calculated for every place in the basic user profile, and the highest similarity score is used as the weight for the new place in the enriched user profile. The process of building the enriched spatial profile is shown in the following procedure 6.1.

Definition 20. ***Enriched Semantic-Temporal User Profile*** *An enriched Semantic-Temporal user profile $\acute{P}_{T_{t_c}}(u)$ of a user u is an extension of the basic profile by tags with the highest degree of similarity to tags in $\bar{P_{T_{t_c}}}(u)$. Let T_u be the set of all tags in $\bar{P_{T_{t_c}}}(u)$ and w_i is the weight associated with tag i in the profile.*

$$\acute{P}_{T_{t_c}}(u) = \{< t_i, w_i > | w_i = \left\{ \begin{array}{ll} w_i & , if t_i \in T_u \\ w_i * Max(Sim(t_i, t_j)) & , \forall (t_i \in \{T - T_u\} \wedge T_j \in T_u) \end{array} \right\}$$

A similar algorithm to that of enriching place profiles is used for choosing the tags and weights.

Definition 21. ***Enriched Spatio-Semantic Temporal User Profile***

An enriched spatio-semantic user profile $\acute{P}_{RT}(u)$ of a user u is an extension of the basic profile by tags and places with the highest degree of similarity to tags in $P_{RT}(u)$. Let T_u be the set of all tags in $\bar{P_T}(u)$, R_u be the set of all places in $\bar{P_R}(u)$ and w_{ij} is the weight associated with tag i and place j in the profile.

$$\acute{P}_{RT}(u) =< [r_i, t_j], w_u(r_i, t_j) > | w_u(r_i, t_j) = \left\{ \begin{array}{ll} w_u(r_i, t_j) & , if\, r_i \in R_u and t_j \in T_u \\ w_u(r_i, t_j) * Max(Sim(r_i, r_k)) & , t_j \in P_T(r_k) \wedge r_k \in \{R - R_u\} \\ 0 & otherwise \end{array} \right\}$$

Algorithm 6.1: TemporalSpatialEnrichment(γ,K,t_i,t_j)

Input: γ,K,t_i,t_j

Output: $\acute{P}_R(u)$

Fetch Spatial Profile $P_R(u)$ between t_i,t_j

for all places r_i in Spatial-Profile $P_R(u)$ **do**

  if $\gamma = 1$ **then**

    Compute $CosSim_{tag}(r_1, r_2)$

  else

    if $\gamma = 0$ **then**

      Compute $CosSim_{User}(r_1, r_2)$

    else

      Compute $Sim(r_1, r_2, \gamma)$

    end if

  end if

  Find top K similar places r_j to each r_i in $P_R(u)$

  for each $< r_j, sim >$ in top similar places **do**

    $w_j = w_i * sim$

    add $< r_j, w_j >$ to $P_R(u)$

  end for

end for

return $\acute{P}_R(u)$

The spatio-semantic profile is extended with the most similar places to the user profile and these are assigned a weight computed using the place similarity value for all tags in their place-tag profiles and 0 *for tags that are not in their profile. Thus the user simply inherits relationships with all the tags and their associated weights from the places that are deemed similar to those in their profile.*

6.4 Experimental Analysis

In this section, our experiments are intended to address the parameters that affects the proposed recommendation approach. Also, it shows a comparison between different time approaches proposed and the basic recommendation methods.

6.4.1 Data Set description

The data set used was described in the previous chapters. The data-set contains two subsets, the check-in users and the tipping users. The time-stamps of the check-ins are distributed throughout 2012 and 2013, while the time-stamps for the tipping users begins in year 2009 and ends in 2015. Figure 6.2 shows the temporal distribution of users across the years. Figure 6.3 shows the distribution of the distinct number of users, places, tags, and the folksonomy relationships in each month during 2012 and, 2013. Figure 6.4 shows the average distribution of users, places, and tags in each month during 2012 and 2013. As the figure shows, the number of users between Feb,2012 and Mar,2013 is larger than the rest of the months. This is because the check-in activity of all users lies in this period of time.

6.4.2 Parameter Setting

To evaluate user profiles, a recommendation method that recommends places or tags based on the similarity between profiles is used. In this subsection, the results of tuning the recommendation parameters are presented. The first parameter is called the given value, which represents the number of places or tags used for the training. So, for example, for the semantic user profile, if a user profile contains 100 tags, not all of them are used for training, only a “given value” is used. The given value value is percentage of the tags used to train the semantic user profile. To set this parameter, the recommendation method is trained with different values to see which value will give

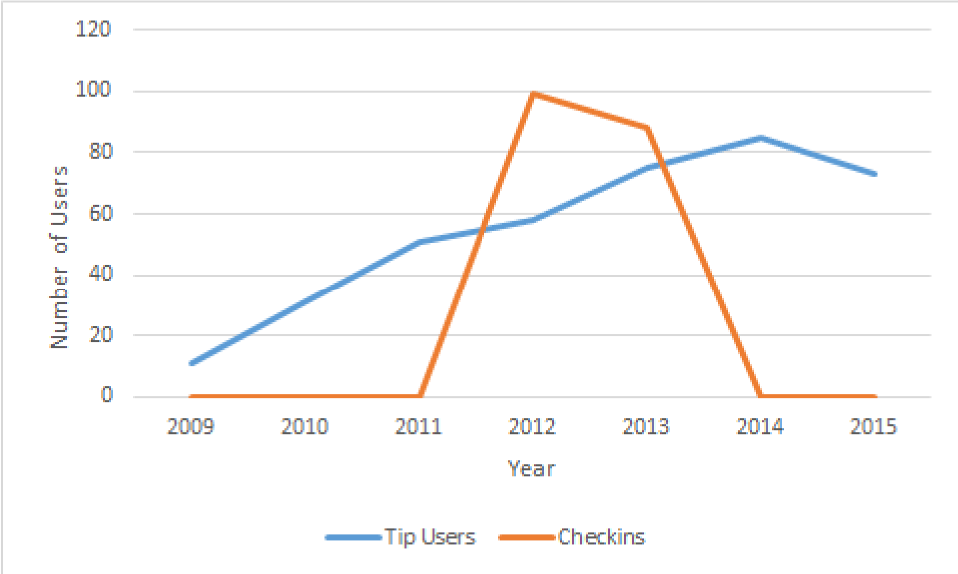


Figure 6.2: Temporal Distribution of Number of Users

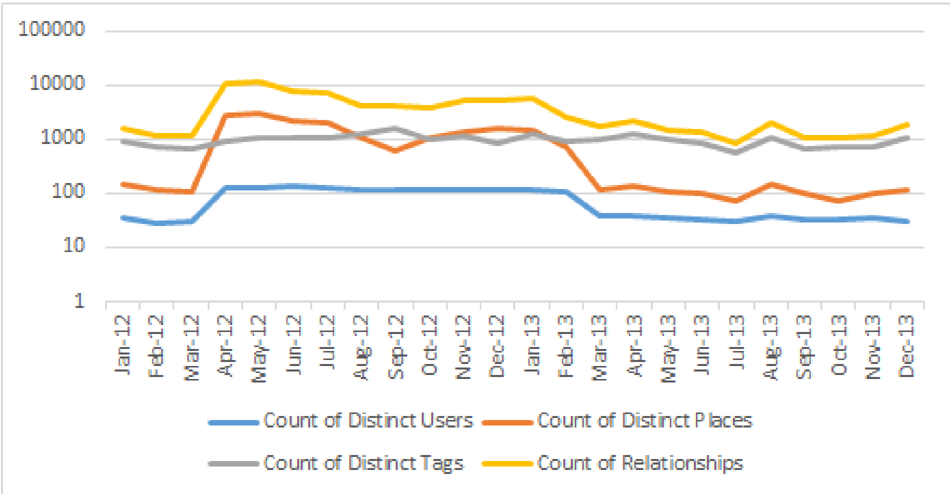


Figure 6.3: Temporal Dataset Statistics

the best results. Figure 6.5, 6.6, and 6.7 show the precision, recall, and F1-measure values when using 1 % 10 %,25 %,50 %, 75 % and 100 % of the tags in the training set for the semantic user profile evaluation. The percision values shows that we increase the number of tags in the training set, the precision values decreases, while the recall values show the increase in the number of tags is directly proportional to the increase of recall values. So, what is required is trad-off between the two values. Figure 6.7

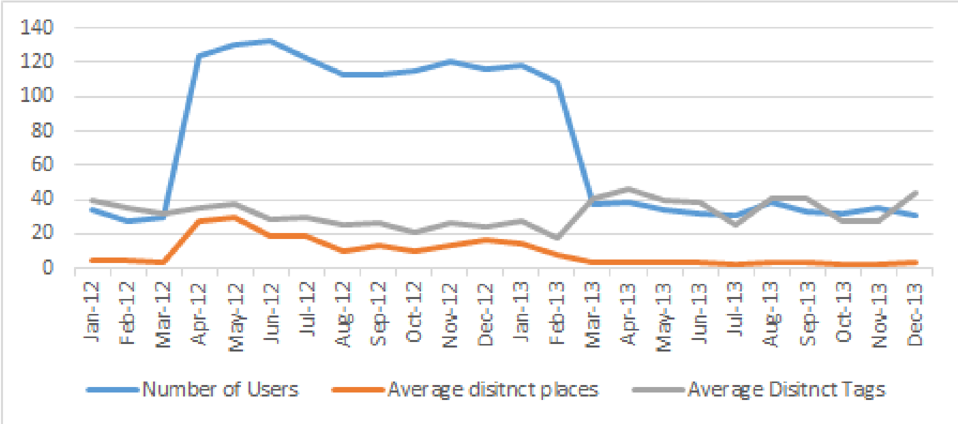


Figure 6.4: Temporal Average Dataset Statistics

shows that using 50% of the tags in the training set gives the best overall F1-measure value. Thus, the given value is set at 50% of the number of tags or places in the training set.

Another parameter that is used for training the recommender is the splitting percent, which randomly assigns a predefined proportion of the users to the training set and all others to the test set. In this experiment, values for the splitting percent ranged from 10% to 90% is experimented. Figures 6.8, 6.9, and 6.10 show the precision, recall, and F1-measure values for different training settings. A radar chart is used to plot the precision and recall values for different top-N recommendations as shown in figures 6.8 and, 6.9. Radar charts are a useful way to display multivariate observations with an arbitrary number of variables. Each star represents an observation of the splitting percent, the bigger the star the higher the values of precision and recall. As the figures show, when using 90% the data-set for training, values of the precision, and recall are the best. Figure 6.10 shows the values of the F1-measure for each splitting percentage with a noted high performance for the 90% setting.

Another important parameter that is used in the recommendation method is the decay factor λ. The decay factor is the value that controls the freshness of the tags. In constructing a dynamic user profile, tags should be ranked according to freshness and frequency. Old tags with no repetitions should decay over time, while recent tags

Algorithm 6.2: Temporal Decay Spatio-semantic Top-K Recommender(γ,K,t_{now})

Input: The combination factor γ,

Number of places to recommend K,

The current timet_{now},

Output: Top-K recommended places$< r_i, w_i >$

for each u_i **do**

Fetch t_0 for u_i

SpatialEnrichment(γ,K,t_0,t_{now})

end for

for all u_i, u_j **do**

Fetch profiles $P_R(u_i), P_R(u_j)$

Compute User-Sim(u_i, u_j) .

end for

for each U_i **do**

Fetch most similar user u_j

Sort $< r_i, w_i >$ of $P_R(u_j)$

Recommend top K r_i that are not in $P_R(u_i)$

end for

return Top-K$< r_i, w_i >$

should have strong weight. In addition, if a recent tag has a small frequency, its weight should be high. Similarly, old tags with high frequency should have a high weight as well. In other words, all the data contribute to the location recommendation, while the most recent data contributes the most. The old data reflects users' previous preferences. It should have small weights in the prediction of recommendation if it was not repeated several times. In the proposed algorithm, an exponential form for the time function is selected to achieve the goal. The exponential time function is widely used in many applications in which it is desirable to gradually decay the history of past behaviour as time goes by. In this experiment the decay value is set to different values depending on

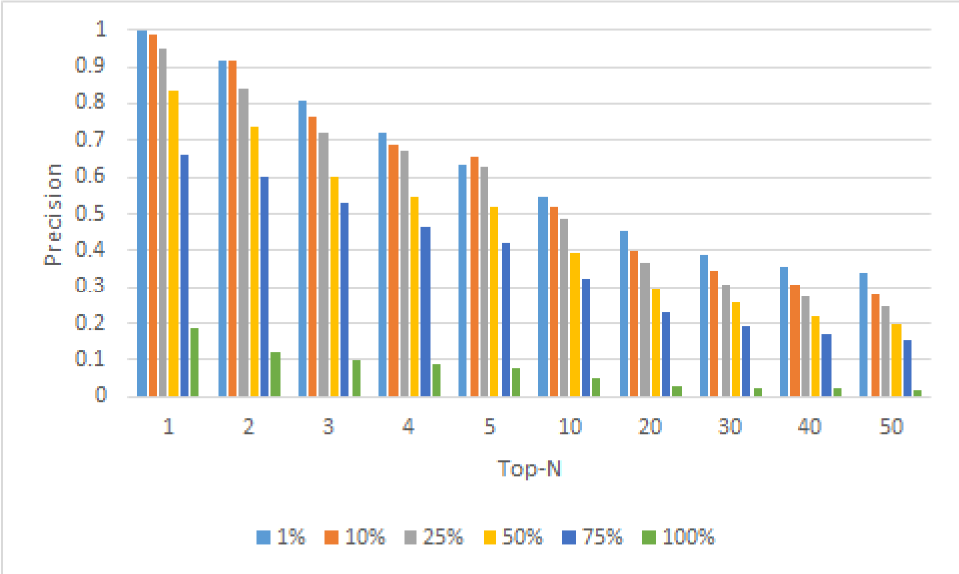


Figure 6.5: Evaluation of the Given Parameter-Precision

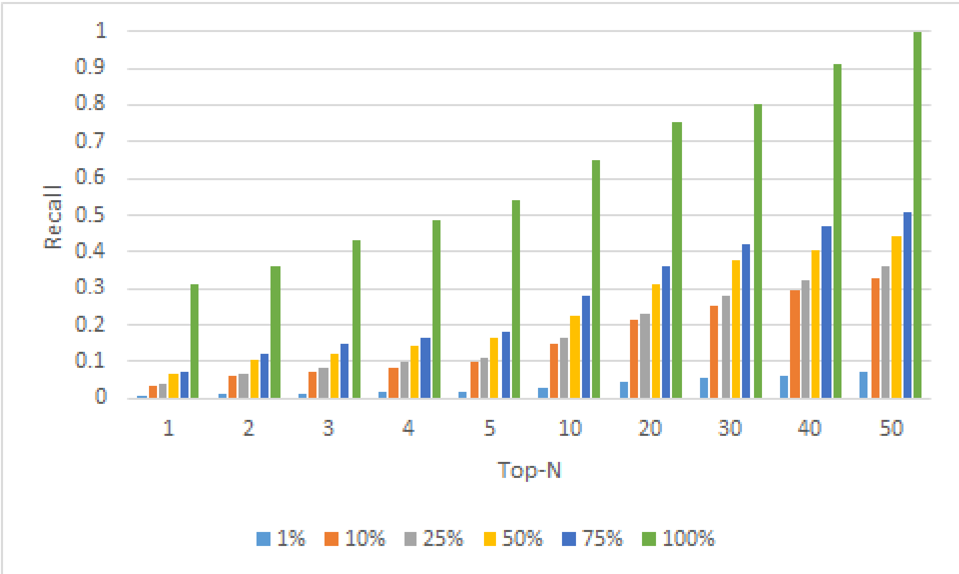


Figure 6.6: Evaluation of the Given Parameter-Recall

the half-life parameter (T_0) which is the time that elapses until the weight of the tag is reduced to half of its amount. In the dataset, the time range is 24 month. The decay value is defined by the following equation.

