TALLINN UNIVERSITY OF TECHNOLOGY
School of Information Technology

Muhammad Nadeem 182467IVSM

POI BASED SERENDIPITOUS RECOMMENDER ALGORITHM

Master Thesis

Supervisor
Prof. Dr. Sadok Ben Yahia
PhD
Co-Supervisor
Imen Ben Sassi
PhD

Tallinn 2021

TALLINNA TEHNIKAÜLIKOOL
Infotehnoloogia teaduskond

Muhammad Nadeem 182467IVSM

HUVIPUNKTIDEL POHINEV SERENDIPITI SOOVITUSALGORITM

Magistritöö

Juhendaja
Prof. Dr. Sadok Ben Yahia
PhD
kaasjuhendaja
Imen Ben Sassi
PhD

Tallinn 2021

Author’s declaration of originality

I hereby certify that I am the sole author of this thesis. All the used materials, references to the literature and the work of others have been referred to. This thesis has not been presented for examination anywhere else.

Author:	Muhammad Nadeem	.... (signature)
Date:	Jan 05, 2021	

Annotatsioon

Kasutajate arvukuse tõus mitmeil erineval võrgul on näinud tõusu soovitussüsteemi vajadusel. POI-l põhinev soovitussüsteem on olnud vajalik osa LBSN teenustest, kuna see aitab mitmetel inimestel külastada ja kasutada erinevaid asukohti nende ümber ning mitmetel kolmanda osapoole tarnijatel teenida tulu, pakkudes asukoha-baasil teenuseid. Mitmeid soovitussüsteeme on loodud läbi aastate, kuid ainsaks limiidiks nende süsteemidega on täpne, aga tuntud või ilmselge soovituste pakkumine. Soovitussüsteemides on sellised probleemid tuntud kui ülespetsialiseerumine.

Oleme õppinud sellist probleemi ning kuidas sellega toime tulla: oleme pakkunud teistmoodi lähenemist, mis on lahenduse leidmine läbi juhuse. See tähendab, et mingi kirje on ebapopulaarne, kuid asjakohane kasutajale. Meie lähenemine aitab ära tunda sarnaseid kasutajaid ning neile soovitada juhusliku lähenemisega asukohti kasutajale temale sarnaselt kasutajalt. Lahenduse toetamiseks oleme kasutanud kolme erinevat hindamismõõdikut, et võrrelda tulemusi teiste algtaseme algoritmidega.

Abstract

An increase in user growth on many social networks has increased the importance of the recommender system. POI-based recommender system has been a crucial part of LBSN services, as it helps many people to visit and enjoy different locations in their surroundings and many third-party vendors to generate revenue by providing location-based services. Many recommender systems have been proposed over the years but the only limitation with these systems is they provide accurate but rather familiar or obvious recommendations. In recommender systems, this problem is known as over-specialization.

We have studied the problem of over-specialization and to deal with it we have proposed an approach called serendipity. Serendipity means an item that is rather unpopular but relevant to a user. Our approach helps in recognizing similar users and then recommending serendipitous locations to a user from similar user. To back our approach we have used three different evaluation metrics to compare our results with some baseline algorithms.

The thesis is in English language and it contains 35 pages of text, 6 chapters, 11 figures, 1 table.

List of abbreviations and terms

RS	Recommender Systems
POI	Point of Interest
AMC	Additive Markov Chain
MF	Matrix Factorization
KG	Knowledge Graph
PFM	Probabilistic Factor Model
LBSN	Location Based Social Network

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1 Research Motivation

For the past decade, the growth in smart devices, network technologies like GPS, and different social media platforms has developed an interest in location-based social networks (LBSN) from different industries. In LBSN, people get connected with their social network of friends and families by sharing information over the internet like photos, and locations of visited places e.g. restaurants, zoo, movie theater, and many other attractions. By using this social interaction among people, LBSN uses this huge pool of data to recommend new locations and attractions to users depending on their interests and activities. This phenomenon of new location recommendation is well known as the point-of-interest (POI) recommender system (RS). In LBSN, POI plays an important role and it has been widely studied and used in big corporate to generate huge revenue by providing location-based services like an advertisement.

