A study of geographical neighborhood influence to business rating prediction

Hu, Longke

2014

Hu, L. (2014). A study of geographical neighborhood influence to business rating prediction.
Master's thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/61791>

<https://doi.org/10.32657/10356/61791>

NANYANG
TECHNOLOGICAL
UNIVERSITY

A STUDY OF GEOGRAPHICAL
NEIGHBORHOOD INFLUENCE TO BUSINESS
RATING PREDICTION

HU LONGKE

SCHOOL OF COMPUTER ENGINEERING

2014

A Study of Geographical Neighborhood Influence to Business Rating Prediction

Hu Longke

School of Computer Engineering

A thesis submitted to the Nanyang Technological University
in fulfilment of the requirement for the degree of
Master of Engineering

2014

Acknowledgments

I would like to express my sincere gratitude to my supervisor Prof. Aixin Sun from School of Computer Engineering, who gives me invaluable assistance and patient guidance. Under his supervision, I am fortunate to learn state-of-the-art techniques on information retrieval and data mining. His support is unforgettable for me, and his suggestions on my research inspire me to work more carefully and harder.

I would also like to thank all my friends, Yong Liu, Zhen Hai, Surendra Sedhai, Yangjie Yao, Chenliang Li, Zongyang Ma, Zhiqiang Xu, Shanshan Feng, Dingxian Wang. Discussions with them were really helpful and pleasant.

Last but not the least, I wish to express my gratitude to my family for their selfless love and understanding through my life.

Contents

Acknowledgments	i
List of Figures	iv
List of Tables	v
Abstract	vi
1 Introduction	1
1.1 Background and Motivation	1
1.2 Contributions	3
1.3 Thesis Organization	4
2 Related Work	6
2.1 Recommendation Systems	6
2.2 Collaborative Filtering	10
2.3 POI Recommendation and Prediction	15
3 Data Analysis on Yelp Dataset	18
3.1 Pearson Correlation	18
3.2 Data Analysis	19
3.2.1 Dataset	19
3.2.2 Observations	20
4 Business Rating Prediction	27
4.1 Matrix Factorization	27
4.2 Incorporating Neighborhood Influence	29
4.3 Incorporating Category Influence	30
4.4 Incorporating Review Content	31
4.5 Popularity and Geo-distance Influences	31

5	Experiment	34
5.1	Experimental Setting	34
5.2	Experimental Results	37
5.2.1	Method Comparison	37
5.2.2	Impact of Neighborhood Size	39
5.2.3	Cold-Start Business Rating Prediction	40
6	Conclusions and Future Work	44
6.1	Conclusion	44
6.2	Limitations and Implications	46
6.3	Future Work	46
A	Parameter Estimation	48
B	Author’s Publications	49
	References	50

List of Figures

3.1	Percentage of businesses having at least 1, 3, 6, 10 neighbors within a distance threshold from 20 to 2000 meters	20
3.2	Pearson’s correlation between a business’s rating and the average rating of its {1, 3, 6, 10} nearest neighbors (NN), within a distance threshold.	21
3.3	Business category distribution	23
3.4	Category-wise rating correlation with nearest neighbors (1NN, 6NN) .	25
3.5	The percentage of nearest neighbors (1NN, 6NN) in the same category	26
5.1	Improvement over GlobalMean (μ)	38
5.2	Prediction error of N -MF for determining geographical neighbors by distance	42
5.3	Prediction error of N -MF for determining geographical neighbors by number	43

List of Tables

3.1	Number of reviews/ratings per user/business	19
4.1	Notations and semantics	28
5.1	MAE and RMSE of all methods, the lower the better.	37
5.2	MAE and RMSE in cold-start setting	40

Abstract

Rating prediction is to predict the preference rating of a user to an item that she has not rated before, and it is one of the most popular and fundamental problems in recommendation systems. Using the business review data from Yelp, we study the problem of business rating prediction in this thesis. A business here can be a restaurant, a shopping mall, a nightlife club or other kind of businesses. Different from most other types of items that have been studied in various recommender systems (*e.g.*, movie, song, book), a business in Yelp physically exists at a geographical location, and most businesses have geographical neighbors within walking distance. When a user visits a business, there is a good chance that she walks by its neighbors. Through data analysis on Yelp, we find that there exists weak positive correlation between a business's ratings and its neighbors' ratings, and the positive correlation in ratings is independent of the categories of the businesses and/or their neighbors. Based on this observation, we assume that a user's rating to a given business is determined by both the intrinsic characteristics of the business and the extrinsic characteristics of its geographical neighbors. Using the widely adopted latent factor model for rating prediction, in our proposed solution, we use two kinds of latent factors to model a business: one for its intrinsic characteristics and the other for its extrinsic characteristics. More specifically, the former encodes the intrinsic characteristics of a business (*e.g.*, taste of food and quality of service) observable by users who have interacted with the business. The latter encodes the extrinsic characteristics of a business (*e.g.*, hygiene standard) in influencing its geographical neighbors observable by the "pass-by" visitors.

We conduct extensive experiments on the Yelp dataset to evaluate the proposed models, and compare the models with state-of-the-art baseline methods. We show that by incorporating geographical neighborhood influences, much lower prediction

error is achieved than the baseline models including Biased MF, SVD++, and Social MF. The prediction error is further reduced by incorporating influences from business category and review content.

Chapter 1

Introduction

1.1 Background and Motivation

Recommender systems have attracted significant attention from both academia and industry since the last decade. Various recommender systems have been developed to facilitate the matching between consumers (*i.e.*, users) with appropriate products or services (*i.e.*, items). Example items include songs, movies from content providers as well as books from E-commerce websites. Recently, recommender systems have also been applied to social network platforms (*e.g.*, Facebook, Twitter, Linkedin) for people recommendation, *e.g.*, recommending friends to a user or recommending who to follow. The prevalence of GPS-enabled devices (*e.g.*, smart phones) in the past few years further extends the landscape of recommender systems in location-based social networks (LBSN), exemplified by Foursquare¹ and Gowalla².

Depending on the application, different recommendation problems have been defined and studied. The top-N item recommendation and rating prediction are two most widely studied categories of recommendation problems. On one hand, top-N item recommendation tasks aim to recommend a user a list of items that she may be interested in. For example, in LBSN, *POI recommendation* aims to recommend unvisited POIs to users, where POI stands for point-of-interest, referring to a focused geographic entity such as district and street, or a specific point location such as landmarks and restaurants [1]. Rating prediction, on the other hand, is to predict the preference rating of a user to a product or service she has not rated before. The

¹<https://foursquare.com/>

²<http://gowalla.com/>

products or services with high predicted ratings are recommended to users. In this study, we are interested in the *business rating prediction* problem with business review data from Yelp.³

Yelp is a business review site and has attracted 47 million reviews to local businesses since 2004. Example businesses in Yelp include restaurants, shopping malls, beauty & spas, etc.. The website reports that it had an average of approximately 117 million monthly unique visitors in the third quarter of 2013.⁴ A Yelp user or Yelper can share her experience with a business by posting a review of the business and also a rating from 1 to 5 stars. Our main task here is to predict how many stars a user would give to a business that she has yet reviewed.

At first glance, rating prediction of business is the same as prediction of user's rating to any other kind of items (*e.g.*, a song or a movie), with the only difference that the item here refers to a business. The key difference, between a business and other kind of items that have been studied in literature, is that a business physically exists at a specific geo-location with latitude/longitude coordinates. More importantly, most businesses (*e.g.*, restaurants and shops) are not geographically isolated from each other. That is, when a user visits a business, there is a good chance that she walks by its neighbors if they are located within walking distance. The overall environment of that region might affect a user's view about a business and subsequently affect user's review and rating to the business. For example, the hygiene standard of a region might affect user's rating to many restaurants located in the region. On the other hand, a region distinguishes itself from other regions and becomes attractive to visitors often because there are a few good businesses located there. For business rating prediction, an interesting research problem here is: *Is there any influence from geographical neighbors to a business's rating prediction?* Further, are the factors (*e.g.*, business category, popularity, review report, and geographical distance) that have been studied in related problems also important factors in business rating prediction?

³<http://www.yelp.com/>

⁴<http://www.yelp.com/about>

1.2 Contributions

To answer the questions above, we first conducted data analysis on Yelp’s business rating data. We observe that there does exist weak positive correlation between the rating of a business and the rating of its neighbors, and this positive correlation in ratings is independent of the categories of the businesses and/or their neighbors. Based on these observations, we incorporate geographical neighborhood influence into our business rating prediction model which is based on the widely adopted latent factor model realized by matrix factorization. Specifically, for each business, we use two character latent factors to model its intrinsic and extrinsic characteristics respectively.

To be reviewed in Chapter 2 in detail, many factors like category, review content, popularity have been studied in related research problems. In Yelp, a business may belong to one or more categories and a review usually together with a rating. The category of a business usually reflects the characteristics of a business and users may use different criteria to evaluate businesses in different categories. The review report further elaborates the reason behind the rating and partially reflects the characteristics of the business. Regional popularity and geographical distance are two important features that have been studied in POI recommendation. We therefore also model the influences from these factors (*i.e.*, business category, user review, popularity and geo-distance) and study their effectiveness in business rating prediction in addition to geographical neighborhood influence. We show that the proposed model outperforms the state-of-the-art baselines, measured by both Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). To the best of our knowledge, this is the first study exploiting geographical neighborhood influence in business rating prediction. We note the different kinds of geographical influence considered here and that considered in POI recommendation and POI prediction. POI recommendation is to recommend unvisited POIs to a user, while POI prediction is to recommend POIs to a user based on her current context (e.g., time and current location) and the POIs to be predicted may or may not be visited. Both problems are top-N recommendation problems and the targeted POIs are unknown. In rating prediction, the task is to predict the rating of a user to a specific POI where the POI

is known. Thus we argue that both POI recommendation and POI prediction are different research problems from business rating prediction (see Chapter 2 for more details). To summarize, the main contributions arise from this study are as follows:

- We conduct data analysis and observe that there exists weak positive correlation between a business’s average rating and the average rating of its neighbors, and the positive correlation in ratings is independent of the categories of the businesses and/or their neighbors. This is an important observation that could be useful for not only business rating prediction, but also related studies, *e.g.*, POI recommendation and sentiment analysis of business reviews.
- We directly model the influence of geographical neighborhood into business rating prediction using matrix factorization. Specifically, for each business, we use two character latent factors to represent its intrinsic and extrinsic characteristics respectively. The influences from geographical neighbors of a business are then modeled as the linear combination of the latent factors for the extrinsic characteristics of its neighbors. In our recommender system, we have also considered other factors including business’s review, category, and popularity. To the best of our knowledge, this is the first study that models geographical neighborhood influence into business rating prediction.
- We conducted extensive experiments to evaluate the effectiveness of incorporating influence of geographical neighborhood and other factors in business rating prediction and compared the prediction accuracy of the proposed model with an array of strong baseline models.
- We further verify that the geographical distance between a user and a business adversely affects business rating prediction, although this is an important and effective factor in POI recommendation problems.

1.3 Thesis Organization

The rest of the thesis is organized as follows.

Chapter 2 In this chapter, we review the related work on recommendation system, mainly focusing on rating prediction, POI recommendation and POI prediction.

Chapter 3 We conduct data analysis and report important observations in this chapter.

Chapter 4 In this chapter, we present the details of the proposed business rating prediction model.

Chapter 5 We conduct experimental evaluation in this chapter.

Chapter 6 We conclude our research work in this chapter.

Chapter 2

Related Work

In this chapter, we review the recent advances in recommendation systems, mainly focusing on the applications of collaborative filtering on rating prediction. Because the major part of this study is the geographical neighborhood influence to rating prediction, we also review the related work in POI recommendation and POI prediction. We then highlight the major differences between our research problem and these two problems.