$$\lambda = \frac{1}{T_0} \tag{6.2}$$

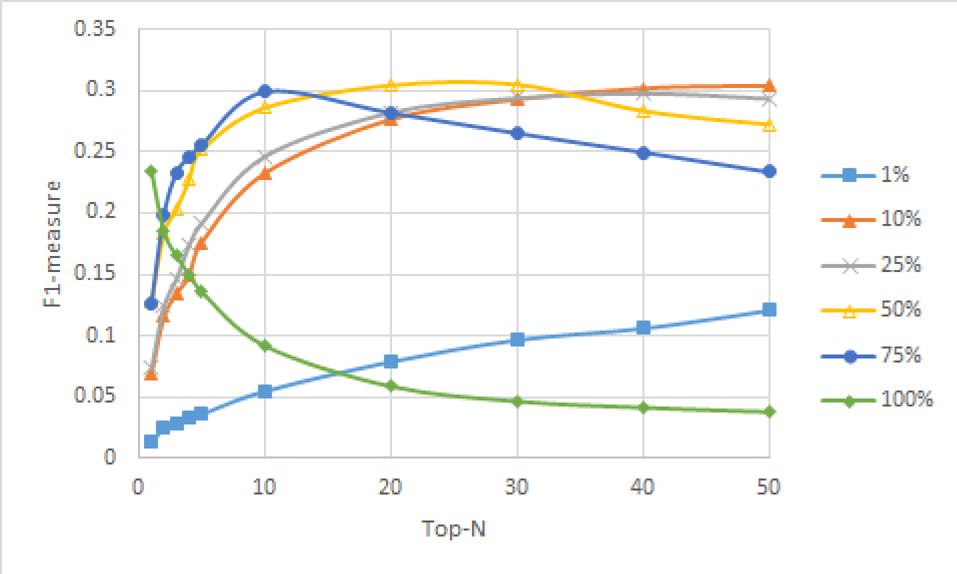


Figure 6.7: Evaluation of the Given Parameter-F1-Measure

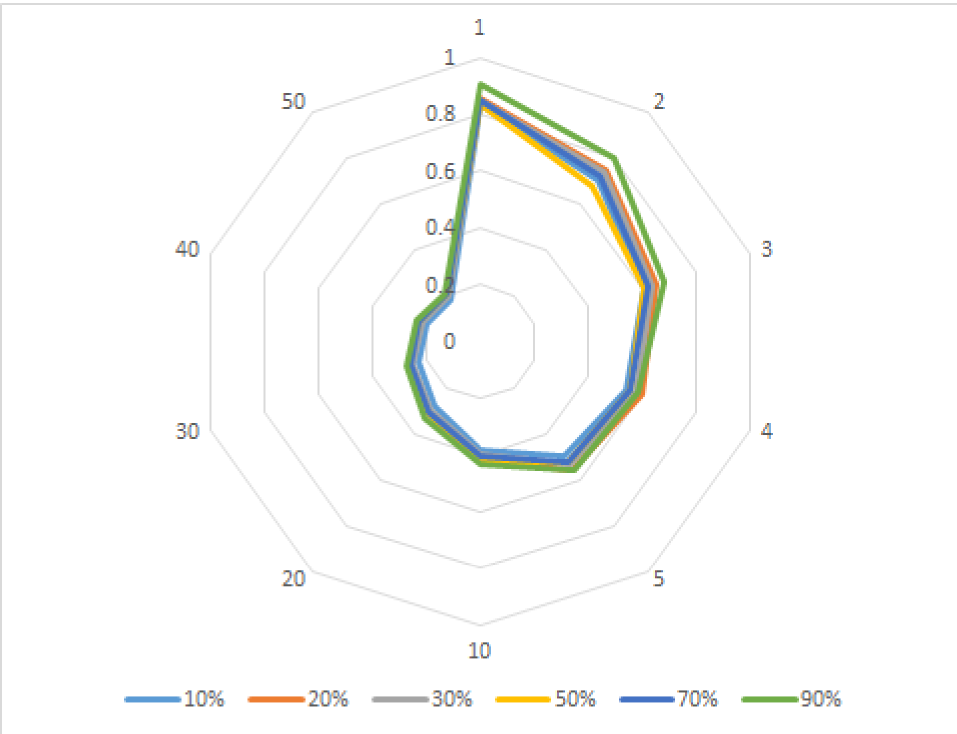


Figure 6.8: Evaluation of the Splitting Parameter-Precision

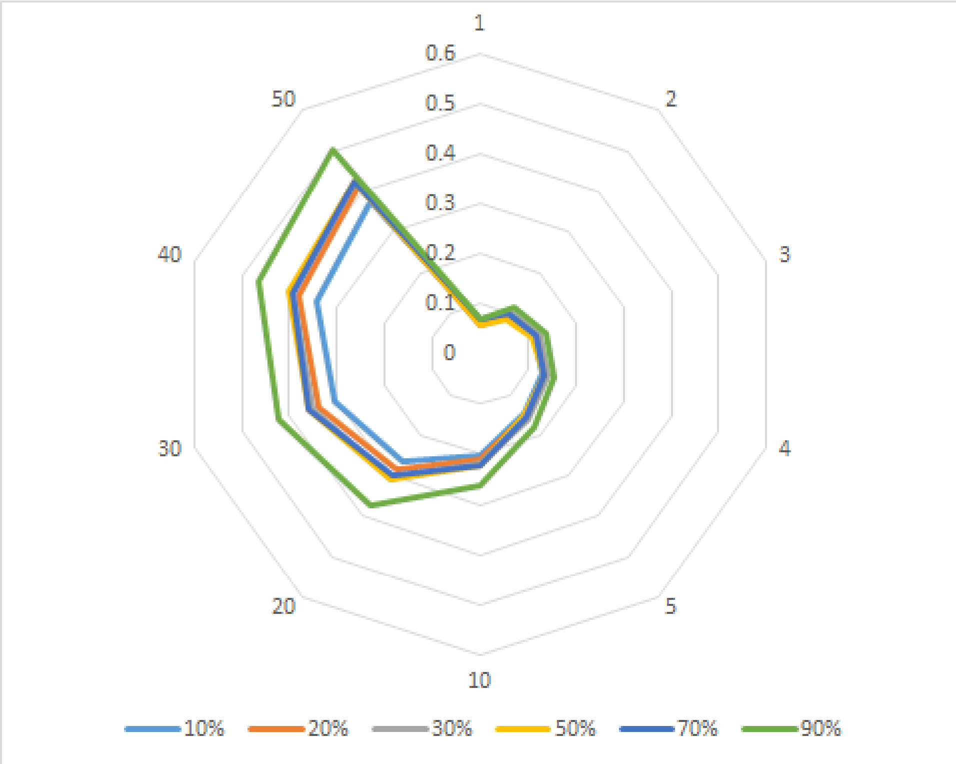


Figure 6.9: Evaluation of the Splitting Parameter-Recall

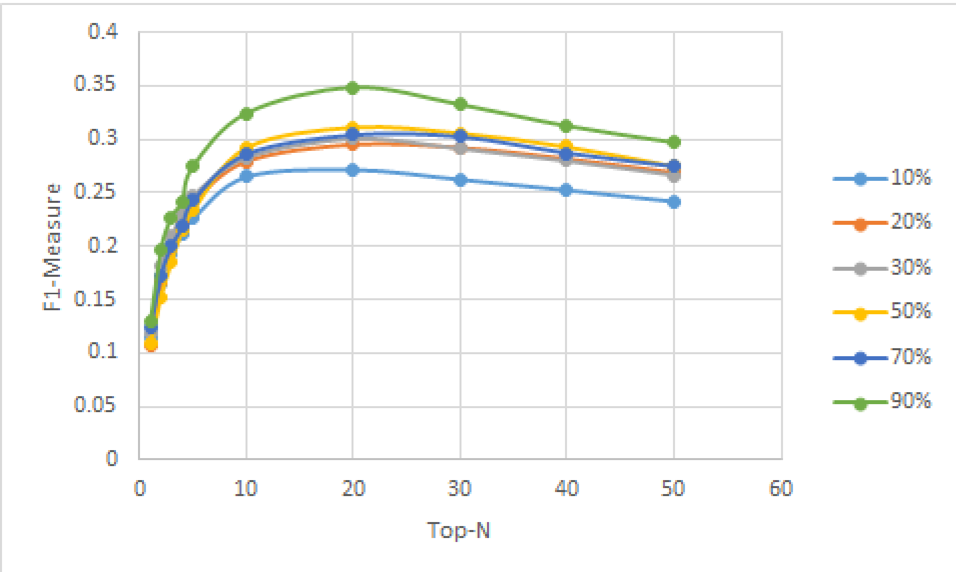


Figure 6.10: Evaluation of the Splitting Parameter-F1-Measure

Table 6.1: Half-Life and Decay Values

Half-life T_0***(in months)***	***Decay Value*** (λ)
1	1
2	0.5
3	0.333333
4	0.25
5	0.2
6	0.166667
7	0.142857
8	0.125
9	0.111111
10	0.1
11	0.090909
12	0.083333

Table 6.1 shows the decay values when using different half-life periods.

Conceptually, the aim of designing a half-life parameter is to define the rate of decay of the weight assigned to each data point. The time function $f(\Delta t) = \cdot e^{-\lambda \Delta t}$ is an exponential function that ranges from 0 to 1.The more recent the data, the higher the value of the time function. This exponential function is suitable for our problem. Figure 6.11 shows different curves of time function when using different values of the half-life parameter T_0. The values of the half are 3 month, 6 month, and 12 month. In the presented problem, the decay rate of old data is decided by how frequently the user interests in places changes. In order to find the appropriate value of parameter T_0 to precisely predict the user's future preference, the recommendation method was trained using different settings of T_0. It is necessary to find the appropriate value of parameter T0 to precisely predict the user's future preference, thereby improving the performance of our proposed algorithm. Figure 6.12 shows the F1-Measure when using different values of T_0. The figures shows that the best value for the decay factor is

'0.33' which corresponds to three month half-life. This means that after six month the tag that was never repeated will vanish from the dataset.

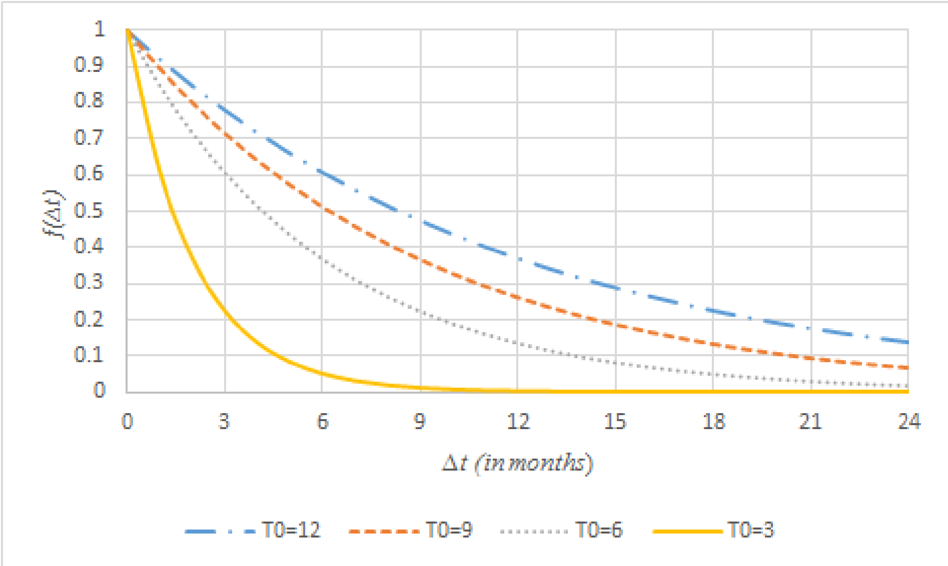


Figure 6.11: Training of Spatial Profile Using Different Half-life Values.

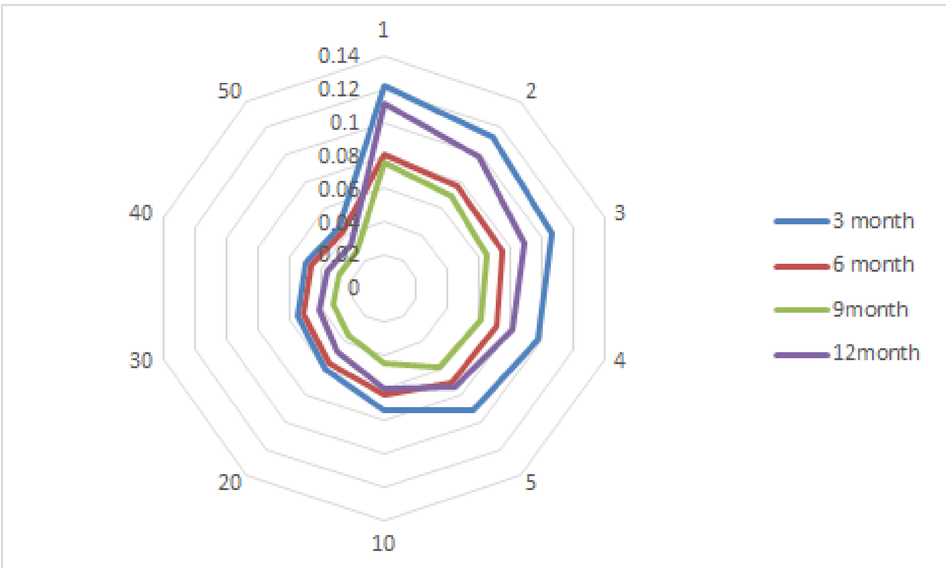


Figure 6.12: F1-Measure of Training Different Values of T_0

6.5 Temporal Decay User-Profile Evaluation

This section will evaluate the decay method proposed with the basic methods of user-based and the item-based collaborative filtering. The time-slices method used in the literature is also compared, with the proposed.