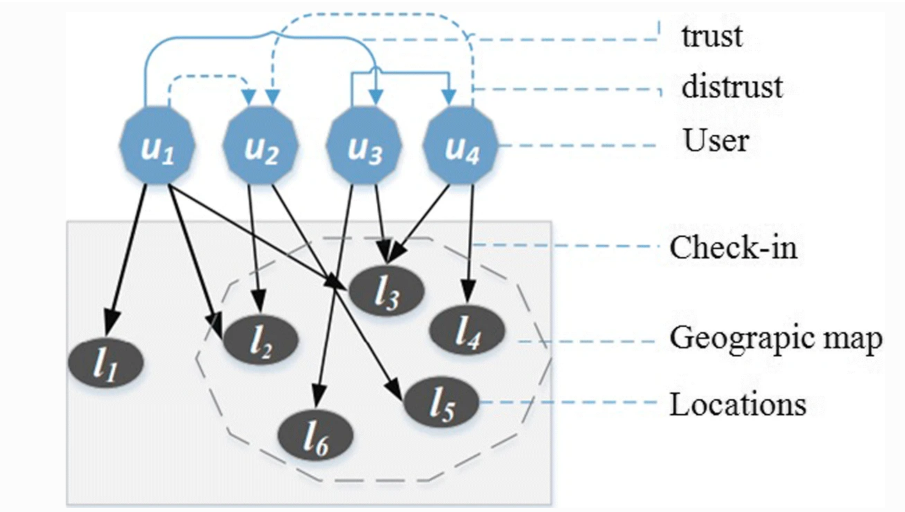

Figure 1. *A*n example of LBSN [1]

Some critics argue, because of some social media and their filters, we are limited to a certain horizon, and we live in our small world. They worry that it should be concerning because those systems only filter news and information from our friend circle which is related to our interest. These systems by doing so affect our innovative, creative, and venturous ideas. To cater to such problems we need a system that keeps our interest and relevance in mind and should filter information that has surprising elements for us. This leads us to the idea of serendipitous recommender systems

1.2 Research Goals

Traditional recommender systems analyze the behavior of users and their interests. Which later develops a user profile, and exploits that profile to recommend similar items to them. But, the constraint with these recommender systems is that they recommend those items that may be quite expected or obvious which sometimes makes it less interesting for users.

By definition, a serendipitous item means an item that is interesting but not popular. So in the case of a recommender system, a serendipitous item is the one that someone will find interesting but due to its popularity, it would have been difficult to discover. Our goal is to deal with this scenario which is called the serendipity problem. In this thesis, we are focusing on a recommender system that will surprise the users with serendipitous items that will be unexpected for them but also relevant to their profile.

1.3 Structure of the Thesis

This thesis is consist of five chapters:

1. **Basic Concepts:** In this chapter, we have discussed basic concepts related to the POI-based recommender system and some of the common POI RS and problems associated with POI RS.
2. **State of the Art:** We have discussed some work related to POI recommender systems. we have analyzed some of the state of the Art RS. we have also given an overview of some of the important models.
3. **Proposed Approach:** This chapter is about our approach for the serendipitous POI-based RS. We have discussed the general idea of our proposed approach and algorithm.
4. **Experimental Evaluation:** This chapter is about the experimental protocol and setup. we talked about the datasets we have used and the evaluation metrics. we have discussed some of the baseline models which we have used for cross reference.

2. Basic Concepts

2.1 Introduction

Recommendation systems are the modern system, which are designed to predict the new things to the end user, on the basis of their interest, activities and many other factors. These new things are products which the user is interested in or he will purchase. Some of the big companies like Amazon and Netflix etc. also uses recommendation systems to assist their audience, which product is suitable to their interest according to their activity on the platforms.