2.1 Recommendation Systems

Recommendation systems are essential tools that help users to deal with the information overload problem and provide personalized services. Basically, recommender systems provide suggestions of items to users. For example, suggesting online news for users to read, or suggesting music for users to listen. The “item” here is a general term according to what the system recommends to users, such as news or music. In general, for one specific type of item, recommendation systems capture users’ preference to items and generate customized recommendations. Recently, many studies have been done on recommender systems including music recommendation [2, 3], movie recommendation [4–6], news recommendation [7–9], App recommendation [10, 11], Q&A recommendation [12], tag recommendation [13–15], tour recommendation [16, 17], community recommendation [18, 19] and personalized tweet recommendation [20, 21].

Generally, there are two kinds of tasks in recommendation systems, namely top-N recommendation [22–26] and rating prediction [27–29]. In top-N recommendation,

the personalized recommendations to one particular user are offered as a ranked list of items. However, for rating prediction, instead of generating a ranked list, the recommender systems predict the ratings of a user to un-rated items. The solutions and evaluation methodologies for these two problems are different accordingly. In top-N item recommendation, the learning objective is to minimize the ranking error for pairs of items, and the evaluation methodologies are commonly based on accuracy metrics like precision and recall. However, the learning objective in rating prediction is to minimize the prediction error in terms of rating differences, and the evaluation methodologies are usually based on error metrics, such as the Mean Absolute Error (MAE) and the Root Mean Square Error (RMSE). Cremonesi *et al.* [24] evaluated the performance of several collaborative filtering algorithms in top-N recommendation task, and they found that recommendation algorithms optimized with MAE/RMSE do not absolutely perform as expected as in top-N recommendation task. In other words, there is no trivial relationship between error metrics and accuracy metrics. Nevertheless, there are some research work for connecting the rating prediction methods with top-N item recommendation. A possible approach to rely on the rating prediction for top-N item recommendation is to modify the objective function of Matrix Factorization. Existing study shows that such modified method can improve recall. Because our work mainly focuses on users' rating behavior, we do not evaluate the proposed method for top-N item recommendation.

There are many approaches for recommendation, such as Collaborative Filtering (CF), which has been widely used in both academia and industry. Latent factor models and neighborhood models are two common methods in CF. In our work, we adopt latent factor model for business rating prediction due to its high performance and good scalability reported in earlier studies. We will discuss more detail in Section 2.2. In addition, there are some other methods for recommendation, such as graph-based models [30–32], knowledge-based models [33–35], content-based models [36–38], trust-aware models [39, 40] and context-aware models [22, 41, 42].

In [31], Baluja *et al.* introduced a video co-view graph for YouTube¹, and then presented the *Adsorption* algorithm, a label propagation algorithm to provide personalized video recommendation. Specifically, in this work, the graph-based recommendation model is based on the ideas that a video is relevant to a user if the user

¹<http://www.youtube.com/>

and the video have a short path or several paths between them on the graph, or the paths between them avoid high-degree nodes. Inspired from the ideas in [31], Xiang *et al.* proposed a session-based temporal graph (STG) in [30], which models users' long-term and short-term preferences simultaneously. Specifically, STG is a directed bipartite graph containing three types of nodes, which are user node, item node and session node. The session node is a combined node which associated with a particular user at a specific time. Based on the graph, they adopted injected preference fusion algorithm to balance the impacts of users' long-term and short-term preferences and make temporal recommendation. Furthermore, they extended personalized Random Walk algorithm based on STG to support temporal recommendation.

Yildirim *et al.* [43] proposed an item-oriented model in which users make random walks according to the similarity of items. The normalized item-similarity matrix is interpreted as a transition probability matrix. The primary motivation behind this model is the success of link analysis algorithms like *PageRank* [44]. The final rating scores of a user for the items are not the actual predictions but are numerical values for sorting these items. In other words, the results identify the top-N recommendation items for each user. Moreover, the proposed algorithm significantly outperforms other top-N algorithm especially when data is sparse.

Linked Open Data (LOD) initiative gave birth to the open knowledge freely accessible on the Web. They provide valuable information for improving the quality of conventional recommendation systems. In [33], Ostuni *et al.* presented a semantic path-based ranking algorithm to compute top-N recommended items from users' implicit feedback. Specifically, they leverage a well-known encyclopedic knowledge base named DBpedia² to extract semantic path-based features. The path-based features are able to capture the complex interaction relationship between users, items and entities. Finally, the problem of generating top-N recommendations was formulated as a *learning to rank* problem, which is similar to the document ranking problem in Information Retrieval [45]. In [34], a semantic-based framework leveraging DBpedia was proposed for cross-domain recommendation.

In [35], Yu *et al.* proposed a generic personalized entity recommendation framework to utilize user click log and the extracted knowledge from Freebase³, a large

²<http://dbpedia.org>

³<http://www.freebase.com>

collaborative entity base. In the proposed framework, various representative features are generated from both user click log data and the entity knowledge base. By utilizing these pairwise features, the proposed models lead to a better recommendation performance.

Content-based recommendation systems recommend items similar to the items previously liked by a given user [36]. Particularly, candidate items are compared with those that users have rated before, and the best matching items are recommended. In practice, items are represented by a number of features, called *attributes* or *properties*. For example, in a book recommendation system, the described features can be *author*, *language* and *genres*. When each item is represented by the same attributes set, there are several ways to define the similarity between items, such as cosine similarity.

The trust-aware recommendation models [39, 46] are proposed to solve data sparsity problem in traditional collaborative filtering. The main idea is searching for trustable users over the trust network instead of searching for similar users as used in collaborative filtering, by exploiting trust propagation in the trust network. For one particular active user, the recommended items are the items appreciated by these trustable users.

Jamali *et al.* [47] proposed a random walk model, namely *TrustWalker*, which combined the trust-based and the item-based collaborative filtering methods to generate recommendation. Trust-based methods improve the coverage of recommendation. On the other hand, trust-based approaches may suffer from noisy data if going far from the origin user within the trust network. The proposed TrustWalker model finds a trade-off between item coverage and recommendation precision. Specifically, the model includes both pure trust-based and item-based collaborative filtering as its special cases. The TrustWalker model also allows users to compute a score to measure the recommendation confidence.

Traditional recommendation systems mainly focus on the entities of items and users, and do not take into consideration of the context information [41]. However, in many recommendation tasks, such as recommending a book, it is reasonable to incorporate the context information into recommendation process. Different

from the traditional two-dimensional (*user*, *item*) recommendation systems, context-aware models deal with the three-dimensional (*user*, *item*, *context*) recommendation process by incorporating available contextual information. The contextual information can be obtained explicitly or implicitly from the data or the environment. Early work in context-aware recommendation incorporated contextual information by pre-filtering, where contextual information controls data selection before building models, or post-filtering, where contextual information adjusts the recommendation results [41]. Recent work has focused on the problem of integrating contextual information into recommendation models directly, and two popular approaches have been proposed, one based on Factorization Machines (FM) [48] and the other on Tensor Factorization (TF) [22, 42]. Factorization Machines is a new model that combines the advantages of Support Vector Machines (SVM) with factorization models. Tensor Factorization is a generalized form of matrix factorization, proposed to deal with context-aware recommendation problems. We will discuss Tensor Factorization in Section 2.2.

2.2 Collaborative Filtering

As mentioned in Section 2.1, collaborative filtering (CF) is the most widely used technique in recommender systems [3, 5, 11, 49]. The underlying assumption of CF is that similar users would rate items similarly or similar items would receive similar ratings from users. Ratings here indicate users' preferences to items.

Collaborative filtering provides three key advantages to recommendation systems that are not provided by content-based models: (i) support for recommending items whose content is not easy to automatically analyze; (ii) the ability to filter items based on experience and quality; and (iii) the ability to provide serendipitous recommendations.

Firstly, in collaborative filtering, users determine the relevance and interest of an item in the systems. Consequently, filtering can be easily performed on items that are hard to understand or analyze by computers, such as music, feelings and ideas.

Secondly, beyond simple content, collaborative filtering can enhance filtering items by measuring how well items meet users' interests or requirement. Humans

are able to analyze items on the dimensions of quality or taste, which are very hard for computer to process. A content-based recommendation system could recommend all news articles related to sports. However, by combining with collaborative filtering together, a system could recommend only the sports news articles that are well written.

Lastly, a collaborative filtering system may sometimes provide serendipitous recommendations. That is, recommending items that are useful for the users, but do not contain users' expected content. In some cases, serendipitous recommendations are very valuable, such as movie recommendation and book recommendation.

According to [50], there are mainly two kinds of approaches for collaborative filtering, namely, *memory-based* CF and *model-based* CF.

The basic idea of memory-based CF is to find similar users or items by using similarity measures (*i.e.*, user-based CF [50, 51] and item-based CF [25, 52]). The commonly used similarity measures are cosine similarity and Pearson's correlation, computed based on user-item ratings as well as other user-/item-specific features depending on the application. The similar users or items are also known as neighbors and therefore memory-based CF is also known as neighborhood-based CF. However "neighbor" in this context is defined by the chosen similarity measure. Therefore, different similarity measures lead to different sets of neighbors for a given user or item.

Although memory-based CF has been proposed for several years, there are still many research work on this topic. Niemann *et al.* in [53] presented a new collaborative filtering method, namely usage context-based collaborative filtering (UC-BCF). This method is proposed to solve the "long tail" problem in recommendation systems, by increasing the aggregate diversity of recommender system. The basic idea is taken from linguistics where relations between words can be inferred from the word's contexts. Specifically, the model takes the user-item matrix as input, and transforms the matrix into an item-item matrix, in which each item is described by top-K co-occurring items in user profile. In this way, the UC-BCF model combines the strength of traditional user-based and item-based CF. Moreover, this model can also capture the similar item vectors by identifying two items that are rated dissimilarly by amount of dissimilar users.

In [29], Desarkar *et al.* proposed a memory-based algorithm for collaborative filtering. Generally, there are many items that have very few ratings from users in real world and traditional similarity measures based on ratings have poor performance in rating prediction. The proposed method assigns similarity weights based on preference relations between items instead of the absolute ratings, and then using these weights to generate an aggregate preference graph for users. Finally, the unknown ratings are inferred from the graph. Experimental results show that the proposed method outperforms other methods in sparse dataset.

However, memory-based CF has some limitations. Because the similarity values in the model are based on common items, therefore they are unreliable when the rating data is sparse. To overcome the shortcomings of memory-based CF and to achieve better recommendation performance, model-based CF [54, 55] models have been proposed. Model-based CF usually employs machine learning techniques to find patterns and build models from the observed user-item rating data, and then to predict the unobserved ratings. There are many model-based CF algorithms, including Bayesian networks [55–57], clustering CF models [58–60], decision trees [61, 62], latent factor models [63] such as singular value decomposition, probabilistic latent semantic analysis [54], and markov decision process based models [64].

In [58], Luo *et al.* extended the existing clustering algorithm for CF, and proposed a flexible mixture model (FMM) to cluster items and users at the same time. This model allows each item and user to belong to multiple clusters, and groups users based on users preference patterns instead of rating patterns. The experimental results indicated that the FMM algorithm outperforms the memory-based CF algorithms.

Liu *et al.* [62] proposed a novel recommendation model named SoCo to incorporate different types of contextual and social network information. Specifically, the model applies random decision trees algorithm to partition the original user-item rating matrix, by taking into account various contextual information. The final prediction is generated by combining the predictions from all the decision trees.

Among all the model-based CF algorithms, latent factor models are probably the most successful model-based CF models. Latent factor CF model introduces latent class variables to discover users’ interest profiles and items’ prototypical profiles. In

the model, users and items are jointly mapped into a shared latent space of much lower dimensionality. In [54], Hofmann proposed a statistical latent class model for CF. The model introduces hidden variables with different states for each user-item pair, and then user and item are rendered conditionally independent. The parameters are learned by using *expectation maximization* (EM) algorithm, which is similar to the probabilistic Latent Semantic Analysis (pLSA) [65] in Information Retrieval.