6.5.1 Spatial Profiles Evaluation

Results for the enriched temporal spatial user profiles using the proposed top-N recommendation method are presented. Different versions of the enriched temporal spatial profiles, using different place similarity measures were created, a) using $\gamma = 0$ (to represent place-tag similarity only), b) using $\gamma = 1$, (to represent place-user similarity only), and c) using $\gamma = 0.5$ for an aggregated view of both effects. Hence, the result sets are shown for the following user profiles. 1. Enriched-Spatial(Tag) 2. Enriched-Spatial(User) 3. Enriched-Spatial(Combined).

The results of the top-N recommendation is compared using the three different profiles with traditional Item-Based Collaborative Filtering (IBCF) [100] and the User-Based collaborative Filtering (UCBF) [89] approaches, applied against the basic spatial user profile. The results of the precision, recall and F1 measures for recommending the top-1, 2, 3, 4, 5, 10, 20, 30, 40, 50 places are shown in Figures 6.13, 6.14 and 6.15, respectively.

6.5.2 Semantic Profile Evaluation

A similar experiment was carried out to evaluate the semantic user profiles. Again, the results were compared to the UBCF and IBCF approaches. Figures 4.9, 4.10 and 4.11 show the results of the top-10, 20, 30, 40, and 50 tag recommendations using the different methods. The results demonstrates the quality of the enriched user profiles, and thus confirm their utility for more accurate representations of user profiles.

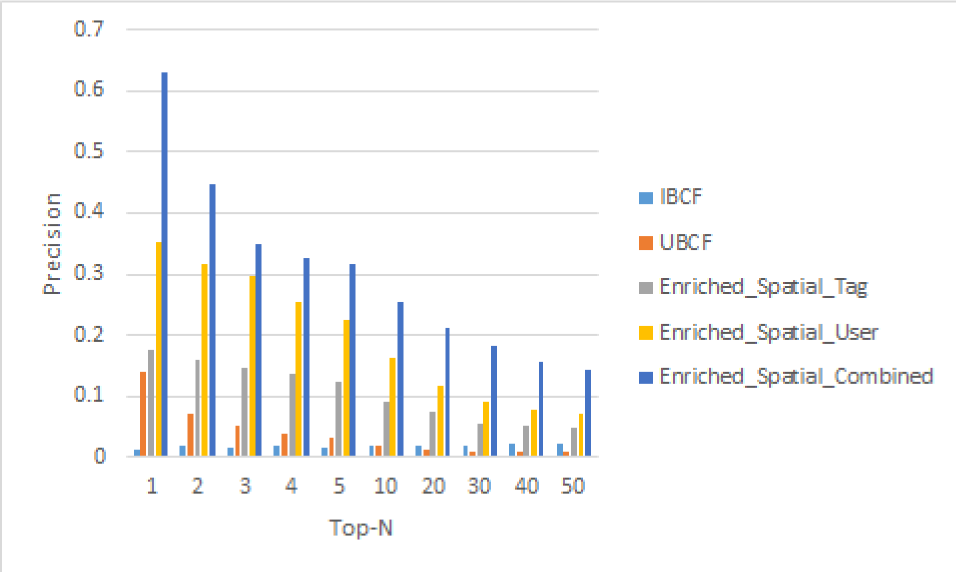


Figure 6.13: Evaluation of the decay Method-Precision

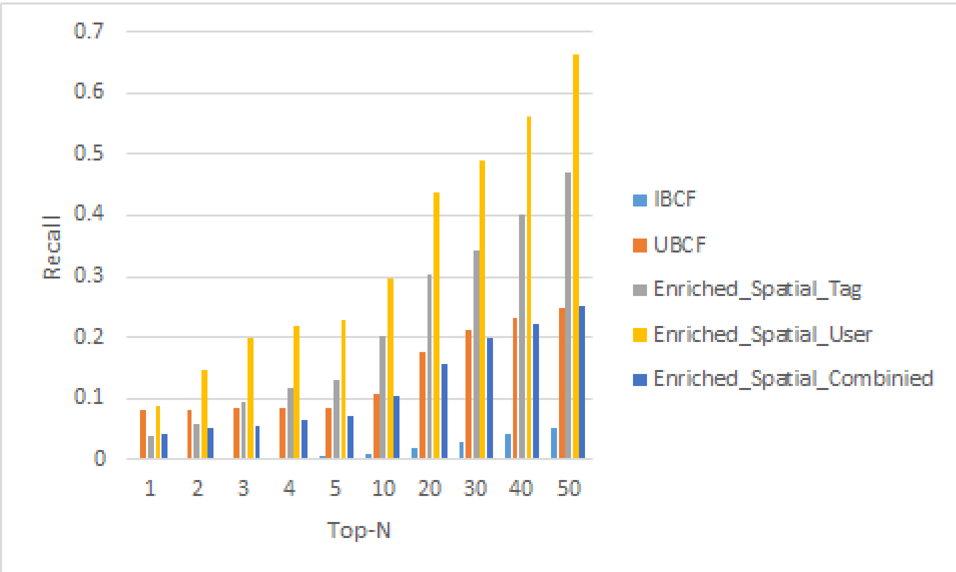


Figure 6.14: Evaluation of the Decay Method-Recall

6.5.3 Decay Method versus Time-Slices evaluation

This subsection will compare the proposed decay method to the time-slices method. The first evaluation metric utilised to compare the methods is the MAP, which is the mean of precision values for all top-N recommendations. The MAP was previously

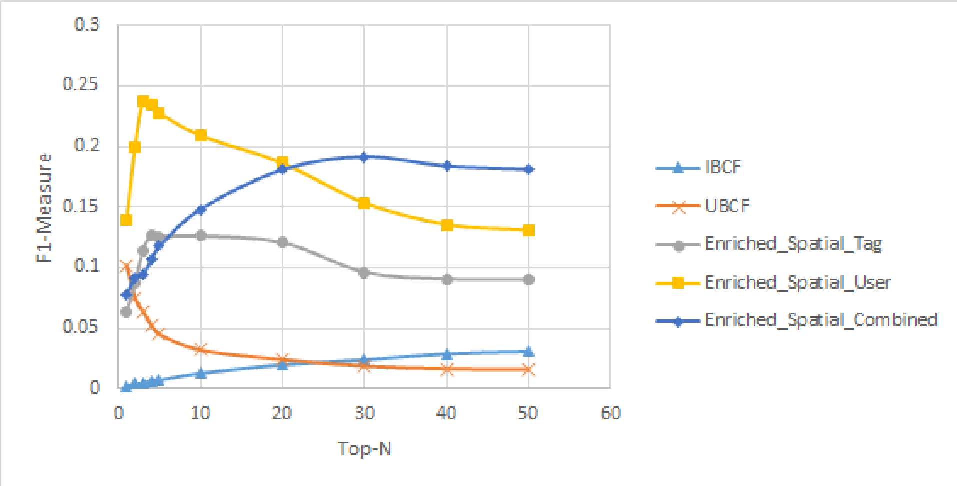


Figure 6.15: Evaluation of the Decay Method-F1-Measure

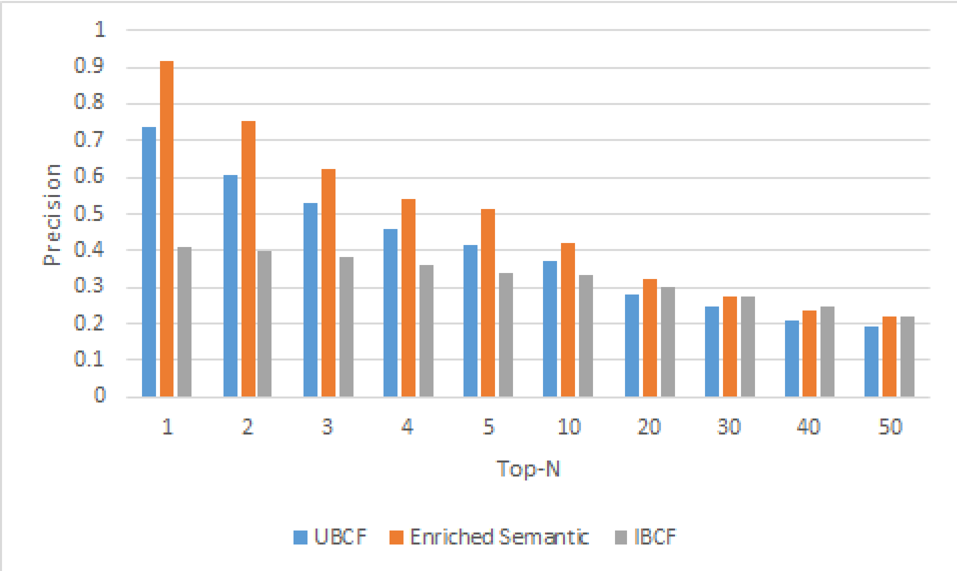


Figure 6.16: Evaluation Semantic Profiles-Precision

introduced in chapter 4 in Equation 7.4. A typical way to evaluate a prediction is to compute the deviation of the prediction from the true value. This is the basis for the Mean Average Error (MAE) described in Equation 6.3.

$$\frac{1}{|K|}\sum_{i,j\in K}|r_{i,j}-\hat{r_{i,j}}| = \frac{FP+TN}{FP+FN+TP+TN} \tag{6.3}$$

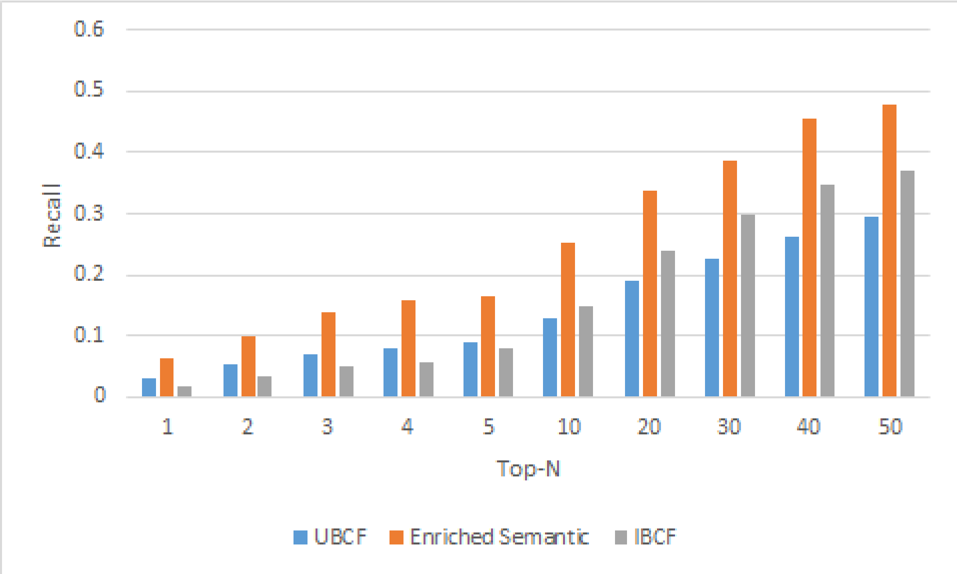


Figure 6.17: Evaluation of Semantic Profiles-Recall

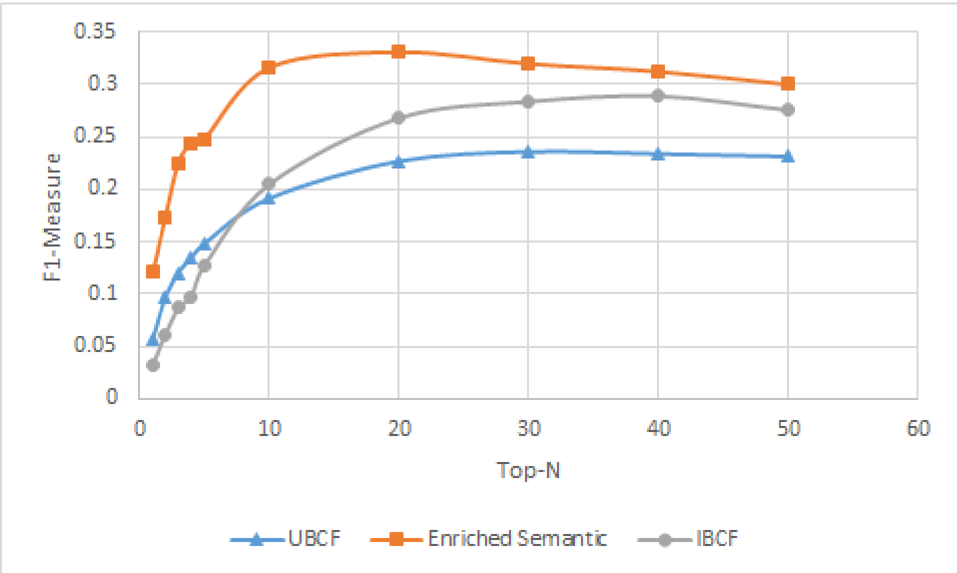


Figure 6.18: Evaluation of the Semantic Profiles-F1-Measure

where K is the set of all user-item pairings (i,j) for which we have a predicted rating $\hat{r_{i,j}}$ and a known rating $r_{i,j}$ which was not used to learn the recommendation model.

Figure 6.19 shows the comparison of different enrichment values using the two proposed method, timeslices and decay. As the figure shows, the decay method outper-

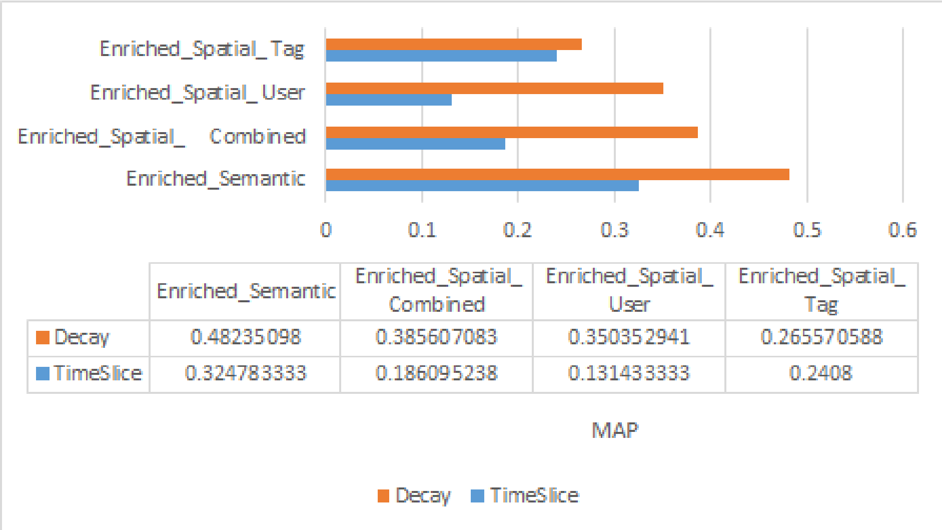


Figure 6.19: Comparing Time-Slice and Decay Methods-MAP

forms the timeslice method in all profiles with best one for the combined enriched spatial profile. The reason for better MAP is, increasing the length of the time-slice makes the data denser, leading to better precision. We also compared the two methods using the MAE metric, but using only the enriched combined profile in Figure 6.20. As figure shows, the decay method error are smaller than the time-slices, which means that rating prediction is better for the decay method. It is observed in the experiments that when increasing the length of time slots, more ground truth places is brought for each user at each time slice. With the number of recommendations (i.e., N) unchanged, poorer recall values are observed with increasing the length of time slot. Thus, the decay method shows poorer recall values because it adds up the profile along the time.