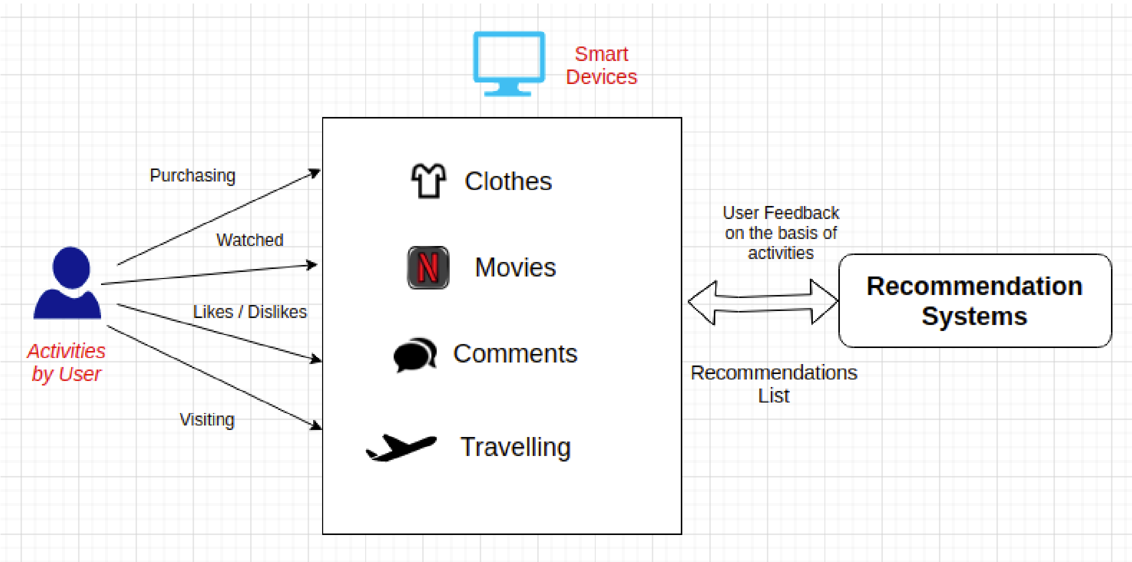

Figure 2. *Overview of recommendation systems*

2.2 General Definitions

In this section we will explain some of the main concepts and their definitions.

2.2.1 Point-of-Interest

A point of interest or POI is a location or attraction. Normally, these locations come under the category of restaurants, museums, parks, etc. POI interest plays an important role in recommendation systems as we recommend different locations to users on the basis of similarity of their interest.

2.2.2 Check-in

Physically visiting a location or arriving at a place like restaurants, airports is called a check-in.

2.3 Recommender Systems Overview

Generally recommendation systems requires a huge chunk of data to train the system for future predictions. This system filters the relevant and important information associated with particular users and then they apply different computations along with provided users preferences and interest.

2.3.1 Background

For the past one decade the world wide web has evolved into a large pool of data and services which has changed our lifestyle, reading books, streaming movies, purchasing different stuff, in fact our whole communication has changed. There is a big sea of services available on the internet which makes users overwhelmed with the choices to make. In such scenarios, recommendation systems are very helpful. They filter down the data into a small set where they can easily make decisions. And the same service provider can generate a source of revenue by adding a business value to their services.

2.3.2 Algorithms Classification

The use of recommendation has helped many businesses to grow by recommending the accurate and relevant item to users, which reduces the headache and trouble of finding the preferred items. Different companies are using different types or mixture of recommendation filter techniques depending on the constraint and limitation on different filtering techniques.

Content-Based Filtering:

Content-based filtering systems depend on the history or actions of users, how they have interacted with similar items. So the similar items are rated based on user interaction with the item., for example, purchasing some product, liking some post, or feedback on some movie. Now once we have a chunk of data available from user history we build a4 user-item profile and then based on that profile we recommend new items to users. Content base filters are mostly used when we recommend news, articles, movies, or some web information to users. For example, we can apply content-based filtering on movie recommender systems. As we have movies with different keywords depending on the type of genre associated with the movie. So if user u likes movies with genre Action, Adventure, Sci-Fi, etc. So users will always get suggestions for movies related to their previous history of watched movies and ratings provided by him on different movies.