As the most successful realization of latent factor model, matrix factorization (MF) [5, 6, 63] is an efficient and popular approach for collaborative filtering. The basic premise behind matrix factorization is that the preferences are influenced by only a small number of factors, and for a particular user, each factor has different weight to determine his overall preference. Practically, by factorizing the original user-item rating matrix into user-specific and item-specific matrix, we can then utilize the two factorized low rank matrices to make further prediction [63, 66]. Recently, MF has been successfully applied to various recommender systems including music rating prediction for Yahoo! Music [3] and Last.fm [2], movie rating prediction for Netflix [4, 5] and Douban [2, 6], and personalized tweet recommendation [20].

Koren [63] gave a very detail introduction about MF. Through matrix factorization, each user and each item is associated with a f -dimensional vector, where $f \ll \min(m, n)$. The inner product of a user vector $\mathbf{p}_u \in \mathbb{R}^{f \times 1}$ and an item vector $\mathbf{q}_i \in \mathbb{R}^{f \times 1}$ is used to approximate the user's preference to the item. Specifically, $\mathbf{p}_u$ and $\mathbf{q}_i$ can be learned from the user-item rating matrix with known ratings. Basically, there are two common approaches to learn the parameters, which are stochastic gradient descent and alternating least squares (ALS) [63].

In dealing with various data aspects and different application requirements, several MF models have been proposed, with increasing complexity and accuracy. For example, users may have certain degree of biases, some users are more lenient and some are very strict about ratings. Similarly, items may also have some degree of biases because of location or branding. To achieve more accurate rating prediction, Biased MF extends the basic matrix factorization by considering the biases factors, which is one of the most widely used baseline methods in recommendation systems.

In [4], a time-aware factor model was proposed to capture the temporal changes in customer preferences. Because the user preferences, user bias and item bias may change over time, we therefore treat different temporal aspects in distinct portions. Specifically, we split the item bias into a set of time-based bins, and model the item bias as a real valued time varying function. At the same time, a linear function is proposed to capture drift of user biases and user preferences. The final factor model is the combination of all the previous components, parameters are learned by using stochastic gradient descent algorithm.

In [5], a successful and widely used MF approach was proposed, namely SVD++, which integrated the MF model and neighborhood-based CF for movie rating prediction. In this integrated model, MF model profiles both users and items, and the neighborhood model captures similarities between users or items. With the integration, the resultant model describes general properties of users and/or items and the user-item interactions with fine-grained adjustments by neighborhood.

MF are flexible and easy to model the influence from other aspects of users or items besides the ratings. For example, Koenigstein *et al.* [3] incorporated rich item bias into MF model to capture taxonomy information of music. Each music has multiple types of information such as track, album, artist and genre. In their proposed model, MF was extended by adding shared bias parameters for items linked by common taxonomy. Moreover, some other work has also shown that popularity is helpful in improving the recommendation accuracy [24, 67].

Recently, with the rapid development of social networks, many research work have been done in recommendation in this area. Ma *et al.* [6] proposed a matrix factorization method to improve the recommendation quality by incorporating social network information. Specifically, the proposed method modeled the social network information as regularization terms, by constraining the matrix factorization objective function. Moreover, the social network information not only means the implicit social relationship between users but the implicit social information that identified by similarity metrics (such as cosine similarity or Pearson's correlation). Chen *et al.* in [20] proposed topic-level latent factors to capture users' interests over tweet content. The model decomposes the tweet latent factors into words latent factors. In other words, instead of directly asking whether a user is interested in a tweet,

the model captures user’s preference over the words in the tweet. Besides the latent factors, the model incorporates explicit features by a linear combination. This model can partially solve the cold-start problem in recommendation.

Ning *et al.* in [68] proposed a set of sparse linear methods with side information to generate recommendation. By involving a regularized optimization process, they learned coefficient matrix based on both user-item rating and item review information, and got better performance.

Contextual information has been recognized as an important factor for recommendation systems. However, most CF methods such as MF do not provide a straightforward way to integrate contextual information. Recently, Tensor Factorization has been introduced to model the data as a user-item-context N-dimensional tensor, and basically, the way of modeling the context variables is the same as the users and items are modeled in MF techniques. In [49], Moghaddam *et al.* proposed a tensor factorization model to personalize review quality prediction. Specifically, the three-dimensional tensor $rater \times reviewer \times product$ is factorized into three matrices and one central tensor. In addition, a extended tensor factorization model was proposed to capture the overall rating of products. All the feature matrices and central tensor are learned by using stochastic gradient descent algorithm as the same processes in MF.

2.3 POI Recommendation and Prediction

POI recommendation has attracted significant attention recently with the popularity of LBSN [69–72]. Because geographical neighborhood is the main focus in our study, there is a need to review recent advances in POI recommendation. The main research issue in POI recommendation is to accurately recommend unvisited POIs to users. Geographical influence and temporal influence are major considerations in POI recommendation [69, 70]. Geographical influence is based on the observation that users tend to visit nearby POIs of their home or office locations, as well as nearby locations of the POI in their favor. Temporal influence is reflected by the observations that users check-in different types of POIs at different time slots of a day (*e.g.*, restaurant in lunch hour and bar in night). POI recommendation has also considered social influence among friends [70].

In [70], Ye *et al.* proposed a unified recommendation framework for POI recommendation by incorporating three complementary factors, which are user preference of POIs, social influence and geographical influence. Specifically, the user preference and social influence obtained from the check-in activities on POIs. In order to capture the geographical influence, a spatial analysis was performed on check-in data and indicated a *geographical clustering phenomenon* in user check-in behavior. Finally, a power-law probabilistic model was proposed to model the geographical influence between POIs. Additional findings based on the results shown that Random Walk and Restart (graph-based) may not be suitable for POI recommendation in LBSNs. Cheng *et al.* [72] proposed a novel fused matrix factorization framework to provide POI recommendation. Specifically, they found users tend to check in around several centers, and modeled the geographical influence of users' check-in behavior as a Multi-center Gaussian Model (MGM).

Yuan *et al.* [69] believe users' activities are often influenced by time and defined a new problem, namely *time-aware POI recommendation*. They analyzed human temporal behavior and spatial behavior from users' check-in history, and developed two POI recommendation methods that incorporate temporal influence and spatial influence respectively. Specifically, a user-based CF was presented to utilize the temporal influence, and a power law distribution was used to model the spatial influence to users' movements. Finally, a unified framework was proposed that fuse the spatial and temporal influences to make the time-aware POI recommendation.

POI prediction aims to predict which POI a user would visit next given her current location and time [71, 73]. Different from POI recommendation, the POI predicted for a user to be visited next may have been visited by the user before. Similarly, both geographical influence and temporal influence have been considered in the modeling. Liu *et al.* in [74] proposed a category-aware POI prediction model by exploiting users' preference transition over location categories. The proposed model consists of two main steps. The first step is preference transition prediction. Specifically, in order to leverage other similar users' preference transition patterns, similar users are clustered based on their activities patterns. In each cluster, collaborative filtering is applied to predict a user's preference transition, by generating top-K1 most likely categories that this user would follow. In step two, in order to

predict users' preference on POI in K1 categories, two factors are considered, which is user's preference on POIs and geographical influence that influence users' check-in behavior. Finally, matrix factorization is employed to predict a user's preference transition over those top-K1 categories. Under each category, top-K2 POIs will be chosen to construct the final recommendation list. The model was evaluated on Gowalla check-in data with three-level category hierarchy.

Noulas *et al.* [75] studied the problem of predicting the next place a mobile user will visit. They proposed a set of 12 features to capture the factors that may drive users' movement, including global features and temporal features, as well as spatial features of user check-in patterns. Moreover, the prediction task was modeled as a binary classification problem. Baraglia *et al.* in [16] studied the tourist movement prediction problem. Based on users' visit history from Flickr⁴, a photo sharing portal, they defined a set of 68 features to capture both the tourist behavior and the characteristics of candidate POIs. Finally, the prediction problem was modeled as a *learning to rank* task by using start-of-the-art learning methods.

In this work, we study the impact of geographical neighborhood influence to business rating prediction. We argue that the geographical neighborhood influence in our research problem is largely different from the geographical influence in POI recommendation and/or prediction. For POI recommendation/prediction, the geographical influence is more related to the cost of travel (*e.g.*, time cost or monetary cost) from a user's point of view. For rating prediction, on the other hand, we predict a user's degree of preference to a yet reviewed business. The cost of travel here is expected to be a less important factor. The contextual factors (*e.g.*, the current location, the time of visiting) which have demonstrated effectiveness in POI recommendation/prediction are also less relevant to our research problem. In our research problem, geographical neighborhood influence is more related to the business environment created by the surrounding businesses to one business located at a geographical region. Next, we conduct data analysis of the Yelp data.

⁴<http://www.flickr.com>

Chapter 3

Data Analysis on Yelp Dataset

In this chapter, we describe preliminaries of our work and then conduct data analysis on Yelp dataset.

3.1 Pearson Correlation

The Pearson's correlation coefficient is a measure of the linear correlation between two variables X and Y . Generally, Pearson's correlation is defined as the covariance of the two variables divided by the product of their standard deviations. The value is between +1 and -1 inclusive, where 1 means the two variables are total positive correlation, while 0 is no correlation, and -1 is total negative correlation.

Specifically, the formulation of Pearson's correlation coefficient is:

$$\gamma = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^n (X_i - \bar{X})^2} \sqrt{\sum_{i=1}^n (Y_i - \bar{Y})^2}} \quad (3.1)$$

where X_i, Y_i are the paired sample data, and $\bar{X}, \bar{Y}$ are the mean value of X and Y , respectively.

Pearson's correlation is one of the most popular similarity measures in collaborative filtering. In next Section, we perform Pearson's correlation coefficient to measure the correlation between the rating of businesses and its neighbors.

Table 3.1: Number of reviews/ratings per user/business

Statistics	User	Business
Minimum number of reviews/ratings	1	3
Maximum number of reviews/ratings	473	540
Average number of reviews/ratings	5.00	19.93

3.2 Data Analysis

3.2.1 Dataset

Our study is based on the recently released Yelp Dataset Challenge.¹ The data is sampled by Yelp from the greater Phoenix, AZ metropolitan area from March 2005 to January 2013. It contains 11,537 businesses, 229,907 reviews by 43,873 users, and 8,282 check-in sets. This dataset was used as training data in ACM RecSys Challenge 2013.²

A *business* has a unique id, name, address, latitude longitude, its categories and some other attributes like city, state, and neighborhoods. However, we observe that the neighborhoods attribute is empty in the dataset. A *review* contains business id, user id, rating from 1 to 5 stars, date, review text, and voting. In this study, we do not use the voting feature for its less relevance to our research problem. Table 3.1 reports the minimum, maximum, and average number of ratings per user and per business respectively. The number of reviews having 1 to 5 stars are: 17516 (1-star), 20957, 35363, 79878, and 76193 (5-star) respectively. A *check-in set* for a business contains the aggregated number of check-ins in every hour from Monday to Sunday. The largest number of check-ins to one business observed from the data is 22,977. Note that, the dataset does not provide detailed check-in records of which user checkin a business at what time, and there are only about 72% of businesses have check-in sets. The percentage becomes 85.13% among all businesses having at least 5 reviews. In other words, a business may have a number of reviews but no check-in records. The dataset does not provide much details of a *user*.

¹http://www.yelp.com/dataset_challenge

²<http://recsys.acm.org/recsys13/recsys-2013-challenge/>

Figure 3.1: Percentage of businesses having at least 1, 3, 6, 10 neighbors within a distance threshold from 20 to 2000 meters

3.2.2 Observations

Tobler’s First Law of Geography states “Everything is related to everything else, but near things are more related than distant things” [76]. Next, we present 3 observations made from the data with respect to geographical neighborhood.