The two proposed methods, time-slice and decay, are used for top-N recommendation as illustrated previously. Each method has it characteristics. In time slice method, the recommended place are predicted based on the short term profile that appears only within the time-slice, while in the decay method, the recommended places are predicted based on the whole historic profile of the user. Figure 6.21 illustrates an example of recommendation when using both methods. The figure shows the profile of a user during three months, January February and March. When we use both meth-

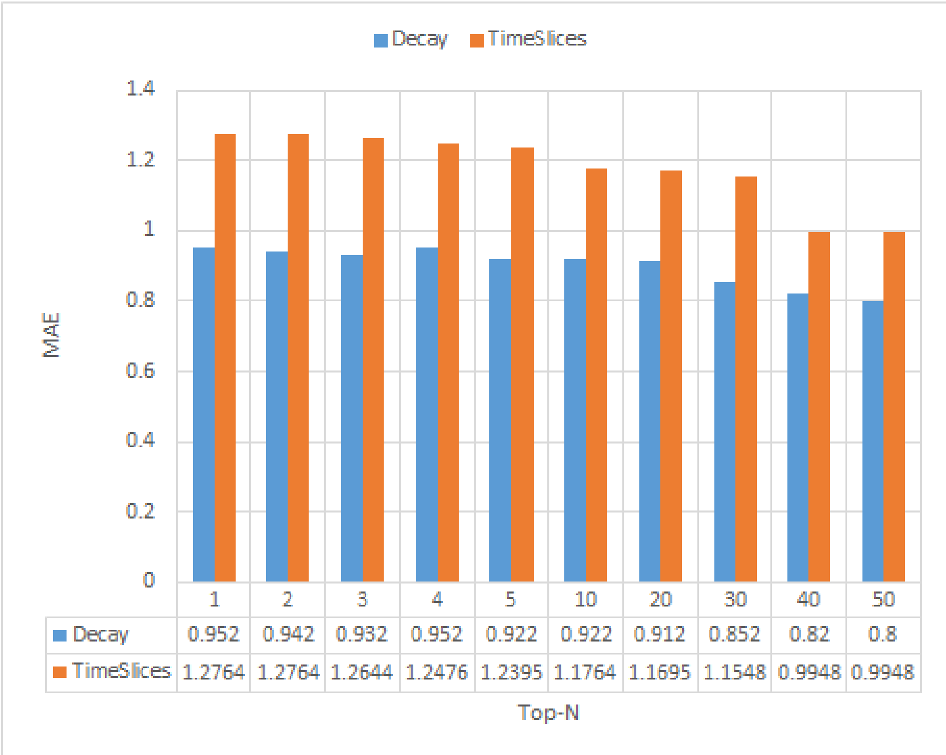


Figure 6.20: Comparing Time-Slice and Decay Methods-MAE

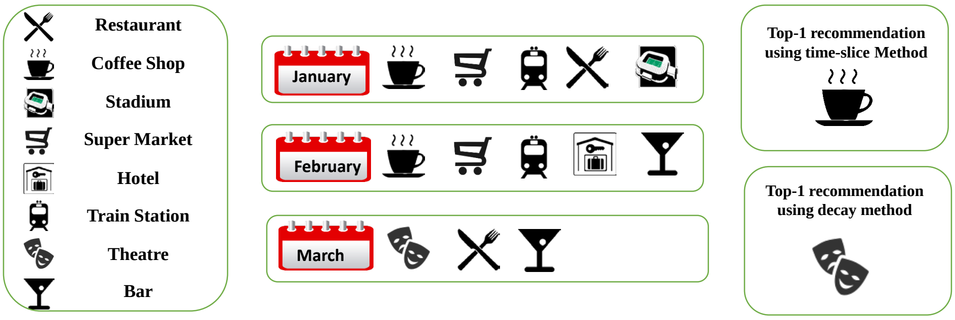


Figure 6.21: A Top-1 Recommendation Example using Time-Slice and Decay Methods.

ods for recommending places to this specific user, the time-slice method recommends the “theatre” because the recommender looks only at the user profile created during March. On the other hand, the decay method recommends the “café”’ because the recommender look at the user profile during the three months together. Thus,the time-slice method is considered as short-term recommendation, while the decay method is considered as a long-term recommendation method. The long-term place recommendation is more realistic than the short term recommendation method because it captures routine activities and interests of the user. The short term recommendation method can be some-how misleading because it does not reflect the actual activities and interests of the user towards the places he visits.

6.6 Summary

This chapter began by describing the temporal geo-folksnomy model and arguing that it is important to study the temporal influence in user modelling. It went on to suggest two methods for representing the time influence. The first method extracts user preferences during different time slots. Then, it calculates user and place similarities in each time interval based on co-occurrences. In the second method, an exponential decay function is used to measure interest drifts. Then, the user's location preferences are predicted using both methods at each time interval. The results of the location recommendation for both methods were discussed. The experimental results show that the proposed decay methods beat all baselines, and improve the accuracy of location recommendation over the time-slice method.The next chapter describes anther method for evaluating the user profiles using user similarity evaluation, it also empathises the conclusions retrieved in this chapter.

Chapter 7

Evaluation of User Similarity

7.1 Introduction

With the growth of location-based social networks, there is a need to calculate the similarities between their users. User similarity has a substantial impact on users, communities, and service providers. In LSBNs, different factors affect users similarity including co-existence in the same place, common interests between users, and temporal dynamics, which monitors the users' behaviour change over time. This Chapter studies and evaluate user similarity based on the users' interests and common POIs and aims to answer Research Question 5: How can different user profiles be evaluated using user similarity measures to assess their quality?. To answer this question, different user similarity measures are proposed and evaluated using a search process.

The rest of the chapter is organised as follows. Section 7.2 explains the different views of user similarity based on different criteria. In Section 7.3 introduces the evaluation method used to evaluate different user similarities. In Section 7.4, the experiment used to evaluate the approach is described and its results presented and discussed. The chapter concludes in Section 7.5.

7.2 Different Views of User Similarity

According to empirical analysis of the different views of user profiles proposed in chapters 3 and 5, the similarity between users A and B is measured using four different types of characteristics: similarity of interests, similarity of co-location , similarity of categories visited and similarity of activities in the place . Thus, there are four types of user similarity depending of the type of the profile used.

- **Semantic user similarity**: This is the cosine similarity between semantic user profiles. Semantic profiles is a conceptual measure of user interests. Similarity of semantic profiles can answer questions such as, "which other user share the same interests as I do?"
- **Spatial user similarity**: This is the cosine similarity between spatial user profiles based on the common places they visited. Similarity of spatial profiles can answer the question of "which other users' visiting habits are similar to mine?"
- **Categorical user similarity**: This is the cosine similarity between category-based user profiles based on the common categories that they visited. Similarity of categorical profiles can answer the question of "which other users visit place categories similar to mine?"
- **Activity user similarity**: This is the cosine similarity between activity-based user profiles based on the common activities that they talked about (or actually took part in) at the places visited. Similarity of Activity user profiles can answer the question of "which other users share the same sort of activities as I do?"

Table 7.1 shows the top most similar users from a sample of five users and a corresponding bar chart of their similarity values. As the figure shows, the most similar user changes according to the view of the profile.

Cosine similarity between semantic user profiles ($Semantic(u_1, u_2)$) and between spatial user profiles ($Spatial(u_1, u_2)$) construct a tag-oriented ranking and place-oriented

Table 7.1: The Top Most Similar Users from a Sample of Five Users

Users	Spatial	Categorical	Semantic	Activity	Similarity Values
1001	417	192	417	192	
1012	519	171	519	742	
1019	208	646	208	153	
1039	823	51	823	51	
1040	353	806	353	806	

ranking, respectively. Moreover, cosine similarity between category-based profiles ($Categorical(u_1, u_2)$) and between activity based profiles ($Activity(u_1, u_2)$) construct an activity based ranking and a category-oriented ranking respectively.

While the basic profiles will discover a map of common places, interests, categories or activities that the users visited or annotated, the enriched spatial profile will produce an extended map of places that are likely to be of interest to both users. Cosine similarity between user profiles can be used to find the recommended user similarity after the enrichment process. The procedure for calculating the recommended user similarity between two users using the spatial and semantic user profiles is shown in algoirthm 7.1.

User interests are not static; contrarily, their interests may change as time goes by. Although u1 and u2 might have common interests, at the same time, u1 may continue to pay attention to the same topic or place, while u2 might change their interest to a different topic or place. In this case, it will be inappropriate to give a constant user similarity measure throughout. Therefore, taking temporal information into consideration may improve the accuracy of user similarity predictions. To this end, a two temporal user similarity, long term user similarity, and short term user similarity for each type of profile is proposed. The short term user similarity is based on the time slices method discussed in Chapter 6. In the time slices method, the data is divided into short periods (weeks/ months) and the user similarity is determined based on interaction in this time slice only. In the long-term user similarity is based on the decay method discussed in Chapter 6. In the decay method, a higher weight is assigned to places visited recently than those that appeared a long time ago, since more recent preferences have greater influence on users' potential interests than earlier preferences.

The decay method adds interactions over time with a forgetting factor that maintains the interactions to be deleted. Thus, the user similarity based on the profile constructed by the decay method expresses all the past history of the user throughout a certain period of time.

Algorithm 7.1: AdaptiveUserSim(u_1, u_2)

1: Fetch Spatial Profiles $P_R(u_1), P_R(u_1)$
2: Compute UserSim($P_R(u_1), P_R(u_1)$)
3: $\acute{P}_R(u_1)$=SpatialEnrichment($P_R(u_1), \gamma$, K)
4: $\acute{P}_R(u_2)$=SpatialEnrichment($P_R(u_2), \gamma$, K)
5: Compute UserSim($\acute{P}_R(u_1), \acute{P}_R(u_2)$)
6: Fetch Spatial Profiles $P_T(u_1), P_T(u_1)$
7: $\acute{P}_T(u_1)$=SemanticEnrichment($P_T(u_1), K$)
8: $\acute{P}_T(u_2)$=SemanticEnrichment($P_T(u_2), K$)
9: Compute Recommended UserSim($\acute{P}_T(u_1), \acute{P}_T(u_2)$)

Figure 7.1 shows a bar chart of similarity values between 'user164' and other users, using their basic spatial and enriched spatial profiles. The figure demonstrates the impact of enrichment on user similarity, where this user appears to become more similar to other users in their profile, given an extended view of their interests in places and their associated concepts.

7.3 User Similarity Evaluation

The measure of user similarity is evaluated as an information retrieval problem where we search for the most similar user to a particular user in question. Each user is represented by his user profile, and the similar user is calculated by finding the cosine similarity between the user and all other users in the data-set, and then find the most

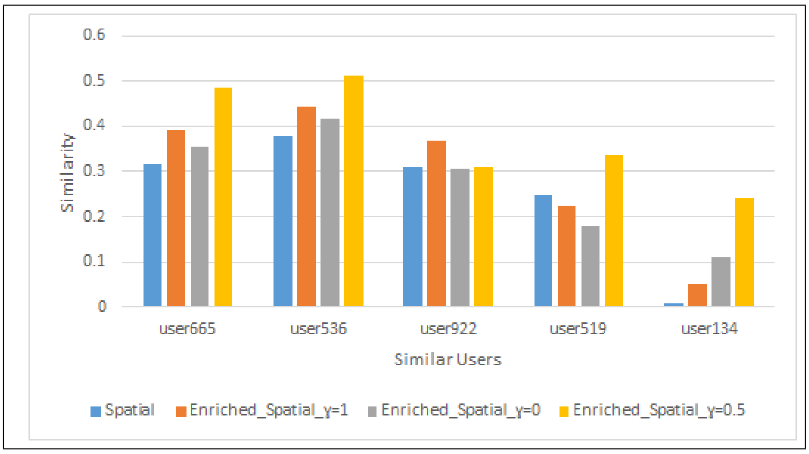


Figure 7.1: Similarity Between 'user164' and Other Users Using Spatial and Enriched Spatial Profiles.

similar user profile with the highest similarity. For instance, using user u_i as a query, the top similar user is retrieved based on their similarity score. To find a ground truth for evaluation, each user, and the most similar user, the place categories are retrieved because similar users tend to visit the same categories of places. Table 7.2 shows an example; where distinct categories for the top-10 most visited place are shown for two sample users (with similarity value of 0.65). Foursquare attaches more than one category to a place, and thus, there may be more than 10 categories for the top-10 places . The highlighted cells show the common categories between the two users.

We use the following evaluation metrics: Precision, Recall, and F-Measure, which are calculated in equations 7.1,7.2, and 7.3

$$Precision = \frac{|f(u) \bigcap f(u_{sim})|}{f(u)} \tag{7.1}$$

$$Recall = \frac{|f(u) \bigcap f(u_{sim})|}{f(u_{sim})} \tag{7.2}$$

where f(u) is the set of the distinct categories of the top-k places of user u, u_{sim} is the most similar user and f(u_{sim}) is the set of distinct categories of the top-k place of the most similar user u_{sim} F1-Score is a combination of recall and precision and is given

Table 7.2: Distinct Categories for Two Users with Similarity Value= 0.65

User 1	User 2
"American Restaurant"	"BBQ Joint"
"Coffee Shop"	"Bagel Shop"
"Shoe Store"	"Train Station"
"Pizza Place"	"Leather Goods Store"
"Office"	"Deli / Bodega"
"Train Station"	"Seafood Restaurant"
"Gym / Fitness Centre"	"Hotel"
"BBQ Joint"	"Clothing Store"
"Deli / Bodega"	"Residential Building"
"Donut Shop"	"Bakery"
"Metro Station"	"Park"
"Leather Goods Store"	"Shoe Store"
	"American Restaurant"
	"Meeting Room"
	"Office"

by

$$F1 - Score = \frac{2 * precision * recall}{precision + recall} \tag{7.3}$$

Hence, precision represents the ratio of common categories between the two users in reference to those of the first user, while recall presents the same ration with respect to the second user. The F1 measure is the harmonic mean of precision and recall. Then, the average of the precisions, recalls, and F-measures for all users is obtained. The precision represents the percentage of common categories between the two users to the total number of categories for the first user. The recall represents the percentage of common categories between the two users to the total number of categories of the second user. The f1 measure is the harmonic mean of precision and recall.

7.4 Experiments and Results

Experiments in this chapter were carried out using a two groups of user. The first group is users with high frequency of check-ins, co-location rate and tips that was previously described in Section 4.4, and the other is a low frequent user dataset that is described in Table 7.3. Figure 7.2 shows the number of places versus the number of users of the collected in the big dataset. As the figure shows, about 94% of the users visited less than 10 places and about 3% of users visited from 11 to 21 places and the remaining 3% visited from 20 to 400 places. The two sample datasets are subsets of the big dataset in the region of New york city. The application of the user similarity evaluation process is constrained by a geographic region of interest to improve the evaluation performance.