Collaborative Filtering: Collaborative filtering is used to overcome shortcomings of content-based filters, which use similarity among the users to recommend items to users. In simple words, in collaborative filters, we can recommend items to U1 based on its similar interest with U2.

To understand the concept of collaborative filtering, we can look into the example of song recommendations. If user A and user B have a similar interest in music in the past. So based on their history of similar music likeness we can recommend user A some music which is in the playlist of user B.

Hybrid Filtering: As the name suggests Hybrid filtering comes from the combination of content-based filtering and collaborative filtering. Limitations and shortcomings in content-based filtering and collaborative filtering systems lead to the birth of a new system that utilizes the strength of each filter to make an analysis. It has been observed that the combination of both filtering systems has increased the common knowledge and resulted in better recommendations.

2.3.3 Challenges

Data Sparsity: In recommendation systems other than POI, we make a user-item matrix based on user direct activity, for example, giving a rating to some song. In such cases, ratings are numerical numbers. A greater number means a higher ranking score. Now in the POI recommender system, we don't have user rating but some user check-in for some location which results in a user-location matrix. User-location check-in has a bigger range than the rating for example users can visit one location way more than some other similar location and because of that, we have to face data sparsity issues. Big corporate-like Netflix has a data sparsity of 99.2 percent and Gowalla has a sparsity of 2.08X10-4. [2]

Scalability: As more and more people are using RS, the size of input user data is also increasing. Despite this fact RS still has to respond to system requests in seconds. In order to deal with such big data we need an efficient algorithm for RS. In a user based collaborative filtering system, the computational complexity in O (n2.m) where n represents user and m represents item rated by user. Similar problem has been found in POI based recommendation systems as the size of data increases with increase in user-location check-in matrix. [3, 4]

Serendipity: By definition serendipity means an item which was discovered accidentally, the discovery can be pleasant or unpleasant. So in RS when we recommend an item to a user which is surprising but not expected it is called a serendipitous item. The serendipitous is also one very prominent issue in RS, as it depends on the mood and time of the user for example one item which was relevant to one user may not be relevant for him after two days. One example can be weather which depends on the context of outdoor activity and music depends on the mood. [5]

2.4 POI Recommendation

For the past decade, There has been extensive research in the field of POI recommendation systems and numerous approaches has been discussed and proposed. In this section we will discuss some of the features of POI recommender system which makes POI recommender systems different from traditional recommender systems.

Geographical influence:

Geography is the most important feature of the POI recommender system as it distinguishes it from the other ones. It has been observed that user preferred to visit a location which is nearby rather than visiting a location far away. So, a user will most probably visit a POI which is nearby to another POI that he prefers.

Frequency data:

In traditional RS, user interest or preferences are measured on the basis of rating or feedback on different items i.e. movies, restaurants, books. But, in POI we use the frequency of check-ins in particular locations, which results in a user-location matrix.

2.4.1 Points-of-interest recommendation problem

Just like other recommendation systems, POI is also a recommendation systems which faces certain constraints which are as follows:

Physical: POI based recommender systems are dependent on check-in of the user on location which is considered a physical constraint as we compare it to streaming shows on Netflix, surfing on Amazon.

Extreme Spareness: Due to the huge chunk of POI data which exists in millions it is difficult to pick some top N values for recommendation.

Complicated relations: Due to the psychology of people, where people in social media are reluctant to make new friends in neighboring geographical locations, the relation among friends on social media generates real complicated relations between location to location and user to locations.

2.4.2 Different POI Recommendation Systems

Next POI recommendation: For the past few years, there has been extensive research on the recommendation of the next POI which users will visit in the near future. These next POI are recommended considering different factors including user-friends, information of POI in text, time interval and time visit. We can define a next POI recommendation as an example, If a user has visited a number of POIs in past

$$L_i^u = {q_t}_1^u, {q_t}_2^u, ..., {q_t}_{i-1}^u$$

for a time interval t_{i-1}. Now we have to give ranking to each POI on the basis of L_i^u for a time t_i. The higher the ranking of POI means the higher the probability of a user to visit that location. Now all the top POI with high ranking will be recommended. [6].