Observation 3.1 *Most businesses have **neighbors within a short geographical distance** from their locations. More than 44% businesses have one neighbor next to it within 20 meters and 95% of businesses have one neighbor within 500 meters.*

Based on the business latitudes and longitudes, we calculate the geographical distance between two businesses using Haversine formula.³ Figure 3.1 plots the percentage of businesses having at least 1, 3, 6, and 10 neighbors respectively, within a geographical distance threshold ranging from 20 to 2000 meters. It shows that most businesses are not isolated geographically from others. Specifically, 44.3% of businesses have one neighbor next to it within 20 meters. The percentage rises to 95.6% if the distance threshold is set to 500 meters, a distance for a 6-minute walk.⁴ Within this walking distance, about 80% of businesses have at least 6 neighbors, and 66% have at least 10 neighbors.

³http://en.wikipedia.org/wiki/Haversine_formula

⁴Human preferred walking speed is about 1.4 m/s or 5km/h.
http://en.wikipedia.org/wiki/Preferred_walking_speed

Figure 3.2: Pearson’s correlation between a business’s rating and the average rating of its {1, 3, 6, 10} nearest neighbors (NN), within a distance threshold.

Observation 3.2 *The average rating of a business is **weakly positively correlated** with the average rating of its neighbors.*

Before exploring a new place, we often receive advice on which region is famous for good food (*e.g.*, a few famous restaurants collocating in that region), or advice on which region to avoid for shopping because the shops there are infamous for cheating. This phenomenon of “things of one kind come together” is also reflected from the positive correlation between the ratings of businesses and their geographical neighbors.

We calculate the average rating of a business from all its user reviews. Hereafter, for easy presentation, we refer this average rating simply as *the rating of a business* or *a business’s rating* when the context is clear. Figure 3.2 plots the Pearson’s correlation coefficient between a business’s rating and the average rating of its 1, 3, 6, and 10 nearest neighbors, at different distance thresholds from 20 to 2000 meters. For a given distance threshold, if a business has no neighbors within the distance, then this business is excluded from the computation. If a business has fewer than the number of nearest neighbors specified (*e.g.*, 10NN) within the distance threshold, then the ratings of all its nearest neighbors within the distance threshold are averaged. For reference, for each business participated in the computation within

a distance threshold, we also randomly sample a business from the dataset (regardless of the distance between the two) and then compute the Pearson’s correlation between the two sets of ratings, labeled “Random” in the plot.

From Figure 3.2, we observe that the rating of a business is weakly positively correlated with the average rating of its nearest neighbor(s). Pearson’s correlation coefficient is in the range of 0.109 to 0.173. The correlation is relatively stronger within a smaller distance (*e.g.*, 20 or 50 meters) and becomes stable for distance threshold of 500 meters or larger. In comparison, correlation coefficient between the ratings of the same set of businesses and their randomly selected counterparts is in the range of -0.002 to 0.002, *i.e.*, no correlation, as expected.

We have two interpretations of the weak positive correlation between ratings of businesses and their neighbors. First, the rating of a business should *mainly* depend on the intrinsic characteristics of the business (*e.g.*, quality of products/services), not its neighbors. Second, a business is not geographically independent from its neighbors. When a user visits one business, she has at least glanced over its neighbors if not visited them. These neighbors give her the sense of the surrounding environment of the business (*e.g.*, hygiene standard). Such extrinsic characteristics may affect her view of the business, leading to the weak positive correlation. Based on the weak positive correlation, we expect more accurate business rating prediction can be achieved by considering the influence of the geographical neighbors of a business.

Observation 3.3 *The weak positive correlation in ratings is **independent of the categories** of the businesses and/or their neighbors.*

A business in Yelp is assigned one or more category labels (*e.g.*, restaurant, shopping). Plotted in Figure 3.3, the number of businesses in categories demonstrate a power-law like distribution, with a few categories each containing a large number of businesses and many small categories each has only one or two businesses. **Restaurants**, the largest category, covers nearly 40% of the 11,537 businesses in the dataset. Next, we study the 5 largest categories as representative examples. They are: **Restaurants** (4,503), **Shopping** (1,681), **Food** (1,616), **Beauty & Spas** (764), and

Figure 3.3: Business category distribution

Nightlife (640), where the numbers in parentheses indicate the number of businesses in each category.

For each of the 5 categories, we calculate the rating correlation between the businesses in the category and their neighbors, regardless of the categories of their neighbors because the neighbors here are defined geographically. The category-wise rating correlations with 1NN and 6NN are plotted in Figures 3.4.a and 3.4.b respectively. The lines labeled “random”, plotted for reference purpose, report the correlation between the ratings of businesses in any of the 5 categories and randomly selected businesses from the dataset (regardless of their categories and geographical distance). Shown in Figure 3.4.a, businesses in **Beauty & Spas** demonstrate slightly stronger rating correlation with their 1NN compared with businesses in **Nightlife** and **Shopping**. When the number of nearest neighbors is enlarged to 6, this pattern becomes less apparent, particularly when the distance threshold reaches 500 meters, plotted in Figure 3.4.b. In both Figures 3.4.a and 3.4.b, weak positive correlation is observed for ratings of businesses in all 5 categories across all distance thresholds.⁵

Figures 3.5.a and 3.5.b show the percentage of the nearest neighbors that are also in the same category. The two figures show very similar patterns: (i) Restaurants are more likely collocated with restaurants with about 50% chance that the neighbors

⁵Similar observation also holds for 3NN and 10NN. The figures are not plotted due to page limit.

of a restaurant are also restaurants; (ii) Beauty & Spas on the other hand is much more distinctive from their neighbors. Below 20% of their neighbors are also in the same category. With respect to Figures 3.4.a and 3.4.b, we argue that weak positive rating correlation between businesses and their neighbors is independent of whether the neighbors are in the same category.

In summary, the three observations made from the data suggest that geographical neighborhood has influences on the rating of a business. Next, we incorporate geographical neighborhood influence into business rating prediction.

3.4.a: Rating correlation of 1NN

3.4.b: Rating correlation of 6NN

Figure 3.4: Category-wise rating correlation with nearest neighbors (1NN, 6NN)

3.5.a: Percentage of 1NN being in the same category

3.5.b: Percentage of 6NN being in the same category

Figure 3.5: The percentage of nearest neighbors (1NN, 6NN) in the same category

Chapter 4

Business Rating Prediction

The problem of rating prediction has been well formulated in literature. In our setting, we use r_{ui} to denote the review rating that user u gives to item i (*i.e.*, a business). r_{ui} is in the range of 1 to 5 stars with more stars indicating higher preference. Given the exiting ratings given by m users to n items, the task is to predict the unknown rating $\hat{r}_{ui}$, if user u has not rated item i before. In the following, we first briefly introduce matrix factorization and then present our proposed model by incorporating various influences in the prediction. Table 4.1 lists the notations used in this paper.

4.1 Matrix Factorization

Our proposed method is based on the latent factor model realized by matrix factorization. Through matrix factorization, each user and each item is associated with a f -dimensional vector, where $f \ll \min(m, n)$. The inner product of a user vector $\mathbf{p}_u \in \mathbb{R}^{f \times 1}$ and an item vector $\mathbf{q}_i \in \mathbb{R}^{f \times 1}$ is used to approximate the user's preference to the item (see [4] for a detailed introduction of matrix factorization). Accordingly, the predicted rating of user u to item i is computed using

$$\hat{r}_{ui} = \mathbf{p}_u^\top \mathbf{q}_i, \quad (4.1)$$

where $\mathbf{p}_u$ and $\mathbf{q}_i$ can be learned from the user-item rating matrix with known ratings. However, users may have certain degree of biases, some users are more lenient and some are very strict about ratings. Similarly, items may also have some degree of

Table 4.1: Notations and semantics

N_i	Set of geographical neighbors of item i
C_i	Set of categories item i belongs to
R_i	Set of words in item i 's review
$\mathcal{K}$	Set of (u, i) pairs with known r_{ui} ratings
$r_{ui}, \hat{r}_{ui}$	Observed and predicted ratings of user u to item i
μ	Mean of all known r_{ui} ratings
b_u, b_i	Bias parameters for user u and item i , respectively
$\mathbf{p}_u$	Latent factors of user u
$\mathbf{q}_i$	Latent factors of item i for its intrinsic characteristics
$\mathbf{v}_i$	Latent factors of item i for its extrinsic characteristics
$\mathbf{d}_c$	Latent factors of category c
$\mathbf{q}_w$	Latent factors of review word w
ρ_i	Normalized popularity of item i
β_i	Popularity weighting parameter for item i
$\tau_{u,i}$	Normalized geo-distance between user u and item i
β_u	Geo-distance weighting parameter for user u

biases because of location or branding. To achieve more accurate rating prediction, Biased MF extends the basic matrix factorization by considering the biases as follows

$$\hat{r}_{ui} = \mu + b_u + b_i + \mathbf{p}_u^\top \mathbf{q}_i, \quad (4.2)$$

where μ is the average rating of all known ratings, b_u and b_i are the user bias and item bias, respectively. Learning the unknown parameters $\mathbf{p}_u$, $\mathbf{q}_i$, b_u , and b_i is an optimization problem to minimize the regularized squared error on the set of known ratings $\mathcal{K}$, shown in Equation 4.3.

$$\min_{\mathbf{p}_*, \mathbf{q}_*, b_*} \sum_{(u,i) \in \mathcal{K}} (r_{ui} - \hat{r}_{ui})^2 + \lambda_1 (\|\mathbf{p}_u\|^2 + \|\mathbf{q}_i\|^2) + \lambda_2 (b_u^2 + b_i^2) \quad (4.3)$$

In Equation 4.3, λ_1 and λ_2 are regularization parameters used to avoid overfitting. Both stochastic gradient descent (SGD) and alternating least squares (ALS) algorithms can be used to solve the optimization function and learn the parameters [4, 5]. Relatively, SGD is easier to implement and faster than ALS in single machine setting. ALS, on the other hand, is more favorable in parallel settings, because in ALS the latent factor for each item and each user is computed independently. In this paper, we adopt SGD to learn the parameters following the algorithm presented in [5].

4.2 Incorporating Neighborhood Influence

Based on our observations in Section 3.2.2, most businesses have neighbors within a short geographical distance, and more importantly, the rating of a business is weakly positively correlated with the rating of its neighbors. These observations suggest that considering the geographical neighborhood influence may improve the accuracy of business rating prediction.

In this paper, to model users' rating behavior on businesses, we first assume that a user's rating to a given business is determined by its *intrinsic characteristics* and the *extrinsic characteristics* of its geographical neighbors. For a business i , we use $\mathbf{q}_i$ and $\mathbf{v}_i$ to model its intrinsic and extrinsic characteristics, respectively. More specifically, $\mathbf{q}_i$ models the intrinsic characteristics of a business (*e.g.*, taste of food and quality of service) observable by users who have interacted with the business. $\mathbf{v}_i$ models the extrinsic characteristics of a business (*e.g.*, hygiene standard) in influencing its geographical neighbors observable by the "pass-by" visitors.