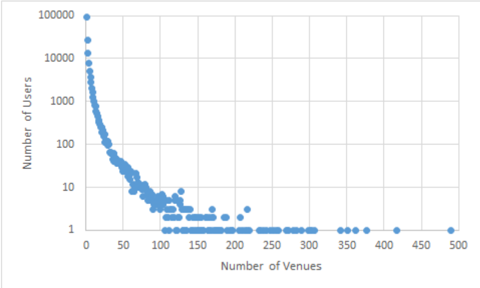


Figure 7.2: The Number of Users Versus the Number of Venues

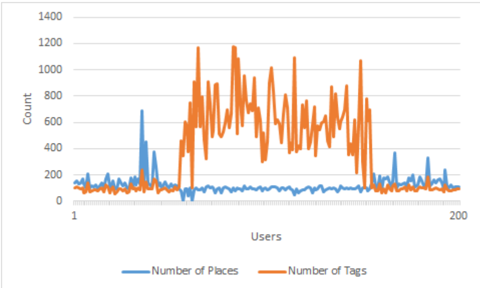


Figure 7.3: The Number of Distinct places and Tags for Users

Table 7.3: Low Frequent User Dataset

Number of Distinct Venues	4,411
Total number of Checkins	4,212
Total Number of Tips	2,900
Total Number of Tags	5,949
Number of users	200
Total Number categories	374
Total Number of Relationships	57,786

7.4.1 Evaluation of different user similarities

In this experiment, we want to evaluate the four basic user similarities coming from the basic user profiles. Thus, we want to evaluate spatial, semantic, categorical, and activity user similarity. Figure 7.4 shows the precision and recall values of comparing different user similarities. The semantic profile showed the best results in terms of precision and f1-measure. This means that the user similarity deduced from the semantic user profile is the most realistic.

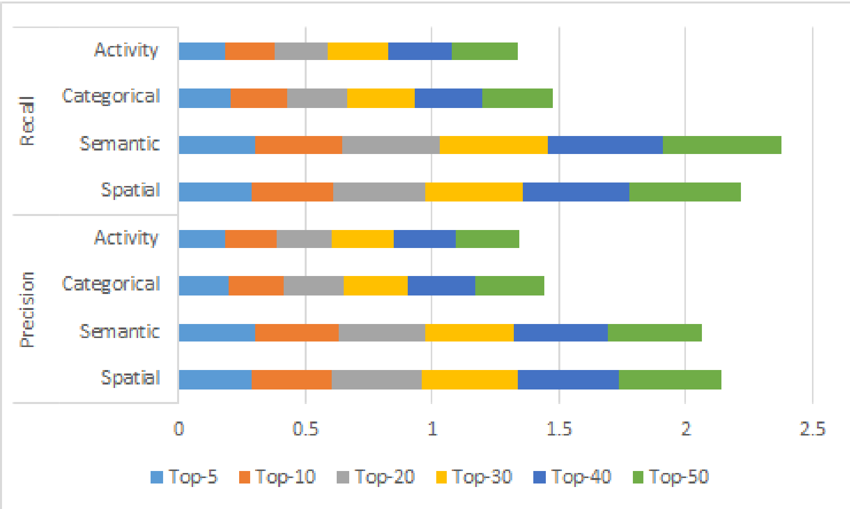


Figure 7.4: Evaluation of Basic User Similarities-Precision and Recall

7.4.2 Evaluation of Enrichment

This section show the effect of enrichment on user similarity. In the previous section, we observed that the spatial and semantic user similarities are the top representations of the different basic user similarity. Thus, in this section we will show the effect of enrichment on the spatial and semantic profiles by evaluating the user similarity between these two profiles.

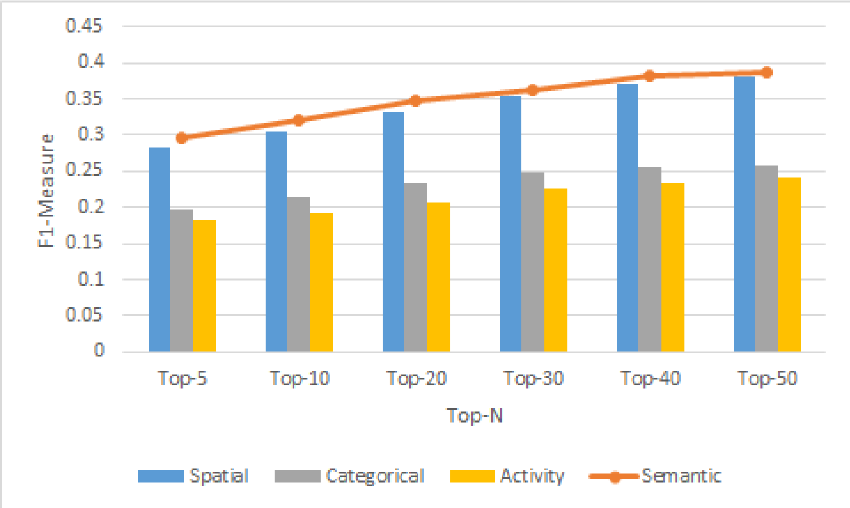


Figure 7.5: Evaluation of Basic User Similarities-F1-Measure

Evaluation of Spatial Profiles

The evaluation experiment aims to measure the impact of using the full range of content captured on LBSN when building user profiles in comparison to using only partial views based on check-in information. The user similarity between the following user profiles is calculated:

1. Spatial User profile
2. Enriched Spatial with $CosSim_{tag}$ (γ=1)
3. Enriched Spatial with $CosSim_{user}$ (γ=0)
4. Enriched Spatial with combined similarity ($\gamma = 0.5$)

The user similarities between the above profiles will be named as $user\_sim$, $user\_sim_{tag}$, $user\_sim_{user}$, $user\_sim_{combined}$, respectively.

Table 7.6 calculated the precision, recall and F1-measure values for the various user similarities. For each profile, frequent top-5, 10, 20, 30, 40, 50 venues are captured, and then categories are evaluated using the equations. As the table shows, enriching spatial profiles and linking similar places to create Enriched spatial user profiles improves the precision, recall and F-measure significantly. The best user similarity results come from the Enriched spatial user profiles with combined place similarity. This shows that although two users might visit very different locations, they can be similar because they carry out the same activities (like shopping), visit the same categories, or talk about same concepts together. To show the improvements in precision and recall, figures 7.6, 7.7, 7.8 the improvement of the enriched profile over the basic spatial profile. A positive value means that when using the enrichment process, the performance improved, whereas a negative value means that the performance dimentioned. It is observed from the figures that the average gain in precision and recall for a spatial profile enriched with combined are dominant, starting from the top-10 to the top-50.

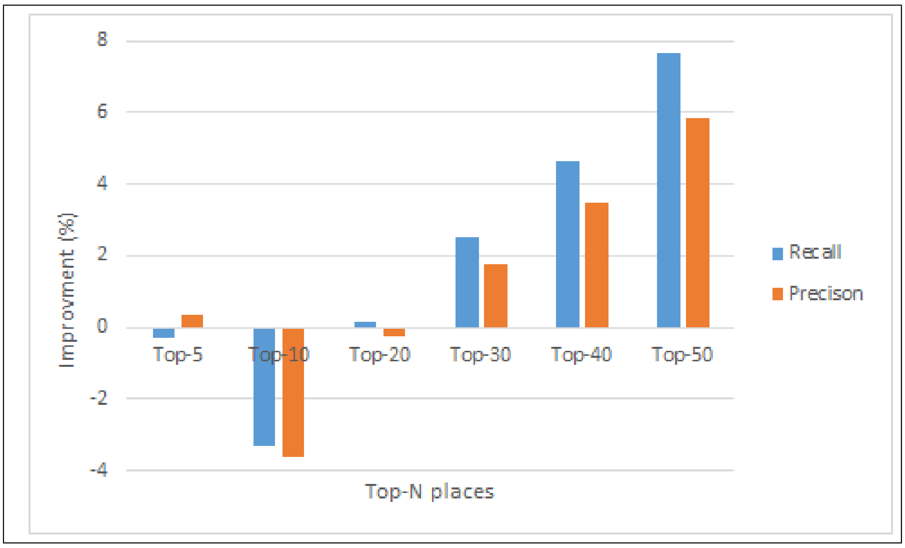


Figure 7.6: *$user\_sim_{tag}$* **Improvement over user_sim**

To evaluate the over all performance,*the Mean Average Precision* (MAP) ,is also employed to evaluate the performance of our method. MAP is the most frequently used summary measure of a ranked retrieval run. In this experiment, it stands for the mean of the precision score after each relevant user is retrieved for different top-N values.

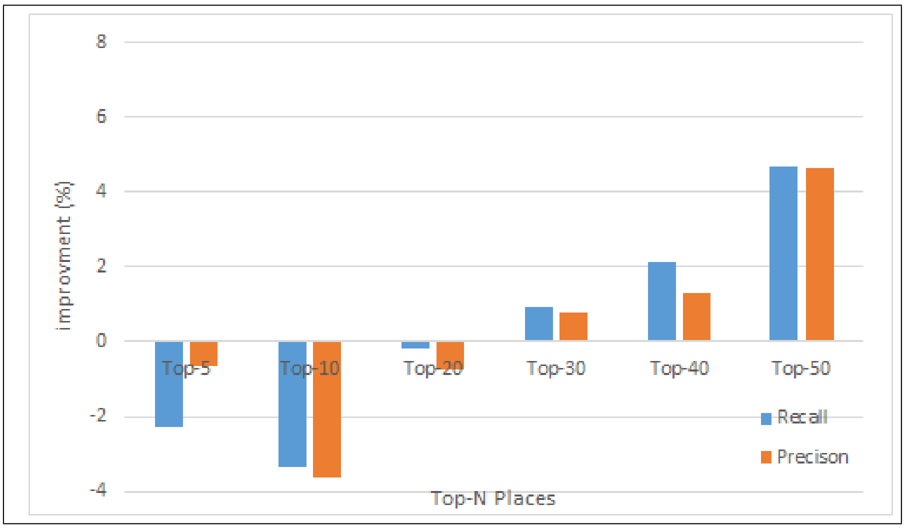


Figure 7.7: $user\_sim_{tag}$ **Improvement over user_sim**

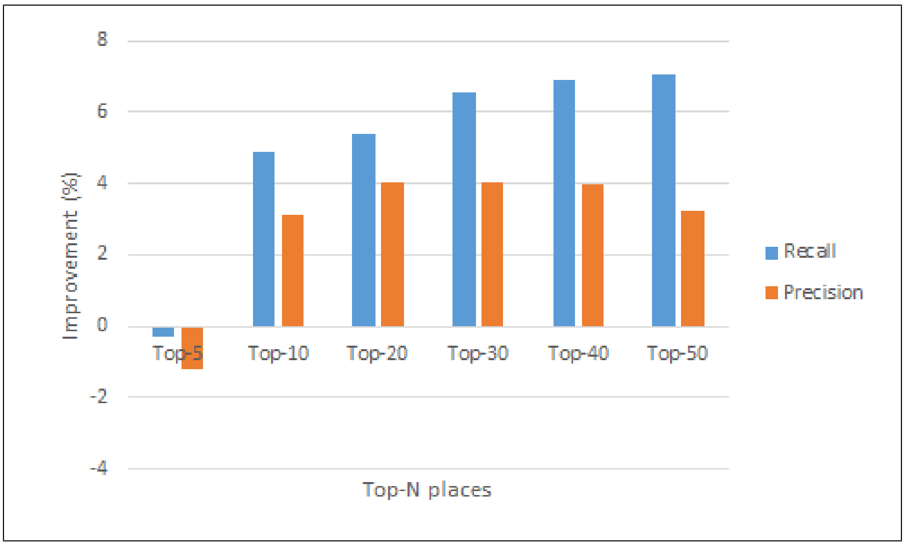


Figure 7.8: $user\_sim_{combined}$ **Improvement over user_sim**

Equation 7.4 shows how the MAP is calculated.

$$MAP = \frac{\sum_1^N p@n}{N} \tag{7.4}$$

Figure 7.9 shows the MAP results using the dataset.The figure shows a comparative study of MAP between different user similarities from different profiles baselines. Again, enriched combined user similarity shows clear advantages over all other similarities. Similarity computation with the enriched spatial profiles produce a higher

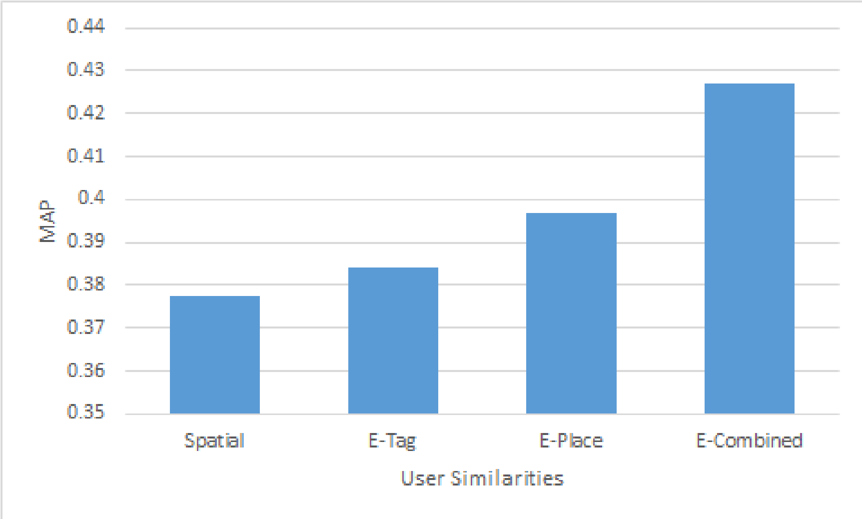


Figure 7.9: Mean Average Precision for Different User Similarities

degree of precision, recall and F-measures in general, whilst the best results are for the enriched profiles with the combined place similarity. Results indicate that location tracks may not be the best basis for finding similar users and that a combined treatment of both the spatial and semantic dimensions can produce more accurate views of user profiles.

Evaluation of semantic Profiles

This section will evaluate the user similarity from semantic profiles. The same experiment evaluating spatial profiles is repeated here. There are two types of semantic profiles: basic and enriched. Table 7.4 shows the precision and recall and F1 measure values of evaluating the basic and enriched semantic similarity between users. As the table shows, the enriched semantic similarity performed better than the basic one when evaluating it against the place categories of each user. This means that the process of enrichment enhances the similarity between users.