POI Itinerary Recommendation:

The aim of recommending multiple POIs with a connected itinerary is quite challenging as it is constrained by many factors including time constraint, time interval, popularity and preferences of users. Itinerary recommendation faces two main problems in general, as we can see the events are supposed to happen in future and for that we have small data to deal with and less user-item interaction as we have in other traditional recommendations.

The other problem we face is the short amount of activities and parallel events, we deal with a scenario where a user may attend an activity of less interest and miss an opportunity of more of his interest, it is called attendance bias. [7].

Time aware recommender system: In time aware recommendation system, time plays an important role in recommendation, as the user visits different places at a particular time. For example, restaurants for lunch in day time, theater for movies at night etc. Time aware recommendation works on two different behaviors, i,e. Temporal behavior where we check the historical behavior of user check-ins. Spatial behavior, in which we analyze the daily check-ins of users where he visits most of the nearby POIs. [8]

3. State of the Art

3.1 Related Work:

In this section we will take a look on some of the relevant and state of the art papers and book references.

A Survey of Point-of-Interest Recommendation in Location-Based Social Networks: [2]

In this paper, the authors have discussed how the rapid growth of social media and the internet has attracted a lot of audiences from different fields of life, especially academics and research. This paper has presented an approach that additional information with check-ins information helps to characterize the POI recommendation in four categories. Paper has described pure check-ins based POI only takes check-in frequency as score or rating and also if user two are visiting a similar up to a certain level, they describe them as similar users. In a geographical-based POI recommender system, in which the distance between two POIs visited by users and distance of user and POI location is considered to recommend a new POI location to the user. They have discussed all the other types of POI recommender Systems in detail.

This survey paper has helped us understand the unique characteristics of POI-based RS and their different classification.

An Experimental Evaluation of Point-of-interest Recommendation in Location-based Social Networks: [9]

The paper has discussed 12 different POI recommendation models, to show a general picture of POI-based RS from multiple aspects by evaluating them with different datasets, datasets of various sources. In this paper different evaluation techniques concerning user modeling, such as matrix factorization, methods for context information have been discussed. They have rephrased the key finding of RankGeoFM, GeoFm, IRenFM, on different datasets and types of users. They have given him an analysis of the shortcomings and strengths of different models. In this paper, some of the other variants of POI recommendation have been discussed

The experimental evaluation has given us a broad vision of cutting-edge algorithms with help of experimental evaluation. They have evaluated 12 state-of-the-art POI recommender models. These findings helped us choosing our models and evaluation metrics for research.

Exploiting sequential influence for location recommendations [10]

In this paper the author has discussed, how the sequential behavior affects the location recommendations. They first represented the sequential patterns as dynamic location-location graphical patterns by mining the sequential patterns from location sequences. Then predicting the probability of users visiting a location using AMC. Finally , generating a recommendation framework by integrating social influence with geographical influence and sequential influence.

This paper is one of the state-of-the-art model which we have used to evaluate our approach with our results.

Point-of-Interest Recommendations: Learning Potential Check-ins from Friends [11]

This paper discussed the cold start problem by splitting common users into three types: social friends, local friends and neighbouring friends. Then a two step framework is designed to tackle the information of friends, to improve the accuracy of recommendations and deal with cold-start problems. So this paper consists of 3 parts. In the first part analysis of correlation among the user and three types of friends. second, developed two approaches where for each individual, learn about the set of locations that their friends has visited and which she will be interested in. At the end, developing matrix factorization with the help of potential check-ins.

This research paper helped us understanding how to deal with challenges like the difficulty deal with address data sparsity and user-location cold-start problem.