Let N_i be the set of geographical neighbors of a business i , satisfying certain selection criteria (*e.g.*, the top-6 nearest neighbors). Let $n \in N_i$ be a neighbor of business i . Incorporated with influence from geographical neighbors, the predicted rating is now computed with both $\mathbf{q}_i$ and $\mathbf{v}_n$'s, shown in Equation 4.4.

$$\hat{r}_{ui} = \mu + b_u + b_i + \mathbf{p}_u^\top \left(\mathbf{q}_i + \frac{\alpha_1}{|N_i|} \sum_{n \in N_i} \mathbf{v}_n \right) \quad (4.4)$$

In above equation, parameter $\alpha_1 \in [0, 1]$ controls the importance of geographical neighborhood influence in business rating prediction, $|\bullet|$ denotes the cardinality of a set. Accordingly, the objective function is updated with regularization components for $\mathbf{v}_n$, shown in Equation 4.5, where λ_3 is the newly introduced regularization parameter, similar to λ_1 and λ_2 .

$$\min_{\mathbf{p}_*, \mathbf{q}_*, b_*, \mathbf{v}_*} \sum_{(u,i) \in \mathcal{K}} (r_{ui} - \hat{r}_{ui})^2 + \lambda_1 (\|\mathbf{p}_u\|^2 + \|\mathbf{q}_i\|^2) + \lambda_2 (b_u^2 + b_i^2) + \lambda_3 \sum_{n \in N_i} \|\mathbf{v}_n\|^2 \quad (4.5)$$

Note that, by considering $\mathbf{v}_n$'s in the prediction, Equation 4.4 partially addresses the cold-start problem for newly established businesses that do not appear in training

data. Although $\mathbf{q}_i$ of a new business is empty, the $\mathbf{v}_n$'s of its geographical neighbors are not empty if they appear in training data. Therefore Equation 4.4 can be applied to make rating prediction of an existing user to a newly established business by using $\mathbf{p}_u$, b_u and $\mathbf{v}_n$'s.

4.3 Incorporating Category Influence

Analyzed in Section 3.2.2, a business in Yelp may belong to one or more categories. The category of a business usually reflects the characteristics of a business, *e.g.*, product/service offered by the business or the way the business is conducted. Intuitively, users may use different criteria to evaluate businesses in different categories. For example, the criteria commonly used for reviewing restaurant (*e.g.*, taste of food) cannot be used to review businesses in beauty & spas category. Moreover, a recent study has also shown that POI recommendation achieves better accuracy by considering the categories of the POIs [74].

In our model, we introduce category latent factors to exploit business categories for more accurate business rating prediction. For a business category c , it is associated with a latent vector $\mathbf{d}_c \in \mathbb{R}^{f \times 1}$. Let C_i be the set of categories a business i belongs to. By incorporating the category influence, the predicted rating $\hat{r}_{ui}$ is now defined in Equation 4.6, where $\alpha_2 \in [0, 1]$ is a parameter that controls the importance of category influence in rating prediction.

$$\hat{r}_{ui} = \mu + b_u + b_i + \mathbf{p}_u^\top \left(\mathbf{q}_i + \frac{\alpha_1}{|N_i|} \sum_{n \in N_i} \mathbf{v}_n + \frac{\alpha_2}{|C_i|} \sum_{c \in C_i} \mathbf{d}_c \right) \quad (4.6)$$

Accordingly, the objective function is updated with regularization components for $\mathbf{d}_c$, shown in Equation 4.7.

$$\begin{aligned} \min_{\mathbf{p}_*, \mathbf{q}_*, b_*, \mathbf{v}_*, \mathbf{d}_*} \sum_{(u,i) \in \mathcal{K}} (r_{ui} - \hat{r}_{ui})^2 + \lambda_1 (\|\mathbf{p}_u\|^2 + \|\mathbf{q}_i\|^2) \\ + \lambda_2 (b_u^2 + b_i^2) + \lambda_3 \left(\sum_{n \in N_i} \|\mathbf{v}_n\|^2 + \sum_{c \in C_i} \|\mathbf{d}_c\|^2 \right) \end{aligned} \quad (4.7)$$

4.4 Incorporating Review Content

In Yelp, when giving a rating to a business from 1 to 5 stars, the user usually writes a review. Typically, the review elaborates the reason behind the rating and partially reflects the characteristics of the business. Collectively, words in all reviews to a business provide a much better description about the business than the learned vector of “latent” factors. However, in order to make use of the review words in the prediction model, the review words have to be mapped to the same f -dimensional vector space. Similar to the topic level decomposition of tweets for tweet recommendation in [20], we decompose the latent factors of a business $\mathbf{q}_i$ into a combination of latent factors of review words. Let R_i be the set of words that appear in business i ’s review, the decomposition is shown as follows, where $\mathbf{q}_w$ denotes the latent factors of word w .

$$\mathbf{q}_i \Rightarrow \frac{1}{|R_i|} \sum_{w \in R_i} \mathbf{q}_w \quad (4.8)$$

With the decomposition, the predicted rating is now shown in Equation 4.9.

$$\begin{aligned} \hat{r}_{ui} = & \mu + b_u + b_i \\ & + \mathbf{p}_u^\top \left(\frac{1}{|R_i|} \sum_{w \in R_i} \mathbf{q}_w + \frac{\alpha_1}{|N_i|} \sum_{n \in N_i} \mathbf{v}_n + \frac{\alpha_2}{|C_i|} \sum_{c \in C_i} \mathbf{d}_c \right) \end{aligned} \quad (4.9)$$

Again, the objective function is updated with regularization components for $\mathbf{q}_w$, shown in Equation 4.10.

$$\begin{aligned} \min_{\mathbf{p}_*, \mathbf{q}_*, b_*, \mathbf{v}_*, \mathbf{d}_*} \sum_{(u,i) \in \mathcal{K}} (r_{ui} - \hat{r}_{ui})^2 + & \lambda_1 \left(\|\mathbf{p}_u\|^2 + \sum_{w \in R_i} \|\mathbf{q}_w\|^2 \right) \\ & + \lambda_2 (b_u^2 + b_i^2) + \lambda_3 \left(\sum_{n \in N_i} \|\mathbf{v}_n\|^2 + \sum_{c \in C_i} \|\mathbf{d}_c\|^2 \right) \end{aligned} \quad (4.10)$$

4.5 Popularity and Geo-distance Influences

In this Section, we discuss two features that have been studied in POI recommendation, namely *popularity* and *geographical distance*. In Chapter 1, we argue that the key difference between a business with other kind of items that have been studied in rating prediction is that a business physically exists at a geographical location.

Although we argue that business rating prediction is different from POI recommendation, it remains interesting to investigate whether the features explored in POI recommendation are useful here.

Regional popularity is mentioned as one of the unique characteristics in LBSNs which distinguish POI recommendation from other recommendation tasks [77]. For example, businesses located in downtown area are likely to receive more visits than those in suburban. In the Yelp dataset, a business has check-in sets and reviews, both are indicators of business popularity. Note that, a user may write one review to a business but may check-in at the business multiple times. However, the Yelp dataset only provides the aggregated check-in numbers on hourly basis from Monday to Sunday but not user-specific check-ins. Some of the businesses in the dataset do not have check-in sets. We therefore simply sum up the number of reviews and the number of check-ins of a business to be its popularity.

In POI recommendation, geographical distance is a major consideration because users tend to visit nearby POIs [69, 70]. With users' review data, we estimate a user's 'home location' by using the recursive grid search algorithm proposed in [78]. Then the geographical distance between a user and an unrated business can be easily calculated based on the estimated user location and the location of the business.

For simplicity, we model both popularity and geographical distance as a rating bias z .

$$z = \beta_i \rho_i + \beta_u \tau_{u,i} \quad (4.11)$$

In above equation, ρ_i is the normalized popularity of item i by taking the common logarithm of its raw popularity value; $\tau_{u,i}$ is the normalized distance between user u and business i by taking the common logarithm of the geographical distance. The two parameters β_i and β_u are the weighting parameters for business i and user u respectively; both are learned from the training data. With rating bias z , the predicted rating is shown in Equation 4.12.

$$\begin{aligned} \hat{r}_{ui} = & \mu + b_u + b_i + z \\ & + \mathbf{p}_u^\top \left(\frac{1}{|R_i|} \sum_{w \in R_i} \mathbf{q}_w + \frac{\alpha_1}{|N_i|} \sum_{n \in N_i} \mathbf{v}_n + \frac{\alpha_2}{|C_i|} \sum_{c \in C_i} \mathbf{d}_c \right) \end{aligned} \quad (4.12)$$

Note that, we do not replace the user bias b_u and item bias b_i by rating bias z . The reason is that the rating bias $\beta_i \rho_i + \beta_u \tau_{u,i}$ captures biases specific to popularity

and geographical distance respectively, while b_u and b_i are used to capture bias from all unknown factors. The objective function considering both popularity and geographical distance is shown in Equation 4.13.

$$\begin{aligned}
 \min_{\mathbf{p}_*, \mathbf{q}_*, b_*, \mathbf{v}_*, \mathbf{d}_*, \beta_*} \sum_{(u,i) \in \mathcal{K}} (r_{ui} - \hat{r}_{ui})^2 &+ \lambda_1 \left(\|\mathbf{p}_u\|^2 + \sum_{w \in R_i} \|\mathbf{q}_w\|^2 \right) \\
 &+ \lambda_2 (b_u^2 + b_i^2 + \beta_i^2 + \beta_u^2) + \lambda_3 \left(\sum_{n \in N_i} \|\mathbf{v}_n\|^2 + \sum_{c \in C_i} \|\mathbf{d}_c\|^2 \right) \quad (4.13)
 \end{aligned}$$

Chapter 5

Experiment

In this chapter, we conduct experiment on Yelp dataset to evaluate the performance of different methods in business rating prediction.

5.1 Experimental Setting

Dataset We use the Yelp dataset that has been studied in Section 3.2.1 in our experiments. The pre-processing of the dataset in our experiments include removal of businesses and users having fewer than 10 reviews and stop words removal and stemming in reviews. After the preprocessing, we have 113,514 ratings by 3,965 users to 3,760 businesses. For each user, we sort her ratings in chronological order. The first 70% of ratings are used for training, and the remaining 30% for testing. As the result, we have 79,309 ratings to build the matrix factorization model for the prediction of the remaining 34,205 ratings. The data sparsity is 99.47%.

Evaluation Metric We adopt two popular evaluation metrics, namely, *Mean Absolute Error* (MAE) and *Root Mean Square Error* (RMSE). The smaller MAE or RMSE value means better rating prediction accuracy. In the following equations, $\mathcal{T}$ is the set of user-item rating pairs (u, i) used for testing.

$$MAE = \frac{1}{|\mathcal{T}|} \sum_{(u,i) \in \mathcal{T}} |r_{ui} - \hat{r}_{ui}|$$
$$RMSE = \sqrt{\frac{1}{|\mathcal{T}|} \sum_{(u,i) \in \mathcal{T}} (r_{ui} - \hat{r}_{ui})^2}$$

Baseline Methods We compare the proposed models with the following baseline methods. All these baseline methods are implemented by using the MyMediaLite library [79].

- (1) *GlobalMean*: this method predicts an unknown rating to be the average of all known ratings, *i.e.*, $\hat{r}_{ui} = \mu$.
- (2) *UserMean*: this method utilizes the mean rating of each user to predict the missing values.
- (3) *ItemMean*: this method uses the mean rating of each item to predict the missing values.
- (4) *UserKNN*: this method is based on user-based collaborative filtering. The unknown ratings are predicted by considering ratings given by similar users. The user similarity is computed by Cosine similarity of ratings and $k = 150$.
- (5) *ItemKNN*: this method is based on item-based collaborative filtering. The unknown ratings are predicted by considering ratings received by similar items. Similarly the item similarity is computed by Cosine similarity of ratings and $k = 150$.
- (6) *BiasedMF*: this is the MF model with user and item biases briefly described in Section 4.1. BiasedMF is widely used as a baseline in recommender systems.
- (7) *SVD++*: this model considers implicit feedback from users for rating prediction [5]. SVD++ usually offers a superior accuracy and is considered as a state-of-the-art matrix factorization algorithm for recommendation.
- (8) *Social MF*: this model considers implicit social information between items and/or users. The implicit social information can be derived from most similar and dissimilar users/items using Pearson’s correlations or cosine similarity of ratings. As our model considers influences from nearest geographical neighbors, for a fair comparison, we include the implicit social information from most similar items in *Social MF*.¹ Following [6], the most similar items are selected by using Pearson’s correlation of ratings greater than 0.75.