Table 7.4: Semantic Similarity Evaluation

	Precision		Recall		Fmeasure	
	Semantic	*Enriched Semantic*	*Semantic*	*Enriched Semantic*	*Semantic*	*Enriched Semantic*
Top-5	0.304228	0.284233	0.302118	0.275948	0.296848	0.277169
Top-10	0.337925	0.283362	0.326494	0.279485	0.319749	0.279089
Top-20	0.389438	0.365204	0.344888	0.360197	0.348238	0.35966
Top-30	0.423739	0.410384	0.351962	0.405662	0.361898	0.402931
Top-40	0.456672	0.452301	0.369331	0.443004	0.382173	0.441523
Top-50	0.468839	0.49545	0.372371	0.474493	0.386104	0.475632

Table 7.5: Descriptive Statistics for Different Users Categories

Descriptive Statistics	Low Frequent Users	High Frequent Users
Mean	26.685	123.455
Median	28	105.5
Mode	29	104
Standard Deviation	6.221682	65.91199
Sample Variance	38.70932	4344.39
Range	29	648
Minimum	9	42
Maximum	38	690
User Count	200	200

7.4.3 Frequent User Evaluation

In this section, two different user groups are used: low frequent users, and high frequent users. This was done by sorting the checking in or tipping activity and then choosing two hundred users from each category. Table 7.5 shows descriptive statistics of distinct venues for user categories.

Figure 7.10 shows the evaluation of high frequent users and low frequent users when using the enriched combined user similarity. As the figure shows, the precision and recall shows high values when considering users with high activity. This makes the model more important for highly frequent users.

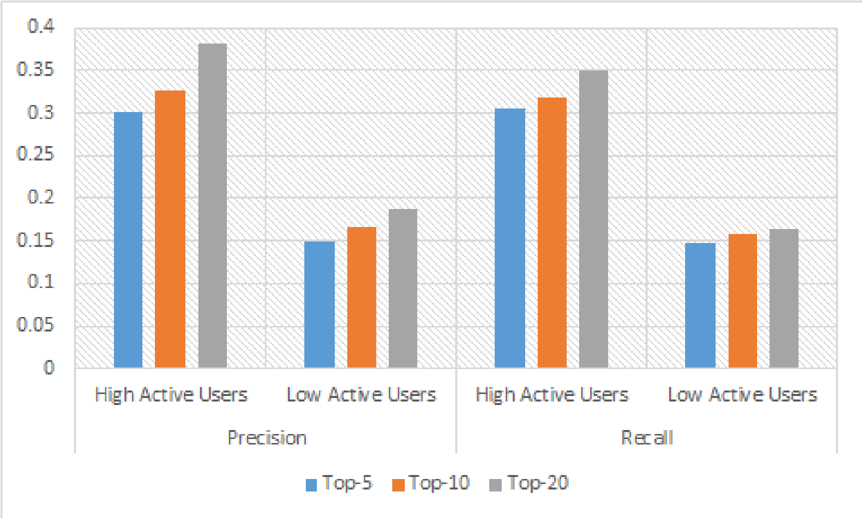


Figure 7.10: Activity Effect on Combined User similarity

7.4.4 Evaluation of Temporal Effect on User similarity

This section will compare short-term user similarity with long term user similarity over a two year period. The data is clustered into months. The different user similarities were calculated for each month by constructing the profiles within the month and then calculating the cosine similarity between each user and all other users throughout this period. The short term user similarity expresses a partial view of the profile within the month period. Thus, it is observed that some of the users in the data set do not have a user profile during this short period, and, accordingly, similarity with these users is zero. On the contrary, in long-term similarity, the profiles express the history of the user in the current month and previous months. For example, if the dataset time-stamp starts in January and one wants to calculate the user similarity values for users in March, one will first calculate the user profiles within January, February and March using the decay method. The user similarity is then calculated based on the user profiles created. Thus, the long-term user similarity expresses not only the similarity over one month, but adds up past similarities as well.

Regarding evaluation, the same metrics were applied for both short-term and long-term similarities to observe their performance. The two methods were compared using the combined spatial user similarity. Figure 7.11 shows the precision and recall of each method. As the Figure shows, the long-term user similarity performed better than the short-term user similarity in terms of both precision and recall. Thus, we conclude that it is better to calculate the user similarity between users based on the history of their interactions, not only the present interactions.

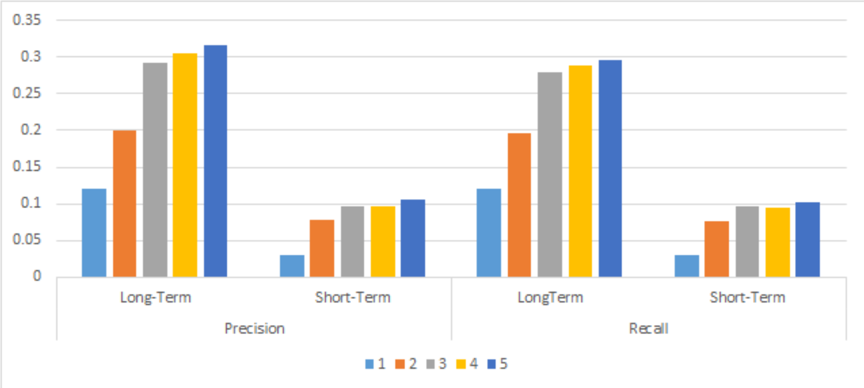


Figure 7.11: Evaluation of Short Term versus Long Term User Similarity

7.5 Summary

This chapter demonstrated different types of user similarity based on four characteristics: similarity of interests, similarity of co-location, similarity of categories visited, and similarity of activities in a place. Two similarity measures based on the temporal effect are also proposed, entitled "short-term user similarity" and "long-term user similarity". The different user similarity measures are evaluated via an information retrieval process in which the user information is utilised to search for the most similar user. The evaluation method showed that calculating user similarity based on semantic user interests is the best way of representing the relationship between users; furthermore, it was evident that user similarity can be more accurately measured when long-term activity regularities are captured to calculate the semantic similarity. Moreover, the results demonstrated that the enrichment process enhances the results. Finally, the results showed that the user similarity model is more important for high frequency users in comparison with low frequency users.

Table 7.6: User Similarity Evaluation: Precision, Recall, F1-measure

Precision				
Top-K Places	$user\_sim$	$user\_sim_{tag}$	$user\_sim_{user}$	$user\_sim_{combined}$
Top-5	0.29016885	0.2836818	**0.2936379**	0.27810863
Top-10	0.32131577	0.28528178	0.28525722	**0.35280818**
Top-20	0.3590904	0.35163218	0.35682544	**0.3996159**
Top-30	0.38940138	0.39706513	0.40721306	**0.42995644**
Top-40	0.41870615	0.43158174	0.45326504	**0.4587258**
Top-50	0.43747735	0.48375404	0.49606603	0.46999252
Recall				
Top-K Places	$user\_sim$	$user\_sim_{tag}$	$user\_sim_{user}$	$user\_sim_{combined}$
Top-5	0.2910496	0.26843706	0.28794414	**0.2881556**
Top-10	0.31549093	0.28207284	0.28236988	**0.36440614**
Top-20	0.35180694	0.34977493	0.35360697	**0.4058911**
Top-30	0.37913677	0.38837454	0.4045744	**0.44445464**
Top-40	0.39773872	0.41915244	0.44400847	**0.4669912**
Top-50	0.40748206	0.45438704	**0.48419788**	0.47782102
F1-measure				
Top-K Places	$user\_sim$	$user\_sim_{tag}$	$user\_sim_{user}$	$user\_sim_{combined}$
Top-5	0.29060855	0.27584896	**0.29076314**	0.28304298
Top-10	0.3183767	0.28366823	0.28380620	**0.35851338**
Top-20	0.35541135	0.35070109	0.35520891	**0.40272906**
Top-30	0.38420052	0.39267175	0.40588944	**0.43708534**
Top-40	0.40795319	0.42527629	0.44858900	**0.46282160**
Top-50	0.42194730	0.46861089	0.49006011	**0.4738744**

Chapter 8

Conclusions & Future Work

Knowledge of users' visits to places is one of the keys to understanding their interest in places. User-contributed annotations of place, as well as other place properties, add a layer of important semantics that if considered, can result in more refined representations of user profiles. In this work, a user and place modelling framework is proposed to represent spatial and semantic relationships between users, places, tags and time. The framework is used to build different views of personalised user profiles that is be used for enhancing place and content recommendation. The main contributions of this work can be roughly summarised as follows:

1. **A static and dynamic location-based user modelling framework from users' direct feedback on venues from LBSNs:** Users' interaction on LBSNs can be regarded as user feedback on geographic places they visited and interacted with. User's visits to places are recorded along with their comments and tags. Modelling different levels of user profiles extracted from the heterogeneous user feedback in LBSNs. User-generated traces at venues in LBSNs include both spatial and implicit semantic content. The location traces are treated equally to the semantic traces inferred from their interaction with the place through tagging and tipping. Collective behaviour of users on the network are also used to understand the place characteristics and these in turn are further used in the modelling of user profiles.

2. **A novel spatial and Semantic Enrichment process:** The proposed approach offers a uniform framework for presenting different views of user profiles, using their direct interactions with the social network or extended with a holistic view of other users interaction with the network in different regions of geographic space. Basic profiles capture direct user interactions, while enriched profiles offer an extended view of user's association with places and tags that take into account relationships in the folksonomy. In particular, it is shown that enriched user profiles can offer potentially more accurate views, than direct profiles, of user's spatial or semantic preferences.

3. **A novel semantic place model**: Both explicit place affordance; the sort of services offered in a place as denoted by its place type, and implicit place affordance; encapsulated in reference to activities in place annotations, are used in building semantic user profiles. Collective user spatial and semantic interactions with places are used to create profiles for geographic places, that in turn provide further enrichment to individual user profiles and is used to recommend activities for users based on user similarity

4. **A new location recommendation method:** The location recommendation method proposed is considered as a hybrid method of recommendation. User-based collaborative filtering is combined with content based filtering to recommend places based on places visited by similar users and places similar to the places visited by the user. Moreover, the semantic and the temporal information is encapsulated in the recommendation method to be more adaptive.

5. **A novel approach to user similarity evaluation:** Similarity between users on LBSN is approached in a uniform manner within the proposed framework, thus providing means of computing spatial, semantic or a combined view of user similarity on these networks.

8.1 Evaluating Research Hypothesis

The research hypothesis for this thesis was presented in Chapter 1. To remind the reader, the hypothesis is reiterated below:

The proliferation of GPS-enabled devices and their utilisation by users for geo-tagging personal resources, actions, and interactions on the Web is leading to the accumulation of a new type of information concerning individual users and user groups. The accumulation of spatiotemporal (ST) user footprints on the Social Web provides an opportunity for deriving profiles for both places, and users that closely reflect users' interests over space and time. Extracting and making sense of such profiles can enhance both place and content recommendation.

The research documented in this thesis, particularly, Chapters 4,5,6 tested this hypothesis to the point where it is possible to say that it is true. The research methodology presented in section 3.2 was followed to achieve this conclusion was to build a framework to a) collect realistic users' spatio-temporal data were extracted from their Social Web "footprints". Mainly, users' direct feedback on venues from LBSNs that captures users' interaction on LBSNs was collected. This can be regarded as user feedback on geographic places they visited and interacted with. User's visits to places are recorded along with their comments and tag; b) analyse the collected spatio-temporal data and design a framework for user and place modelling; c) evaluate the proposed framework to measure its quality using the well known information retrieval evaluation metrics: precision and recall.

Evaluation experiments are carried out using samples of realistic data sets from Foursquare LBSN. The influence from the user visits to places and the user behaviour in places are combined to define user and places. The framework proposed could also applied to other social networks such as twitter. However tweet analysis needs a process of identifying the places visited by users.

8.2 Answers of Research Questions

In this section, the research questions previously identified in Section 1.2 will be discussed in relation to the research undertaken in this thesis. Each research question will be repeated and the relevant research will be discussed including any related analysis, evaluation approaches and new knowledge that has been acquired.

Research Question 1: How can different views of user profiles be constructed from user footprints collected on LSBNs that emphasis the different facets of collected data? To answer this question, Chapter 4 introduced a user modelling framework entitled the 'geo-folksonomy model', which produces user profiles from LBSN data. The proposed approach provides users with the ability to project different views of their profiles using their direct interactions with the social network. This approach to modelling users in LBSN mainly represents a user's spatial, semantic and combined spatio-semantic association with place. A spatial user profile represents a user's interest in places, while a semantic profile describes his or her association with concepts associated with places in the folksonomy model. Finally, a spatio-semantic profile describes the user's specific interest in certain concepts associated with places in his or her profile. The word-net was also used to divide the tags into classes, which were then used to capture the semantics of tags in Chapter 5. In addition, Chapter 6 proposed two different methods for temporal user modelling : the *time-slice* approach and the *decay* approach. In the time-slice approach, user profiles are simply computed from the geo-folksonomy temporal graph ${\mathbb{G}_{\mathbb{F}}}^{i}$ for any time slot of interest t_i, while the other folksonomy graphs for $t \neq t_i$ are discarded. The decay approach considers the historical interactions in all sub-graphs of the folksonomy before the time point of interest.The results in Chapter 6 revealed the consideration of temporal constraints for the construction of user profiles improves the performance of personalised recommender systems. Furthermore, in comparison with the time-slice , the high quality decay method produced a better performance. This reveals that when designing user profiles, it is beneficial to learn and add historical interactions through time rather than simply taking a snapshot time window.

Research Question 2: How does this enrichment process impact the quality of personal user profiles? To improve the understanding of user profiles, an enrichment process was created to extend the basic profiles by using data from other users' interactions with the LBSN relating to various regions of geographic space. In Chapter 4, an extended profile that describes "recommended" associations given the overall interactions between users, places and concepts in the data-set was constructed. To model such interactions separately in the extended profile by controlling the similarity function used to create the profile. For example, one can focus on modelling the types of places visited or take into account the visit behaviour of other users whose profiles overlap with the user. Two types of enrichment have been undertaken: a) semantic enrichment (based on tag similarity calculations); and b) spatial enrichment (based on place similarity calculations). These calculations were used to enrich the basic profiles and to build different views of these enriched user profiles; these were then evaluated using a recommendation method. The results showed that enriched profiles perform better than basic profiles.

Research Question 3: How can implicit semantics of place profiles be used to reflect users experience in geographic places through the activities they carry out in those places?

Chapter 5 introduced a semantic place model that uses three primary concepts: place, place categories and place activities. In Chapter 5, a behavioural place category model was constructed. Moreover, motivated by our observation of users' interests in exploring new location categories, a category-based user profile that represents these interests in categories rather than locations was proposed. This changes user-place high-dimensional rating data into user-category reduced rating data. Using this new user model, an enhancement in category recommendation quality was evident in the results. Human related activities were also studied in Chapter 5 and activity related user and place profiles were proposed. Finally, an activity-aware category recommendation method was proposed, and the results shows effectiveness of the proposed algorithm.

Research Question 4: How can we construct a new location recommendation method using different dimension of LBSNs and evaluate it existing methods? In this thesis, the location recommendation problem was studied to evaluate the effectiveness of the user and place models proposed.To answer this question , we proposed in Chapter 4 a location recommendation method that uses user similarity and a combined place similarity to recommend locations for users. We also proposed in Chapter 6 a temporal decay location recommendation method that considers the four dimensions of the data: user, location, content, and time. The recommendation methods shows the effectiveness of the user and place modelling techniques used. These methods were compared to basic collaborative filtering recommendation methods and it showed better precision and recall results.