Fused Matrix Factorization with Geographical and Social Influence in Location-Based Social Networks [12]

As many social media allowed users to share their check-ins with friends. Using check-ins information many personalized POI recommender systems have been proposed over the years. In this paper, authors proposed an idea of integrating matrix factorization with geographical and social influence to recommend new POIs in LBSNs. To get the geographical influence they used the probability of user check-ins on location as a Multi center Gaussian model. In the next part they integrated the social and geographical influence in MF.

In the paper, they have discussed the impact of the fused matrix along with geographical influence in depth. Their research was very helpful in understanding the fused matrix factorization framework and its impact and efficiency in LBSNs

Exploring Temporal Effects for Location Recommendation on Location-Based Social Networks [13]

This paper has described the temporal effects in terms of temporal regularization and aggregation. On the basis of observed temporal properties, they have proposed a location recommendation framework with temporal effects. They have used four different strategies related to temporal aggregation to fuse the user's check-in preferences for temporal states. Later the evaluation of temporal effects proved that time-dependent check-ins are more preferable over static check-ins and they give better location recommendations.

This paper helped to understand the temporal patterns via user behavior, and also how we can improve the location recommendation using temporal effects.

Location Recommendation in Location-based Social Networks using User Check-in Data [14]

By studying the LBSN we can affirm that social relation and traveling distance have influence over the location recommendation. In this paper, they have studied the LBSN a little more in-depth to observe that people are most likely to visit the nearest location to the last visited location and it is influenced by their social network. This paper also proves if we add more features to predict locations to users it will improve the accuracy of predicted locations. On the basis of these assumptions, they have proposed a recommendation system that outshines other traditional RS. They have used real data-sets from Gowalla and Brightkite, and they find some correlation between past visited locations and social information of the user's network. The results have shown that using correlations of the data helps the algorithm to recommend location superiorly as compared to the other state of the art RS

This paper has explained the idea of user-location relation and similar users in depth.

A Survey on Knowledge Graph-Based Recommender Systems [15]

We can retrieve side information using the knowledge graph as this area of research has gathered attention from researchers more recently. A KG is a heterogeneous graph, where nodes act as entities, and edges will be the relations among the entities. Item attributes are used to understand the relationship between items and Moreover, users and user side information is also fused into the KG. Now these techniques help in capturing relations between users and items, as well as the user preference. KG based recommender systems are applied in three ways, the embedding-based method, the path-based method and the unified method. This survey paper investigated KG-based recommender systems and gave the overview of the efforts being done in this field. This survey demonstrates different techniques using the KG as extra information to increase the efficiency of the recommendation result. Finally they have discussed what are the future aspects of the KG based RS and what domain researchers are aiming to improve them to increase the performance of RS.

A General Geographical Probabilistic Factor Model for Point of Interest Recommendation [16]

This paper addresses the geographical scenario of POI recommendation which can be useful in recommending better locations to users. This paper proposed a general geographical probabilistic factor model framework also called a Geo-PFM framework recommended by taking different strategies in consideration.This framework helps in getting the effect of geography on the check-ins patterns of the users along with user mobility which can be very effective in the recommendation model. On the basis of the Geo-PFM framework, they created a Poisson Geo-PFM, which results in a more effective process of generating more probabilistic processes for complete models and is more effective in better POI recommendation. They have analysed the algorithm on three real world LBSN data-sets which proves this recommendation systems outclass other state of the art models by a significant margin.

3.2 Overview of Important Models

POI recommendation is a vital utility to LBSNs as it helps a lot of businesses and end-users to get benefits. Some of the recommender systems are discussed over the years, but still there's room for improvement. In this section, we will discuss state-of-the-art POI recommendation models. Our goal is to provide an overview of the pioneering research on POI recommendation.