¹Note that, our model can also be extended to include the other kinds of implicit social information as in *Social MF*. However, the main purpose of our experimental evaluation is to evaluate the influence from geographical neighbors in comparison to the influence from similar items measured by rating-based similarity.

Proposed Methods We extended BiasedMF to incorporate influences from multiple factors: geographical neighborhood (N), business category (C), review content (R), business popularity (P), and geographical distance (D). The proposed methods are denoted using the letters in the parentheses to indicate the influences considered in each method.

- N -MF: this method incorporates geographical neighborhood influence (Section 4.2, Equation 4.4).
- NC -MF: this method incorporates both geographical neighborhood and category influence (Section 4.3, Equation 4.6).
- NCR -MF: this model incorporates neighborhood, category and review content influence (Section 4.4, Equation 4.9).
- $NCRP$ -MF: this model incorporates neighborhood, category, review content and popularity influence, by setting $z = \beta_i \rho_i$ in Equation 4.11 (Section 4.5).
- $NCRPD$ -MF: this model incorporates all factors: neighborhood, category, review content, popularity and geographical distance (Section 4.5).

We also evaluate another two methods: CRP -MF and $CRPD$ -MF. These two methods do not incorporate geographical neighborhood influence but incorporate influences from other factors (*i.e.*, C , R , P , and D) indicated by the method names.

Parameter Setting We performed 5-fold cross-validation on the training set to empirically set the hyperparameters. The number of latent factors $f = 20$; the relative importance of neighborhood and category influences are set to $\alpha_1 = 0.8$, $\alpha_2 = 0.6$; The regularization parameters: $\lambda_1 = 0.8$, $\lambda_2 = 0.4$, $\lambda_3 = 0.6$. The latent factors are learned by SGD with initial learning rate $\gamma = 0.008$, which decreases by a factor of 0.9 after each iteration (see Appendix A). The same parameters are used in all methods for fair comparison for all our proposed methods and the baseline methods whenever applicable. For example, the number of latent factors is also set to 20 in baseline methods *Biased MF*, *SVD++* and *Social MF*. For geographical neighborhood influence, by default, the proposed methods use the 6 nearest neighbors

Table 5.1: MAE and RMSE of all methods, the lower the better.

ID	Method	MAE	RMSE
1	Global Mean (μ)	0.8854	1.0962
2	Item Mean	0.8369	1.0939
3	User Mean	0.8599	1.0838
4	Item KNN	0.8208	1.0574
5	User KNN	0.8110	1.0429
6	Biased MF	0.8237	1.0483
7	SVD++	0.8120	1.0352
8	Social MF	0.8123	1.0303
9	N-MF	0.7952	1.0110
10	NC-MF	0.7929	1.0096
11	NCR-MF	0.7923	1.0078
12	NCRP-MF	0.7920	1.0072
13	NCRPD-MF	0.7958	1.0132
14	CRP-MF	0.7956	1.0138
15	CRPD-MF	0.8062	1.0191

for each business. For all the methods based on matrix factorization, the reported results are averaged over 5 runs to avoid the impact of initialization in parameter learning.

5.2 Experimental Results

We first compare the proposed methods with baseline methods and then evaluate the two schemes for defining the set of geographical neighbors for a business. Lastly, we evaluate the proposed methods with cold-start setting.

5.2.1 Method Comparison

The prediction errors measured by MAE and RMSE of all methods are reported in Table 5.1 with best results highlighted in boldface. The relative improvement over GlobalMean method is also plotted in Figure 5.1. We make four observations from the results.

First, incorporating geographical neighborhood influence into business rating prediction greatly reduces prediction errors measured by both MAE and RMSE. All the proposed methods with geographical neighborhood influence (*i.e.*, methods 9 - 13)

Figure 5.1: Improvement over GlobalMean (μ)

outperform all baseline methods (methods 1 - 8). The best prediction accuracy is achieved by *NCRP-MF* which considers geographical neighborhood (N), business category (C), review content (R), and business popularity (P). With geographical neighborhood influence alone, *N-MF* outperforms all baselines including state-of-the-art methods *SVD++* and *Social MF*. This result suggests that geographical neighbors are more effective than the “neighbors” derived from rating-based similarity as in *Social MF*.

Second, compared with geographical neighborhood, further considering factors like business category (*C*), review content (*R*), and business popularity (*P*) leads to relatively small additional reduction in prediction errors. The geographical distance factor (*D*) adversely affects business rating prediction accuracy. Method *NCRPD-MF* performs the worst among all methods with geographical neighborhood influence measured by both MAE and RMSE. This result suggests that the geographical distance factor introduces noise in the prediction model. Such result supports our earlier discussion that geographical distance between users and items is less relevant to the problem of business rating prediction.

Third, without incorporating geographical influence, *CRP-MF* performs poorer than most methods with geographical neighborhood influence including *N-MF*, *NC-MF*, *NCR-MF*, and *NCRP-MF*. The poorer performance of *CRP-MF* against *N-MF* suggests that the geographical neighborhood influence is more effective than

the combination of the three factors (C , R , and P) in business rating prediction. The much poorer performance of $CRPD-MF$ in comparison with $CRP-MF$ again suggests the adverse effect of geographical distance factor (D). On the other hand, the effectiveness of geographical neighborhood influence is also reflected from the better performance of $NCRPD-MF$ compared with $CRPD-MF$.

Last, among the 8 baseline methods, $SVD++$ and $Social\ MF$, the two state-of-the-art methods, perform the best evaluated by RMSE. By MAE measure, $User\ KNN$ gives surprisingly low error rate and the two methods $SVD++$ and $Social\ MF$ remain among the three best methods. This tells us that, as one of the most successful model-based CF models, latent factor models always outperform other collaborative filtering methods.

5.2.2 Impact of Neighborhood Size

In this Section, we study the impact of the size of geographical neighbors. Basically, given a business, its geographical neighbors can be defined based on either a geographical distance threshold or the number of nearest neighbors. In this set of experiments, we evaluate the two schemes on their impact to the prediction error of the $N-MF$ method. The $N-MF$ method is selected as the method for evaluation because it only considers the geographical neighborhood factor.

Figures 5.2.a and 5.2.b respectively plot MAE and RMSE of the $N-MF$ method by defining geographical neighbors with a distance threshold ranging from 20 to 2000 meters. Figures 5.3.a and 5.3.b respectively plot the MAE and RMSE of the method by taking the number of nearest neighbors ranging from 1 to 10. The y -axes of the two sets of figures are plotted in the same scale for easy comparison. From the two sets of figures, either the distance threshold of 1000 meters or the top-6 nearest neighbors give the best prediction accuracy by considering both MAE and RMSE. Plotted in Figure 3.1, with a distance threshold of 1000 meters, more than 90% of businesses have at least 6 neighbors and about 84% of the businesses have at least 10 neighbors. Because some businesses may have a large number of neighbors within 1000 meters, we choose to use 6 nearest neighbors mainly for minimizing computational cost. We have also evaluated the method by using at most 6 nearest neighbors within a 1000-meter distance threshold. However, poorer

Table 5.2: MAE and RMSE in cold-start setting

Method	MAE	RMSE
Global Mean	1.0319	1.2749
User Mean	0.9963	1.2566
Biased MF	1.0020	1.2539
N-MF	0.9956	1.2538
NC-MF	0.9936	1.2535

prediction accuracy was obtained compared with simply using 6 nearest neighbors without a distance threshold.

5.2.3 Cold-Start Business Rating Prediction

Cold-start is a challenging issue in any recommender systems when there is not enough information about either users or items. We now evaluate the performance of the proposed methods in predicting ratings by *existing users* to *newly established* businesses. Recall that in our data preprocessing (see Section 5.1), businesses and users with fewer than 10 reviews were removed. In this set of experiments, we try to predict the ratings by existing users in the training data to the businesses that were removed in data preprocessing. That is, from existing training data, we have the user vector of latent factors $\mathbf{p}_u$, but not the item vector $\mathbf{q}_i$ because the “newly established” businesses were not the training data. In total, there are 20,395 ratings made by 3,319 existing users to 6,939 new businesses. This collection of ratings will be used as test set.

Among all baseline methods, *Global Mean* and *User Mean* can be easily applied. *Item Mean*, *User KNN* and *Item KNN* cannot be applied as we assume that the newly established businesses have no reviews yet. Both *Biased MF* and *SVD++* are reduced to $\mu + b_u$ because b_i and $\mathbf{q}_i$ are both unknown (see Equation 4.2). *Social MF* is not applicable here because it relies on similar items by some similarity functions. Among the proposed methods, both the latent factors for the extrinsic characteristics of geographical neighbors (*i.e.*, $\mathbf{v}_n$ ’s of existing businesses) and business category (*i.e.*, $\mathbf{d}_c$) can be utilized in prediction. Here, we assume the category of the newly established business is known (*e.g.*, restaurant or bookstore). The prediction is made by Equation 4.6 after removing b_i and $\mathbf{q}_i$ from the equation.

Experimental result are reported in Table 5.2. From the result, we can see that the proposed method *NC-MF* achieves the best prediction accuracy by considering both geographical neighborhood and business category. Utilizing geographical influence alone, *N-MF* is the second best performing methods by both MAE and RMSE. Specifically, both *User Mean* and *Biased MF* are better than *Global Mean* with the former achieves better MAE and the latter achieves better RMSE. In summary, the proposed methods predict more accurate ratings for existing users to newly established businesses in cold-start setting.

5.2.a: Neighbors by distance (MAE)

5.2.b: Neighbors by distance (RMSE)

Figure 5.2: Prediction error of N -MF for determining geographical neighbors by distance

5.3.a: Neighbors by neighborhood size (MAE)

5.3.b: Neighbors by neighborhood size (RMSE)

Figure 5.3: Prediction error of N -MF for determining geographical neighbors by number

Chapter 6

Conclusions and Future Work

In this chapter, we first summarize the work that has been done and then we detail the future work to be further explored.

6.1 Conclusion

Rating prediction is one of the most popular and fundamental problems in recommendation systems. The main task is to predict the preference rating of a user to an item that she has not rated before. In this thesis, we study the business rating prediction problem with business review data from Yelp, a business review site. A Yelp user or Yelper can share her experience with a business by posting a review of the business and also a rating from 1 to 5 stars. We study the geographical influences to business rating prediction through the task of predicting how many stars a user would give to an un-reviewed business by considering neighbors of business and other factors.

Different from most other types of items that have been studied in various recommender systems (*e.g.*, movie, song, book), a business here physically exists at a geographical location, and most businesses have geographical neighbors within walking distance. When a user visits a business, there is a good chance that she walks by its neighbors. Through data analysis on Yelp, we observe that most businesses have neighbors within a short geographical distance from their locations. On the other hand, we find that there exists weak positive correlation between a business's ratings and its neighbors' ratings, and the positive correlation in ratings is independent of the categories of the businesses and/or their neighbors. Based on

these important observations, we assume that a user’s rating to a given business is determined by both the intrinsic characteristics of the business and the extrinsic characteristics of its geographical neighbors. Using the widely adopted latent factor model for rating prediction, in our proposed solution, we use two kinds of latent factors to model a business: one for its intrinsic characteristics and the other for its extrinsic characteristics. The latter encodes the neighborhood influence of this business to its geographical neighbors. Because of considering geographical neighbors, this problem is also related to point-of-interests recommendation and prediction, where various factors have been evaluated. These factors include category, review content, popularity and distance. We therefore also considered these factors in our study for the completeness of the evaluation. The category of a business usually reflects the characteristics of a business and users may use different criteria to evaluate businesses in different categories. The review elaborates the reason behind the rating and partially reflects the characteristics of the business. Regional popularity and geographical distance are two important features that has been studied in POI recommendation. We additionally model the influences from such those factors including business category, user review, popularity and geo-distance.