Research Question 5: How can different user profiles be evaluated using user similarity measures to assess their quality?

Studying user similarity from LBSN data is useful, as information available about users, their locations and activities is considered to be sparse. User similarities can be exploited to predict types of activities and places preferred by a user based on those of users with similar preferences. To answer the above question, different kinds of user similarity based on four characteristics: similarity of interests were created: similarity of co-location, similarity of categories visited and similarity of activities in the place. Two similarity measures based on the temporal effect were also proposed, entitled “short-term user similarity” and “long term user similarity”. The different user similarity measures are evaluated via an information retrieval process in which the user information is used to search for the most similar user. The evaluation method showed that calculating user similarity based on semantic user interests is the best way of representing the relationship between users; furthermore, it was evident that user similarity can be more accurate when long-term activity regularities were captured to calculate the semantic similarity. Moreover, the results demonstrated that the enrichment process enhances the results. Finally, the user similarity model was shown to be more effective when considering active users of the LBSN in comparison to low frequency or occasional users.

8.3 Future Work

This section discusses some points learned during our research and that highlights possible future directions.

1. **Big data challenge** Nowadays, with the increasing level of social and geo-spatial data, it is important to address the problem of data scalability. It is a challenge to handle large volume datasets efficiently and quickly because most computer systems do not have sufficient memory or computational power. In this research, the framework proposed has two problems: high complexity of algorithms proposed and high memory usage. To overcome these problems, a supercomputer that contain sufficient memory (Raven) was used to run our experiments. Another way to solve big data challenge is to use parallel algorithms such as map reduce to decrease the running time. Another possible directions for future work is to explore matrix factorisation reduction algorithms that transforms the data in the high-dimensional space to a space of fewer dimensions. The key advantages of reduction algorithms is that it reduces the time and storage space required. Moreover, it improves the performance of the machine learning model.

2. **Integrating data sources:** As users are now addicted to using social networks (e.g. Twitter, Flicker, Yelp, Foursquare), collecting data about a user from different data sources could be a promising future direction. In this research, we

used the foursquare only as our data source, but it is observed that one user can have multiple profiles in other social networks. The fusion and integration of social data from unstructured sources to extract value knowledge is an extremely difficult task which has not been completely explored. User profiles from such sources can be integrated and consequently enhance the problem of data sparsity and the performance of the location recommendation in general.

3. **Sentiment analysis:** In the proposed framework, the user opinion towards a place is mapped as words. A possible future direct is to extend the our proposed framework by integrating the sentiments or opinions extracted from the text of users' comments on POIs to improve the quality of location recommendation.

4. **Context awareness** In this research, the spatial, semantic, temporal, and social aspects were used. Another possible approach for the future is to add user context to location recommendations. Examples of user context include age, gender, profession, and income. An Environmental context could also help in improving recommendations as it includes information about weather, traffic and events.

Bibliography

[1] Topic model. https://en.wikipedia.org/wiki/Topic_model. Accessed: 2016-07-18.

[2] Fabian Abel. *Contextualization, User Modeling and Personalization in the Social Web*. Phd thesis, Gottfried Wilhelm Leibniz University Hannover, April 2011.

[3] Fabian Abel, Q Gao, GJ Houben, and K Tao. Contextualization, user modeling and personalization in the social web. In *Proceedings of Extended Semantic Web Conference (ESWC), Heraklion, Greece. Springer, Heidelberg*. Citeseer, 2011.

[4] Fabian Abel, Qi Gao, Geert-Jan Houben, and Ke Tao. Analyzing user modeling on twitter for personalized news recommendations. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 1–12. Springer, 2011.

[5] Fabian Abel, Qi Gao, Geert-Jan Houben, and Ke Tao. Semantic enrichment of twitter posts for user profile construction on the social web. In *Extended Semantic Web Conference*, pages 375–389. Springer, 2011.

[6] Fabian Abel, Claudia Hauff, Geert-Jan Houben, and Ke Tao. Leveraging user modeling on the social web with linked data. In *International Conference on Web Engineering*, pages 378–385. Springer, 2012.

[7] Gediminas Adomavicius and Alexander Tuzhilin. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE transactions on knowledge and data engineering*, 17(6):734–749, 2005.

[8] Gediminas Adomavicius and Alexander Tuzhilin. Context-aware recommender systems. In *Recommender systems handbook*, pages 217–253. Springer, 2011.

[9] Ching-man Au Yeung, Nicholas Gibbins, and Nigel Shadbolt. Contextualising tags in collaborative tagging systems. In *Proceedings of the 20th ACM conference on Hypertext and hypermedia*, pages 251–260. ACM, 2009.

[10] Sören Auer, Christian Bizer, Georgi Kobilarov, Jens Lehmann, Richard Cyganiak, and Zachary Ives. Dbpedia: A nucleus for a web of open data. In *The semantic web*, pages 722–735. Springer, 2007.

[11] Andrea Ballatore, Gavin McArdle, Caitriona Kelly, and Michela Bertolotto. Recomap: an interactive and adaptive map-based recommender. In *Proceedings of the 2010 ACM Symposium on Applied Computing*, pages 887–891. ACM, 2010.

[12] Linas Baltrunas and Xavier Amatriain. Towards time-dependant recommendation based on implicit feedback. In *Workshop on context-aware recommender systems (CARS'09)*, 2009.

[13] Jie Bao, Yu Zheng, David Wilkie, and Mohamed Mokbel. Recommendations in location-based social networks: a survey. *GeoInformatica*, 19(3):525–565, 2015.

[14] Samia Beldjoudi, Hassina Seridi, and Catherine Faron-Zucker. Ambiguity in tagging and the community effect in researching relevant resources in folksonomies. In *Proc. of ESWC Workshop User Profile Data on the Social Semantic Web*, 2011.

[15] Paul N Bennett, Ryen W White, Wei Chu, Susan T Dumais, Peter Bailey, Fedor Borisyuk, and Xiaoyuan Cui. Modeling the impact of short-and long-term be-

havior on search personalization. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 185–194. ACM, 2012.

[16] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1):107–117, 1998.

[17] Peter Brusilovsky and Eva Millán. User models for adaptive hypermedia and adaptive educational systems. In *The adaptive web*, pages 3–53. Springer-Verlag, 2007.

[18] Ceren Budak, Anitha Kannan, Rakesh Agrawal, and Jan Pedersen. Inferring user interests from microblogs. Technical report, Tech. Rep, 2014.

[19] Robin Burke. Hybrid web recommender systems. In *The adaptive web*, pages 377–408. Springer, 2007.

[20] Pedro G Campos, Fernando Díez, and Iván Cantador. Time-aware recommender systems: a comprehensive survey and analysis of existing evaluation protocols. *User Modeling and User-Adapted Interaction*, 24(1-2):67–119, 2014.

[21] Xin Cao, Gao Cong, and Christian S Jensen. Mining significant semantic locations from gps data. *Proceedings of the VLDB Endowment*, 3(1-2):1009–1020, 2010.

[22] Ciro Cattuto, Alain Barrat, Andrea Baldassarri, Gregory Schehr, and Vittorio Loreto. Collective dynamics of social annotation. *Proceedings of the National Academy of Sciences*, 106(26):10511–10515, 2009.

[23] Ciro Cattuto, Dominik Benz, Andreas Hotho, and Gerd Stumme. Semantic analysis of tag similarity measures in collaborative tagging systems. *arXiv preprint arXiv:0805.2045*, 2008.

[24] Jilin Chen, Rowan Nairn, Les Nelson, Michael Bernstein, and Ed Chi. Short and tweet: experiments on recommending content from information streams.

In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1185–1194. ACM, 2010.

[25] Chen Cheng, Haiqin Yang, Irwin King, and Michael R Lyu. Fused matrix factorization with geographical and social influence in location-based social networks. In *Aaai*, volume 12, pages 17–23, 2012.

[26] Eunjoon Cho, Seth A Myers, and Jure Leskovec. Friendship and mobility: user movement in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1082–1090. ACM, 2011.

[27] Chi-Yin Chow, Jie Bao, and Mohamed F Mokbel. Towards location-based social networking services. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, pages 31–38. ACM, 2010.

[28] Lucia Del Prete and Licia Capra. differs: A mobile recommender service. In *2010 Eleventh International Conference on Mobile Data Management*, pages 21–26. IEEE, 2010.

[29] Yi Ding and Xue Li. Time weight collaborative filtering. In *Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 485–492. ACM, 2005.

[30] Cynthia Dwork, Ravi Kumar, Moni Naor, and Dandapani Sivakumar. Rank aggregation methods for the web. In *Proceedings of the 10th international conference on World Wide Web*, pages 613–622. ACM, 2001.

[31] Ehab ElGindy and Alia Abdelmoty. Enriching user profiles using geo-social place semantics in geo-folksonomies. *International Journal of Geographical Information Science*, 28(7):1439–1458, 2014.

[32] Gregory Ference, Mao Ye, and Wang-Chien Lee. Location recommendation for out-of-town users in location-based social networks. In *Proceedings of the*

22nd ACM international conference on Information & Knowledge Management, pages 721–726. ACM, 2013.

[33] L Fonda, GC Ghirardi, and A Rimini. Decay theory of unstable quantum systems. *Reports on Progress in Physics*, 41(4):587, 1978.

[34] Huiji Gao, Jiliang Tang, Xia Hu, and Huan Liu. Exploring temporal effects for location recommendation on location-based social networks. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 93–100. ACM, 2013.

[35] Huiji Gao, Jiliang Tang, Xia Hu, and Huan Liu. Content-aware point of interest recommendation on location-based social networks. In *AAAI*, pages 1721–1727. Citeseer, 2015.

[36] Huiji Gao, Jiliang Tang, and Huan Liu. Exploring social-historical ties on location-based social networks. In *ICWSM*, 2012.

[37] Huiji Gao, Jiliang Tang, and Huan Liu. gscorr: modeling geo-social correlations for new check-ins on location-based social networks. In *Proceedings of the 21st ACM international conference on Information and knowledge management*, pages 1582–1586. ACM, 2012.

[38] Anna Lisa Gentile, Vitaveska Lanfranchi, Suvodeep Mazumdar, and Fabio Ciravegna. Extracting semantic user networks from informal communication exchanges. In *International Semantic Web Conference*, pages 209–224. Springer, 2011.

[39] Gianluigi Gentili, Alessandro Micarelli, and Filippo Sciarrone. Infoweb: An adaptive information filtering system for the cultural heritage domain. *Applied Artificial Intelligence*, 17(8-9):715–744, 2003.

[40] M Rami Ghorab, Dong Zhou, Alexander O' Connor, and Vincent Wade. Personalised information retrieval: survey and classification. *User Modeling and User-Adapted Interaction*, 23(4):381–443, 2013.

[41] Eric Gilbert and Karrie Karahalios. Predicting tie strength with social media. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 211–220. ACM, 2009.

[42] Tom Gross, Chris Stary, and Alex Totter. User-centered awareness in computer-supported cooperative work-systems: Structured embedding of findings from social sciences. *International Journal of Human-Computer Interaction*, 18(3):323–360, 2005.

[43] Şule Gündüz-Ögüdücü. Web page recommendation models: theory and algorithms. *Synthesis Lectures on Data Management*, 2(1):1–85, 2010.

[44] Harry Halpin, Valentin Robu, and Hana Shepherd. The complex dynamics of collaborative tagging. In *Proceedings of the 16th international conference on World Wide Web*, pages 211–220. ACM, 2007.

[45] Morgan Harvey, Fabio Crestani, and Mark J Carman. Building user profiles from topic models for personalised search. In *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, pages 2309–2314. ACM, 2013.

[46] Claudia Hauff, Marcel Berthold, Geert-Jan Houben, Christina M Steiner, and Dietrich Albert. Tweets reveal more than you know: a learning style analysis on twitter. In *European Conference on Technology Enhanced Learning*, pages 140–152. Springer, 2012.

[47] Jonathan L Herlocker, Joseph A Konstan, Al Borchers, and John Riedl. An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 230–237. ACM, 1999.

[48] Christoph Hermann. Time-based recommendations for lecture materials. In *2010 World Conference on Educational Multimedia, Hypermedia and Telecommunications*, pages 1028–1033, 2010.

[49] Paul Heymann, Daniel Ramage, and Hector Garcia-Molina. Social tag prediction. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, pages 531–538. ACM, 2008.

[50] Tzvetan Horozov, Nitya Narasimhan, and Venu Vasudevan. Using location for personalized poi recommendations in mobile environments. In *International Symposium on Applications and the Internet (SAINT'06)*, pages 6–pp. IEEE, 2006.

[51] Andreas Hotho, Robert Jäschke, Christoph Schmitz, and Gerd Stumme. Information retrieval in folksonomies: Search and ranking. In *Semantic web: research and applications, proceedings*, pages 411–426. Springer, 2006.

[52] Andreas Hotho, Robert Jäschke, Christoph Schmitz, Gerd Stumme, and Klaus-Dieter Althoff. Folkrank: A ranking algorithm for folksonomies. In *LWA*, volume 1, pages 111–114, 2006.

[53] Bo Hu and Martin Ester. Spatial topic modeling in online social media for location recommendation. In *Proceedings of the 7th ACM conference on Recommender systems*, pages 25–32. ACM, 2013.

[54] Longke Hu, Aixin Sun, and Yong Liu. Your neighbors affect your ratings: on geographical neighborhood influence to rating prediction. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 345–354. ACM, 2014.

[55] Anthony Jameson, Cécile Paris, and Carlo Tasso. 'reader's guide,". In *User Modeling, Proceedings of the Sixth International Conference UM97*, 1997.

[56] Zhiyang Jia, Yuting Yang, Wei Gao, and Xu Chen. User-based collaborative filtering for tourist attraction recommendations. In *Computational Intelligence & Communication Technology (CICT), 2015 IEEE International Conference on*, pages 22–25. IEEE, 2015.

[57] Alexandros Karatzoglou. Collaborative temporal order modeling. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 313–316. ACM, 2011.

[58] Adam Kilgarriff and Christiane Fellbaum. Wordnet: An electronic lexical database, 2000.

[59] Sheila Kinsella, Mengjiao Wang, John G Breslin, and Conor Hayes. Improving categorisation in social media using hyperlinks to structured data sources. In *Extended Semantic Web Conference*, pages 390–404. Springer, 2011.

[60] Kazuki Kodama, Yuichi Iijima, Xi Guo, and Yoshiharu Ishikawa. Skyline queries based on user locations and preferences for making location-based recommendations. In *Proceedings of the 2009 International Workshop on Location Based Social Networks*, pages 9–16. ACM, 2009.

[61] Noam Koenigstein, Gideon Dror, and Yehuda Koren. Yahoo! music recommendations: modeling music ratings with temporal dynamics and item taxonomy. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 165–172. ACM, 2011.

[62] Yehuda Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 426–434. ACM, 2008.