3.2.1 Exploring Temporal Effects for Location Recommendation on Location-Based Social Networks [13]

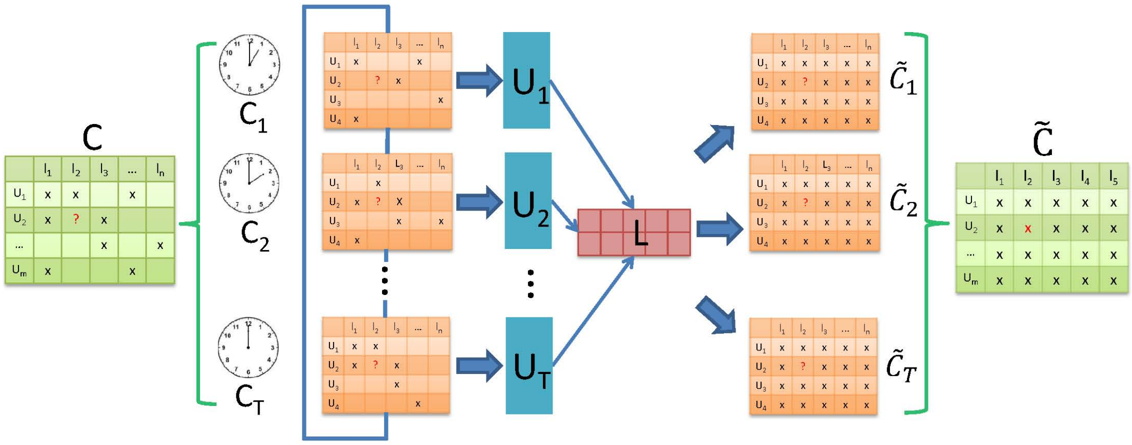

Figure 3. *Exploring Temporal Effects for Location Recommendation on Location-Based Social Networks [13]*

In this model the temporal effects is described in terms of temporal regularization and aggregation. Derived from observed temporal properties, they have proposed a location recommendation framework with temporal effects.

The diagram describes the workflow of the framework where ‘x’ shows observed check-in frequency and ‘?’ represents preferences for user unvisited locations which framework will deduce. The framework workflow is divided in three sections, first is temporal division, this section is responsible for further division of matrix C into further sub matrices where each sub matrix gives check-ins actions which are in correspondence to temporal state T. Second is temporal factorization, in which each C_t is separated into U_t known as check-in preference and L which is the location Characteristics. L is shared by all of U_t. Last section is temporal aggregation, low-rank approximation C_t is added up into C, which represents user check-in preferences for each location.

3.2.2 Fused Matrix Factorization with Geographical and Social Influence in Location-Based Social Networks [12]

A personalized RS is defined as a relation between a location and user check-in frequency matrix, where U is known as users, L defines locations, F is the user social relation and the goal is to recommend top k locations to users.

Multi-center Gaussian Model

To achieve the results, a Multi-center Gaussian model is proposed to gather the geographical influence on the users check-in and by applying matrix factorization, along with social information. Below is the probability of a user u,with multi-center set Cu, l is the POI location.

$$P(l|C_u) = \sum_{c_u=1}^{|C_u|} P(l \in c_u) \frac{f_{c_u}^{\alpha}}{\sum_{i \in C_u} f_i^{\alpha}} \frac{\mathcal{N}(l|\mu_{c_u}, \Sigma_{c_u})}{\sum_{i \in C_u} \mathcal{N}(l|\mu_i, \Sigma_i)}.$$

Figure 4. *Fused Matrix Factorization with Geographical and Social Influence in Location-Based Social Networks [12]*

Probabilistic Factor Model

Probabilistic factor model can model the frequency data by applying Beta distribution on Matrices U and V, poison distribution over frequency.

$$\begin{aligned}
\Psi(\cdot,\cdot;\cdot) \;=\; & \sum_{i=1}^{|\mathcal{U}|} \sum_{k=1}^{K} ((\alpha_k - 1)\ln(U_{ik}/\beta_k) - U_{ik}/\beta_k) \\
& + \sum_{j=1}^{|\mathcal{L}|} \sum_{k=1}^{K} ((\alpha_k - 1)\ln(L_{jk}/\beta_k) - L_{jk}/\beta_k) \\
& + \sum_{i=1}^{|\mathcal{U}|} \sum_{j=1}^{|\mathcal{L}|} (F_{ij}\ln(U^T L)_{ij} - (U^T L)_{ij}) + c,
\end{aligned}$$

Figure 5. *Fused Matrix Factorization with Geographical and Social Influence in Location-Based Social Networks [12]*

Results outlined from different sets of recommendations have shown that MGM performs better then PFM in all metrics which implies that geographical influence has a significant impact in POI recommendation.