We conduct extensive experiments on the Yelp dataset to evaluate the proposed models, and compare the models with state-of-the-art baseline methods. We show that by incorporating geographical neighborhood influences, much lower prediction error is achieved than the baseline models including Biased MF, SVD++, and Social MF. Additional factors like business category, review content, and popularity can further improve the rating prediction accuracy. Nevertheless, the geographical distance between a user and a business adversely affect the prediction accuracy, although it is an important factor in POI recommendation and POI prediction. The incorporation of geographical neighborhood influence also partially enable our methods to better handle cold-start situation, for rating prediction of newly established businesses based on both their geographical neighbors and business category.

To the best of our knowledge, this is the first study on geographical neighborhood influence to users’ business rating behavior. We believe the observation that a business’s rating weakly positively correlated with its geographical neighbors’ rating is an important finding.

6.2 Limitations and Implications

The proposed methods in this research are relatively specific, because an item in this rating predication problem must has a static geographical location like business. Other example datasets include Gowalla and Foursquare check-in data and hotel reviews collected from TripAdvisor or other review sites. The observations made in this study therefore may not be applicable to other kinds of items (*e.g.*, movie, book, music). Another limitation is that currently two simple schemes are used to determine the “geographical neighborhood” *e.g.*, by distance or by number of neighbors. There might exists a much more complicated relationship between a business and its neighbors. In other words, for different businesses, a different sets of businesses might be considered for their influences. This will be part of our future study.

The observation that a business’s rating weakly positively correlates with its geographical neighbors rating is an important finding. This could be useful for not only in business rating prediction, but also in related studies. POI recommendation and sentiment analysis of business reviews are example areas which may find this research finding useful.

6.3 Future Work

As a part of our future work, we are interested in investigating the influence of geographical neighborhood in POI recommendation. On the one hand, users’ decisions are influenced by the extrinsic characteristics of its geographical neighbors. On the other hand, for two close businesses, users have personal preference as they usually prefer one than the other. We believe this is a pairwise choice and is very similar to the problems in Computational Advertising [80]. By treating the businesses as ads, we can introduce *learning to rank* method to automatically construct a ranking model from the business pairwise preference set. After learning the parameters of the ranking model, we can simply use it to rank the unvisited businesses and recommend the top- N businesses to users.

Another possible direction of our research is finding the influence of geographical neighborhood in sentiment analysis of business reviews. Since users may express

the feelings about the extrinsic characteristics of business in reviews, we believe geographical neighborhood does not only affect users' business rating behavior, but also have influence to users' opinions on businesses.

Appendix A

Parameter Estimation

All the objective functions (*e.g.*, Equations 4.5, 4.7, 4.10, and 4.13) in the proposed models share the same form. Next, we detail the parameter estimation for Equation 4.13 (where $z = \beta_i \rho_i + \beta_u \tau_{u,i}$) as an example using Stochastic Gradient Descent (SGD) algorithm [5]. Let e_{ui} be the error associated with the prediction $e_{ui} = r_{ui} - \hat{r}_{ui}$. The parameters are learned by moving in the opposite direction of the gradient with a learning rate γ in an iterative manner. In our experiments, learning rate γ is set to 0.008 in the first iteration and is decreased by a factor of 0.9 after each subsequent iteration.

$$\begin{aligned}
b_u &\leftarrow b_u + \gamma \cdot (e_{ui} - \lambda_2 \cdot b_u) \\
b_i &\leftarrow b_i + \gamma \cdot (e_{ui} - \lambda_2 \cdot b_i) \\
\beta_i^\rho &\leftarrow \beta_i^\rho + \gamma \cdot (e_{ui} \cdot \rho_i - \lambda_2 \cdot \beta_i^\rho) \\
\beta_u^\tau &\leftarrow \beta_u^\tau + \gamma \cdot (e_{ui} \cdot \tau_{u,i} - \lambda_2 \cdot \beta_u^\tau) \\
\mathbf{p}_u &\leftarrow \mathbf{p}_u + \gamma \cdot \left(e_{ui} \cdot \left(\frac{1}{|R_i|} \sum_{w \in R_i} \mathbf{q}_w + \frac{\alpha_1}{|N_i|} \sum_{n^{(j)} \in N_i} \mathbf{v}_{n^{(j)}} \right. \right. \\
&\quad \left. \left. + \frac{\alpha_2}{|C_i|} \sum_{c^{(j)} \in C_i} \mathbf{d}_{c^{(j)}} \right) - \lambda_1 \cdot \mathbf{p}_u \right) \\
\forall w \in R_i & : \mathbf{q}_w \leftarrow \mathbf{q}_w + \gamma \cdot \left(e_{ui} \cdot \frac{1}{|R_i|} \cdot \mathbf{p}_u - \lambda_1 \cdot \mathbf{q}_w \right) \\
\forall n^{(j)} \in N_i & : \mathbf{v}_{n^{(j)}} \leftarrow \mathbf{v}_{n^{(j)}} + \gamma \cdot \left(e_{ui} \cdot \frac{\alpha_1}{|N_i|} \cdot \mathbf{p}_u - \lambda_3 \cdot \mathbf{v}_{n^{(j)}} \right) \\
\forall c^{(j)} \in C_i & : \mathbf{d}_{c^{(j)}} \leftarrow \mathbf{d}_{c^{(j)}} + \gamma \cdot \left(e_{ui} \cdot \frac{\alpha_2}{|C_i|} \cdot \mathbf{p}_u - \lambda_3 \cdot \mathbf{d}_{c^{(j)}} \right)
\end{aligned}$$

Appendix B

Author's Publications

- **Longke Hu**, Aixin Sun, Yong Liu, “Your Neighbors Affect Your Ratings: On Geographical Neighborhood Influence to Rating Prediction”, *In Proc. of 37th Annual ACM SIGIR Conference (SIGIR14)*. Pages 345-354. Gold Coast, Australia. July 2014.

References

- [1] A. Rae, V. Murdock, A. Popescu, and H. Bouchard, “Mining the web for points of interest,” in *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR ’12, pp. 711–720, 2012.
- [2] D. Yang, T. Chen, W. Zhang, Q. Lu, and Y. Yu, “Local implicit feedback mining for music recommendation,” in *Proceedings of the Sixth ACM Conference on Recommender Systems*, RecSys ’12, pp. 91–98, ACM, 2012.
- [3] N. Koenigstein, G. Dror, and Y. Koren, “Yahoo! music recommendations: Modeling music ratings with temporal dynamics and item taxonomy,” in *Proceedings of the Fifth ACM Conference on Recommender Systems*, RecSys ’11, pp. 165–172, ACM, 2011.
- [4] Y. Koren, “Collaborative filtering with temporal dynamics,” in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’09, pp. 447–456, ACM, 2009.
- [5] Y. 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*, KDD ’08, pp. 426–434, ACM, 2008.
- [6] H. Ma, “An experimental study on implicit social recommendation,” in *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR ’13, pp. 73–82, ACM, 2013.

- [7] J. Liu, P. Dolan, and E. R. Pedersen, “Personalized news recommendation based on click behavior,” in *Proceedings of the 15th International Conference on Intelligent User Interfaces*, IUI ’10, pp. 31–40, ACM, 2010.
- [8] F. Abel, Q. Gao, G.-J. Houben, and K. Tao, “Twitter-based user modeling for news recommendations,” in *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, IJCAI’13, pp. 2962–2966, AAAI Press, 2013.
- [9] L. Li and T. Li, “News recommendation via hypergraph learning: Encapsulation of user behavior and news content,” in *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, WSDM ’13, pp. 305–314, ACM, 2013.
- [10] P. Yin, P. Luo, W.-C. Lee, and M. Wang, “App recommendation: A contest between satisfaction and temptation,” in *Proceedings of the Sixth ACM International Conference on Web Search and Data Mining*, WSDM ’13, pp. 395–404, ACM, 2013.
- [11] N. Natarajan, D. Shin, and I. S. Dhillon, “Which app will you use next?: Collaborative filtering with interactional context,” in *Proceedings of the 7th ACM Conference on Recommender Systems*, RecSys ’13, pp. 201–208, ACM, 2013.
- [12] G. Dror, Y. Koren, Y. Maarek, and I. Szpektor, “I want to answer , who has a question ? yahoo ! answers recommender system,” in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’11, pp. 1109–1117, ACM, 2011.
- [13] S. Rendle and L. Schmidt-Thieme, “Pairwise interaction tensor factorization for personalized tag recommendation,” in *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, WSDM ’10, pp. 81–90, ACM, 2010.
- [14] B. Sigurbjörnsson and R. van Zwol, “Flickr tag recommendation based on collective knowledge,” in *Proceedings of the 17th International Conference on World Wide Web*, WWW ’08, pp. 327–336, ACM, 2008.

- [15] Z. Guan, J. Bu, Q. Mei, C. Chen, and C. Wang, “Personalized tag recommendation using graph-based ranking on multi-type interrelated objects,” in *Proceedings of the 32Nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR ’09, pp. 540–547, ACM, 2009.
- [16] R. Baraglia, C. I. Muntean, F. M. Nardini, and F. Silvestri, “Learnnext: Learning to predict tourists movements,” in *Proceedings of the 22Nd ACM International Conference on Conference on Information & Knowledge Management*, CIKM ’13, pp. 751–756, ACM, 2013.
- [17] A. Gionis, T. Lappas, K. Pelechrinis, and E. Terzi, “Customized tour recommendations in urban areas,” in *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, WSDM ’14, pp. 313–322, ACM, 2014.
- [18] A. Sharma and B. Yan, “Pairwise learning in recommendation: Experiments with community recommendation on linkedin,” in *Proceedings of the 7th ACM Conference on Recommender Systems*, RecSys ’13, pp. 193–200, ACM, 2013.
- [19] W.-Y. Chen, D. Zhang, and E. Y. Chang, “Combinational collaborative filtering for personalized community recommendation,” in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’08, pp. 115–123, ACM, 2008.
- [20] K. Chen, T. Chen, G. Zheng, O. Jin, E. Yao, and Y. Yu, “Collaborative personalized tweet recommendation,” in *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR ’12, pp. 661–670, ACM, 2012.
- [21] C. Lu, W. Lam, and Y. Zhang, “Twitter user modeling and tweets recommendation based on wikipedia concept graph,” in *Workshops at the Twenty-Sixth AAAI Conference on Artificial Intelligence*, pp. 33–38, 2012.
- [22] Y. Shi, A. Karatzoglou, L. Baltrunas, M. Larson, A. Hanjalic, and N. Oliver, “Tfmap: Optimizing map for top-n context-aware recommendation,” in *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR ’12, pp. 155–164, ACM, 2012.

- [23] E. Diaz-Aviles, L. Drumond, L. Schmidt-Thieme, and W. Nejdl, “Real-time top-n recommendation in social streams,” in *Proceedings of the Sixth ACM Conference on Recommender Systems*, RecSys ’12, pp. 59–66, ACM, 2012.
- [24] P. Cremonesi, Y. Koren, and R. Turrin, “Performance of recommender algorithms on top-n recommendation tasks,” in *Proceedings of the Fourth ACM Conference on Recommender Systems*, RecSys ’10, pp. 39–46, ACM, 2010.
- [25] M. Deshpande and G. Karypis, “Item-based top-n recommendation algorithms,” *ACM Trans. Inf. Syst.*, vol. 22, no. 1, pp. 143–177, 2004.
- [26] S. Kabbur, X. Ning, and G. Karypis, “Fism: Factored item similarity models for top-n recommender systems,” in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’13, pp. 659–667, ACM, 2013.
- [27] P. Symeonidis, E. Tiakas, and Y. Manolopoulos, “Product recommendation and rating prediction based on multi-modal social networks,” in *Proceedings of the Fifth ACM Conference on Recommender Systems*, RecSys ’11, pp. 61–68, ACM, 2011.
- [28] J. Tang, H. Gao, X. Hu, and H. Liu, “Context-aware review helpfulness rating prediction,” in *Proceedings of the 7th ACM Conference on Recommender Systems*, RecSys ’13, pp. 1–8, ACM, 2013.
- [29] M. S. Desarkar, S. Sarkar, and P. Mitra, “Aggregating preference graphs for collaborative rating prediction,” in *Proceedings of the Fourth ACM Conference on Recommender Systems*, RecSys ’10, pp. 21–28, ACM, 2010.
- [30] L. Xiang, Q. Yuan, S. Zhao, L. Chen, X. Zhang, Q. Yang, and J. 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*, KDD ’10, pp. 723–732, ACM, 2010.