[63] Yehuda Koren. Collaborative filtering with temporal dynamics. *Communications of the ACM*, 53(4):89–97, 2010.

[64] Yehuda Koren, Robert Bell, Chris Volinsky, et al. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.

[65] Georgia Koutrika, Benjamin Bercovitz, and Hector Garcia-Molina. Flexrecs: expressing and combining flexible recommendations. In *Proceedings of the*

2009 ACM SIGMOD International Conference on Management of data, pages 745–758. ACM, 2009.

[66] Dominik Kowald, Paul Seitlinger, Christoph Trattner, and Tobias Ley. Long time no see: The probability of reusing tags as a function of frequency and recency. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 463–468. ACM, 2014.

[67] Nikolas Landia. *Content-awareness and graph-based ranking for tag recommendation in folksonomies*. PhD thesis, University of Warwick, 2013.

[68] Nikolas Landia, Sarabjot Singh Anand, Andreas Hotho, Robert Jäschke, Stephan Doerfel, and Folke Mitzlaff. Extending folkrank with content data. In *Proceedings of the 4th ACM RecSys workshop on Recommender systems and the social web*, pages 1–8. ACM, 2012.

[69] Santiago Larrain, Christoph Trattner, Denis Parra, Eduardo Graells-Garrido, and Kjetil Nørvåg. Good times bad times: A study on recency effects in collaborative filtering for social tagging. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 269–272. ACM, 2015.

[70] Dongjoo Lee, Sung Eun Park, Minsuk Kahng, Sangkeun Lee, and Sang-goo Lee. Exploiting contextual information from event logs for personalized recommendation. In *Computer and Information Science 2010*, pages 121–139. Springer, 2010.

[71] Min-Joong Lee and Chin-Wan Chung. A user similarity calculation based on the location for social network services. In *Database Systems for Advanced Applications*, pages 38–52. Springer, 2011.

[72] Tong Queue Lee, Young Park, and Yong-Tae Park. A time-based approach to effective recommender systems using implicit feedback. *Expert systems with applications*, 34(4):3055–3062, 2008.

[73] Tong Queue Lee, Young Park, and Yong-Tae Park. An empirical study on effectiveness of temporal information as implicit ratings. *Expert systems with Applications*, 36(2):1315–1321, 2009.

[74] Wee Sun Lee. Collaborative learning for recommender systems. In *ICML*, volume 1, pages 314–321, 2001.

[75] Lei Li, Li Zheng, and Tao Li. Logo: a long-short user interest integration in personalized news recommendation. In *Proceedings of the fifth ACM conference on Recommender systems*, pages 317–320. ACM, 2011.

[76] Quannan Li, Yu Zheng, Xing Xie, Yukun Chen, Wenyu Liu, and Wei-Ying Ma. Mining user similarity based on location history. In *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*, page 34. ACM, 2008.

[77] Bin Liu, Yanjie Fu, Zijun Yao, and Hui Xiong. Learning geographical preferences for point-of-interest recommendation. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1043–1051. ACM, 2013.

[78] Bin Liu, Hui Xiong, Spiros Papadimitriou, Yanjie Fu, and Zijun Yao. A general geographical probabilistic factor model for point of interest recommendation. *Knowledge and Data Engineering, IEEE Transactions on*, 27(5):1167–1179, 2015.

[79] Nathan N Liu, Luheng He, and Min Zhao. Social temporal collaborative ranking for context aware movie recommendation. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 4(1):15, 2013.

[80] Zhiyuan Liu, Xinxiong Chen, and Maosong Sun. Mining the interests of chinese microbloggers via keyword extraction. *Frontiers of Computer Science*, 6(1):76–87, 2012.

[81] Pasquale Lops, Marco De Gemmis, and Giovanni Semeraro. Content-based recommender systems: State of the art and trends. In *Recommender systems handbook*, pages 73–105. Springer, 2011.

[82] Grant McKenzie, Benjamin Adams, and Krzysztof Janowicz. A thematic approach to user similarity built on geosocial check-ins. In *Geographic Information Science at the Heart of Europe*, pages 39–53. Springer, 2013.

[83] Alan Mislove, Bimal Viswanath, Krishna P Gummadi, and Peter Druschel. You are who you know: inferring user profiles in online social networks. In *Proceedings of the third ACM international conference on Web search and data mining*, pages 251–260. ACM, 2010.

[84] David Newman, Youn Noh, Edmund Talley, Sarvnaz Karimi, and Timothy Baldwin. Evaluating topic models for digital libraries. In *Proceedings of the 10th annual joint conference on Digital libraries*, pages 215–224. ACM, 2010.

[85] Satoshi Niwa, Takuo Doi, and Shinichi Honiden. Web page recommender system based on folksonomy mining for itng'06 submissions. In *Third International Conference on Information Technology: New Generations (ITNG'06)*, pages 388–393. IEEE, 2006.

[86] Anastasios Noulas and Cecilia Mascolo. Exploiting foursquare and cellular data to infer user activity in urban environments. In *14th International Conference on Mobile Data Management*, volume 1, pages 167–176. IEEE, 2013.

[87] Fabrizio Orlandi, John Breslin, and Alexandre Passant. Aggregated, interoperable and multi-domain user profiles for the social web. In *Proceedings of the 8th International Conference on Semantic Systems*, pages 41–48. ACM, 2012.

[88] Moon-Hee Park, Jin-Hyuk Hong, and Sung-Bae Cho. Location-based recommendation system using bayesian user's preference model in mobile devices. In *Ubiquitous Intelligence and Computing*, pages 1130–1139. Springer, 2007.

[89] Michael J Pazzani. A framework for collaborative, content-based and demographic filtering. *Artificial Intelligence Review*, 13(5-6):393–408, 1999.

[90] Ken Peffers, Tuure Tuunanen, Charles E Gengler, Matti Rossi, Wendy Hui, Ville Virtanen, and Johanna Bragge. The design science research process: a model for producing and presenting information systems research. In *Proceedings of the first international conference on design science research in information systems and technology (DESRIST 2006)*, pages 83–106, 2006.

[91] H Pragarauskas and Oliver Gross. Temporal collaborative filtering with bayesian probabilistic tensor factorization. 2010.

[92] Zhiming Qian, Ping Zhong, and Runsheng Wang. Tag refinement for user-contributed images via graph learning and nonnegative tensor factorization. *Signal Processing Letters, IEEE*, 22(9):1302–1305, 2015.

[93] Giovanni Quattrone, Emilio Ferrara, Pasquale De Meo, and Licia Capra. Measuring similarity in large-scale folksonomies. *CoRR*, abs/1207.6037, 2012.

[94] Reza Rafeh and Arash Bahrehmand. An adaptive approach to dealing with unstable behaviour of users in collaborative filtering systems. *Journal of Information Science*, 38(3):205–221, 2012.

[95] Seyyed Mohammadreza Rahimi and Xin Wang. Location recommendation based on periodicity of human activities and location categories. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, pages 377–389. Springer, 2013.

[96] Lakshmish Ramaswamy, P Deepak, Ramana Polavarapu, Kutila Gunasekera, Dinesh Garg, Karthik Visweswariah, and Shivkumar Kalyanaraman. Caesar: A context-aware, social recommender system for low-end mobile devices. In *2009 Tenth International Conference on Mobile Data Management: Systems, Services and Middleware*, pages 338–347. IEEE, 2009.

[97] Steffen Rendle, Zeno Gantner, Christoph Freudenthaler, and Lars Schmidt-Thieme. Fast context-aware recommendations with factorization machines. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 635–644. ACM, 2011.

[98] Paul Resnick, Neophytos Iacovou, Mitesh Suchak, Peter Bergstrom, and John Riedl. Grouplens: an open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM conference on Computer supported cooperative work*, pages 175–186. ACM, 1994.

[99] Gerard Salton, Anita Wong, and Chung-Shu Yang. A vector space model for automatic indexing. *Communications of the ACM*, 18(11):613–620, 1975.

[100] Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl. Item-based collaborative filtering recommendation algorithms. In *Proceedings of the 10th international conference on World Wide Web*, pages 285–295. ACM, 2001.

[101] J Ben Schafer, Dan Frankowski, Jon Herlocker, and Shilad Sen. Collaborative filtering recommender systems. In *The adaptive web*, pages 291–324. Springer, 2007.

[102] Upendra Shardanand and Pattie Maes. Social information filtering: algorithms for automating word of mouth. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 210–217. ACM Press/Addison-Wesley Publishing Co., 1995.

[103] Yue Shi, Pavel Serdyukov, Alan Hanjalic, and Martha Larson. Personalized landmark recommendation based on geotags from photo sharing sites. *ICWSM*, 11:622–625, 2011.

[104] Panagiotis Symeonidis, Alexandros Nanopoulos, and Yannis Manolopoulos. Tag recommendations based on tensor dimensionality reduction. In *Proceedings of the 2008 ACM conference on Recommender systems*, pages 43–50. ACM, 2008.

[105] Yuichiro Takeuchi and Masanori Sugimoto. Cityvoyager: an outdoor recommendation system based on user location history. In *International Conference on Ubiquitous Intelligence and Computing*, pages 625–636. Springer, 2006.

[106] Jian Tang, Zhaoshi Meng, Xuanlong Nguyen, Qiaozhu Mei, and Ming Zhang. Understanding the limiting factors of topic modeling via posterior contraction analysis. In *ICML*, pages 190–198, 2014.

[107] Thomas Vanderwal. Folksonomy, url:http://www.vanderwal.net/folksonomy.html. 2007.

[108] Petros Venetis, Hector Gonzalez, Christian S Jensen, and Alon Halevy. Hyper-local, directions-based ranking of places. *Proceedings of the VLDB Endowment*, 4(5):290–301, 2011.

[109] Thanh Tien Vu, Dawei Song, Alistair Willis, Son Ngoc Tran, and Jingfei Li. Improving search personalisation with dynamic group formation. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval*, pages 951–954. ACM, 2014.

[110] Hao Wang, Manolis Terrovitis, and Nikos Mamoulis. Location recommendation in location-based social networks using user check-in data. In *Proceedings of the 21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 374–383. ACM, 2013.

[111] Joe Weakliam, Michela Bertolotto, and David Wilson. Implicit interaction profiling for recommending spatial content. In *Proceedings of the 13th annual ACM international workshop on Geographic information systems*, pages 285–294. ACM, 2005.

[112] wiki pedia. Folksonomy, url:http://en.wikipedia.org/wiki/folksonomy. 2014.

[113] Dengneng Wu, Zhenming Yuan, Kai Yu, and Hong Pan. Temporal social tagging based collaborative filtering recommender for digital library. In *International Conference on Asian Digital Libraries*, pages 199–208. Springer, 2012.

[114] Liang Xiang, Quan Yuan, Shiwan Zhao, Li Chen, Xiatian Zhang, Qing Yang, and Jimeng Sun. Temporal recommendation on graphs via long-and short-term preference fusion. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 723–732. ACM, 2010.

[115] Rongjing Xiang, Jennifer Neville, and Monica Rogati. Modeling relationship strength in online social networks. In *Proceedings of the 19th international conference on World wide web*, pages 981–990. ACM, 2010.

[116] Dingqi Yang, Daqing Zhang, Zhiyong Yu, and Zhu Wang. A sentiment-enhanced personalized location recommendation system. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, pages 119–128. ACM, 2013.

[117] Xiwang Yang, Yang Guo, Yong Liu, and Harald Steck. A survey of collaborative filtering based social recommender systems. *Computer Communications*, 41:1–10, 2014.

[118] Junjie Yao, Bin Cui, Yuxin Huang, and Xin Jin. Temporal and social context based burst detection from folksonomies. In *AAAI*, 2010.

[119] Mao Ye, Krzysztof Janowicz, Christoph Mülligann, and Wang-Chien Lee. What you are is when you are: the temporal dimension of feature types in location-based social networks. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 102–111. ACM, 2011.

[120] Mao Ye, Xingjie Liu, and Wang-Chien Lee. Exploring social influence for recommendation-a probabilistic generative model approach. *arXiv preprint arXiv:1109.0758*, 2011.

[121] Mao Ye, Dong Shou, Wang-Chien Lee, Peifeng Yin, and Krzysztof Janowicz. On the semantic annotation of places in location-based social networks. In *Proceedings of the 17th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 520–528. ACM, 2011.

[122] Mao Ye, Peifeng Yin, and Wang-Chien Lee. Location recommendation for location-based social networks. In *Proceedings of the 18th SIGSPATIAL International Conference on Advances in Geographic Information Systems*, pages 458–461. ACM, 2010.

[123] Mao Ye, Peifeng Yin, Wang-Chien Lee, and Dik-Lun Lee. Exploiting geographical influence for collaborative point-of-interest recommendation. In *Proceedings of the 34th international ACM SIGIR conference on Research and development in Information Retrieval*, pages 325–334. ACM, 2011.

[124] Josh Jia-Ching Ying, Wen-Ning Kuo, Vincent S Tseng, and Eric Hsueh-Chan Lu. Mining user check-in behavior with a random walk for urban point-of-interest recommendations. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(3):40, 2014.

[125] Josh Jia-Ching Ying, Eric Hsueh-Chan Lu, Wen-Ning Kuo, and Vincent S Tseng. Urban point-of-interest recommendation by mining user check-in behaviors. In *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, pages 63–70. ACM, 2012.

[126] Quan Yuan, Gao Cong, Zongyang Ma, Aixin Sun, and Nadia Magnenat Thalmann. Time-aware point-of-interest recommendation. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*, pages 363–372. ACM, 2013.

[127] Jia-Dong Zhang and Chi-Yin Chow. igslr: personalized geo-social location recommendation: a kernel density estimation approach. In *Proceedings of the*

21st ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, pages 334–343. ACM, 2013.

[128] Jia-Dong Zhang and Chi-Yin Chow. Ticrec: A probabilistic framework to utilize temporal influence correlations for time-aware location recommendations. 2015.

[129] Zhongfeng Zhang, Qiudan Li, and Daniel Zeng. Mining evolutionary topic patterns in community question answering systems. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 41(5):828–833, 2011.

[130] Zi-Ke Zhang and Chuang Liu. A hypergraph model of social tagging networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(10):P10005, 2010.

[131] Yi-Liang Zhao, Liqiang Nie, Xiangyu Wang, and Tat-Seng Chua. Personalized recommendations of locally interesting venues to tourists via cross-region community matching. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(3):50, 2014.

[132] Vincent Wenchen Zheng, Bin Cao, Yu Zheng, Xing Xie, and Qiang Yang. Collaborative filtering meets mobile recommendation: A user-centered approach. In *AAAI*, volume 10, pages 236–241, 2010.

[133] Dequan Zhou, Bin Wang, Seyyed Mohammadreza Rahimi, and Xin Wang. A study of recommending locations on location-based social network by collaborative filtering. In *Advances in Artificial Intelligence*, pages 255–266. Springer, 2012.

[134] Vinko Zlatić, Gourab Ghoshal, and Guido Caldarelli. Hypergraph topological quantities for tagged social networks. *Physical Review E*, 80(3):036118, 2009.