- [31] S. Baluja, R. Seth, D. Sivakumar, Y. Jing, J. Yagnik, S. Kumar, D. Ravichandran, and M. Aly, “Video suggestion and discovery for youtube: Taking random walks through the view graph,” in *Proceedings of the 17th International Conference on World Wide Web*, WWW ’08, pp. 895–904, ACM, 2008.
- [32] S. Lee, S.-i. Song, M. Kahng, D. Lee, and S.-g. Lee, “Random walk based entity ranking on graph for multidimensional recommendation,” in *Proceedings of the Fifth ACM Conference on Recommender Systems*, RecSys ’11, pp. 93–100, ACM, 2011.
- [33] V. C. Ostuni, T. Di Noia, E. Di Sciascio, and R. Mirizzi, “Top-n recommendations from implicit feedback leveraging linked open data,” in *Proceedings of the 7th ACM Conference on Recommender Systems*, RecSys ’13, pp. 85–92, ACM, 2013.
- [34] I. Fernández-Tobías, I. Cantador, M. Kaminskis, and F. Ricci, “A generic semantic-based framework for cross-domain recommendation,” in *Proceedings of the 2Nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems*, HetRec ’11, pp. 25–32, ACM, 2011.
- [35] X. Yu, H. Ma, B.-J. P. Hsu, and J. Han, “On building entity recommender systems using user click log and freebase knowledge,” in *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, WSDM ’14, pp. 263–272, ACM, 2014.
- [36] M. Balabanović and Y. Shoham, “Fab: Content-based, collaborative recommendation,” *Commun. ACM*, vol. 40, no. 3, pp. 66–72, 1997.
- [37] A. M. Ferman, J. H. Errico, P. v. Beek, and M. I. Sezan, “Content-based filtering and personalization using structured metadata,” in *Proceedings of the 2Nd ACM/IEEE-CS Joint Conference on Digital Libraries*, JCDL ’02, pp. 393–393, ACM, 2002.
- [38] M. J. Pazzani and D. Billsus, “The adaptive web,” ch. Content-based Recommendation Systems, pp. 325–341, Springer-Verlag, 2007.

REFERENCES

- [39] S. Ray and A. Mahanti, “Improving prediction accuracy in trust-aware recommender systems,” in *Proceedings of the 2010 43rd Hawaii International Conference on System Sciences*, HICSS ’10, pp. 1–9, IEEE, 2010.
- [40] X. Yang, H. Steck, and Y. Liu, “Circle-based recommendation in online social networks,” in *Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’12, pp. 1267–1275, ACM, 2012.
- [41] G. Adomavicius and A. Tuzhilin, “Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions,” *IEEE Trans. on Knowl. and Data Eng.*, vol. 17, no. 6, pp. 734–749, 2005.
- [42] A. Karatzoglou, X. Amatriain, L. Baltrunas, and N. Oliver, “Multiverse recommendation: N-dimensional tensor factorization for context-aware collaborative filtering,” in *Proceedings of the Fourth ACM Conference on Recommender Systems*, RecSys ’10, pp. 79–86, ACM, 2010.
- [43] H. Yildirim and M. S. Krishnamoorthy, “A random walk method for alleviating the sparsity problem in collaborative filtering,” in *Proceedings of the 2008 ACM Conference on Recommender Systems*, RecSys ’08, pp. 131–138, ACM, 2008.
- [44] L. Page, S. Brin, R. Motwani, and T. Winograd, “The pagerank citation ranking: Bringing order to the web,” in *Proceedings of the 7th International World Wide Web Conference*, WWW ’99, pp. 161–172, ACM, 1999.
- [45] T.-Y. Liu, “Learning to rank for information retrieval,” *Found. Trends Inf. Retr.*, vol. 3, no. 3, pp. 225–331, 2009.
- [46] P. Massa and P. Avesani, “Trust-aware recommender systems,” in *Proceedings of the 2007 ACM Conference on Recommender Systems*, RecSys ’07, pp. 17–24, ACM, 2007.
- [47] M. Jamali and M. Ester, “Trustwalker: A random walk model for combining trust-based and item-based recommendation,” in *Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’09, pp. 397–406, ACM, 2009.

- [48] S. Rendle, Z. Gantner, C. Freudenthaler, and L. 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*, SIGIR ’11, pp. 635–644, ACM, 2011.
- [49] S. Moghaddam, M. Jamali, and M. Ester, “ETF: Extended tensor factorization model for personalizing prediction of review helpfulness,” in *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining*, WSDM ’12, pp. 163–172, ACM, 2012.
- [50] J. S. Breese, D. Heckerman, and C. Kadie, “Empirical analysis of predictive algorithms for collaborative filtering,” in *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, UAI’98, pp. 43–52, Morgan Kaufmann Publishers Inc., 1998.
- [51] R. Jin, J. Y. Chai, and L. Si, “An automatic weighting scheme for collaborative filtering,” in *Proceedings of the 27th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR ’04, pp. 337–344, ACM.
- [52] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, “Item-based collaborative filtering recommendation algorithms,” in *Proceedings of the 10th International Conference on World Wide Web*, WWW ’01, pp. 285–295, ACM, 2001.
- [53] K. Niemann and M. Wolpers, “A new collaborative filtering approach for increasing the aggregate diversity of recommender systems,” in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’13, pp. 955–963, ACM, 2013.
- [54] T. Hofmann, “Latent semantic models for collaborative filtering,” *ACM Trans. Inf. Syst.*, vol. 22, pp. 89–115, 2004.
- [55] K. Miyahara and M. J. Pazzani, “Collaborative filtering with the simple bayesian classifier,” in *Proceedings of the 6th Pacific Rim International Conference on Artificial Intelligence*, PRICAI’00, pp. 679–689, Springer-Verlag, 2000.

REFERENCES

- [56] X. Su and T. M. Khoshgoftaar, “Collaborative filtering for multi-class data using belief nets algorithms,” in *Proceedings of the 18th IEEE International Conference on Tools with Artificial Intelligence*, ICTAI ’06, pp. 497–504, IEEE, 2006.
- [57] D. H. Stern, R. Herbrich, and T. Graepel, “Matchbox: Large scale online bayesian recommendations,” in *Proceedings of the 18th International Conference on World Wide Web*, WWW ’09, pp. 111–120, ACM, 2009.
- [58] L. Si and R. Jin, “Flexible mixture model for collaborative filtering,” in *Proceedings of the Twentieth International Conference on Machine Learning*, ICML ’03, pp. 704–711, AAAI Press, 2003.
- [59] S. H. S. Chee, J. Han, and K. Wang, “Rectree: An efficient collaborative filtering method,” in *Proceedings of the Third International Conference on Data Warehousing and Knowledge Discovery*, DaWaK ’01, pp. 141–151, Springer-Verlag, 2001.
- [60] L. H. Ungar and D. P. Foster, “Clustering methods for collaborative filtering,” in *Workshop on Recommender Systems at the 15th National Conference on Artificial Intelligence (AAAI’98)*, pp. 114–129, AAAI Press, 1998.
- [61] N. Golbandi, Y. Koren, and R. Lempel, “Adaptive bootstrapping of recommender systems using decision trees,” in *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM ’11, pp. 595–604, ACM, 2011.
- [62] X. Liu and K. Aberer, “Soco: A social network aided context-aware recommender system,” in *Proceedings of the 22Nd International Conference on World Wide Web*, WWW ’13, pp. 781–802, International World Wide Web Conferences Steering Committee, 2013.
- [63] Y. Koren, R. Bell, and C. Volinsky, “Matrix factorization techniques for recommender systems,” *Computer*, vol. 42, no. 8, pp. 30–37, 2009.

REFERENCES

- [64] G. Shani, D. Heckerman, and R. I. Brafman, “An mdp-based recommender system,” *J. Mach. Learn. Res.*, vol. 6, pp. 1265–1295, 2005.
- [65] T. Hofmann, “Unsupervised learning by probabilistic latent semantic analysis,” *Mach. Learn.*, vol. 42, pp. 177–196, 2001.
- [66] R. Salakhutdinov and A. Mnih, “Bayesian probabilistic matrix factorization using markov chain monte carlo,” in *Proceedings of the 25th International Conference on Machine Learning, ICML ’08*, pp. 880–887, ACM, 2008.
- [67] H. Steck, “Item popularity and recommendation accuracy,” in *Proceedings of the Fifth ACM Conference on Recommender Systems, RecSys ’11*, pp. 125–132, ACM, 2011.
- [68] X. Ning and G. Karypis, “Sparse linear methods with side information for top-n recommendations,” in *Proceedings of the Sixth ACM Conference on Recommender Systems, RecSys ’12*, pp. 155–162, ACM, 2012.
- [69] Q. Yuan, G. Cong, Z. Ma, A. Sun, and N. M. Thalmann, “Time-aware point-of-interest recommendation,” in *Proceedings of the 36th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR ’13*, pp. 363–372, ACM, 2013.
- [70] M. Ye, P. Yin, W.-C. Lee, and D.-L. 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, SIGIR ’11*, pp. 325–334, ACM, 2011.
- [71] E. Cho, S. A. Myers, and J. 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, KDD ’11*, pp. 1082–1090, ACM, 2011.
- [72] C. Cheng, H. Yang, I. King, and M. R. Lyu, “Fused matrix factorization with geographical and social influence in location-based social networks,” in *AAAI Conference on Artificial Intelligence, AAAI ’12*, pp. 17–23, AAAI Press, 2012.

REFERENCES

- [73] C. Cheng, H. Yang, M. R. Lyu, and I. King, “Where you like to go next: Successive point-of-interest recommendation,” in *Proceedings of the Twenty-Third International Joint Conference on Artificial Intelligence*, IJCAI’13, pp. 2605–2611, AAAI Press, 2013.
- [74] X. Liu, Y. Liu, K. Aberer, and C. Miao, “Personalized point-of-interest recommendation by mining users’ preference transition,” in *Proceedings of the 22Nd ACM International Conference on Conference on Information & Knowledge Management*, CIKM ’13, pp. 733–738, ACM, 2013.
- [75] A. Noulas, S. Scellato, N. Lathia, and C. Mascolo, “Mining user mobility features for next place prediction in location-based services,” in *Proceedings of the 2012 IEEE 12th International Conference on Data Mining*, ICDM ’12, pp. 1038–1043, IEEE Computer Society, 2012.
- [76] W. R. Tobler, “A computer movie simulating urban growth in the detroit region,” *Economic geography*, vol. 46, pp. 234–240, 1970.
- [77] B. Liu, Y. Fu, Z. Yao, and H. Xiong, “Learning geographical preferences for point-of-interest recommendation,” in *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD ’13, pp. 1043–1051, ACM, 2013.
- [78] Z. Cheng, J. Caverlee, K. Lee, and D. Z. Sui, “Exploring millions of footprints in location sharing services,” in *ICWSM*, pp. 81–88, The AAAI Press, 2011.
- [79] Z. Gantner, S. Rendle, C. Freudenthaler, and L. Schmidt-Thieme, “MyMediaLite: A free recommender system library,” in *Proceedings of the Fifth ACM Conference on Recommender Systems*, RecSys ’11, pp. 305–308, ACM, 2011.
- [80] A. Z. Broder, “Computational advertising and recommender systems,” in *Proceedings of the 2008 ACM Conference on Recommender Systems*, RecSys ’08, pp. 1–2, ACM, 2008.