3.2.3 LORE: Exploiting sequential influence for location recommendations [10]

In this model location recommendations in LBSNs are improved by considering the sequential influence on users' check-in behaviors. The model with sequential influence along with additive Markov chain (AMC), called LORE. In LORE, they first analyze the check-in location sequences of all users to incrementally mine sequential patterns as location-location transition graphs. A location sequence contains check-in locations for a user at a certain check-in time.

For sequential probability prediction of a user who is visiting a new POI, they have used an nth-order additive Markov chain. It has been observed the prediction of new locations not only depends on newly visited locations but also previously visited locations as well. Finally, they integrated the derived sequential, geographical probability, and social rating of the new location visited by the user into a new single score to recommend top-k new POIs.

For sequential probability prediction of a user who is visiting a new POI, they have used an nth-order additive Markov chain. It has been observed the prediction of new locations not only depends on newly visited locations but also previously visited locations as well. Finally, they integrated the derived sequential, geographical probability, and social rating of the new location visited by the user into a new single score to recommend top-k new POIs.

To formally commend the quality of the recommended location, we check how many locations visited locations matches the results discovered from a recommender system. To find these results mostly two performance metrics are used precision and recall. Precision is defined as the ratio of relevant locations to the k recommended locations.

$$Precision = \frac{number\ of\ relevant\ locations\ in\ top - k}{total\ recommended\ locations}$$

Figure 6. *Precision metric*

Recall defined as the ratio of relevant locations to the total relevant locations.

$$Recall = \frac{number\ of\ relevant\ locations\ in\ top - k}{total\ relevant\ locations}$$

Figure 7. *Recall metric*

3.2.4 Conclusion

In this chapter we have discussed some relevant work regarding POI recommender systems. We have given an overview of some of the state of the art models with respect to their results and performances

4. Proposed Approach

In this chapter, we will discuss the basic idea of our thesis, and our approach leading to the algorithm which we proposed.

4.1 Introduction

Recommender systems are also known as filters which provide related information to targeted users depending on their interest and history. First, they analyze the user activities on the platform from history and behavior. Then they build a profile for the user which contains information regarding their interest. In the end, by exploiting their profile they recommend new items to the user which will interest them.

It has been observed that this workflow of the traditional recommender system is responsible for the scenario called overspecialization also known as serendipity. To deal with the serendipity problem we have proposed a strategy where we find serendipitous items against each user, and then recommend those serendipitous items to similar users.

4.2 Approach

We aim to surprise the user with serendipitous locations, which are positively surprising and unexpected at the same time from other users which are similar to a user.

4.2.1 General Idea

Some RS recommend items that are very accurate and useful for certain purposes like a grocery store or music store where RS recommends products according to the user interest. However, in some cases, these traditional RS result in obvious or redundant recommendations. For the past few years, researchers have investigated the aspect of serendipity in the recommender system to avoid the scenario of redundant or over-specialized recommendations.

To move further ahead, we need to first explain serendipity. In the oxford dictionary, serendipity is defined as *The faculty of making happy and unexpected discoveries by* *accident*. In different literature i.e; studies of art, social science, and humanities serendipity is considered a vital part of creativity.

Various definition of serendipity in the aspect of RS has been proposed over the years. For example, serendipity is a measure of the degree to which the RS is both positively surprising and attractive to users. In some papers, serendipity has been defined as the extent to which the recommendations are surprising and successful.

Also, serendipity has been described as *the most closely related concept to unexpectedness involves a positive emotional response of the user about a previously unknown item and measures how surprising these recommendations are* [17]

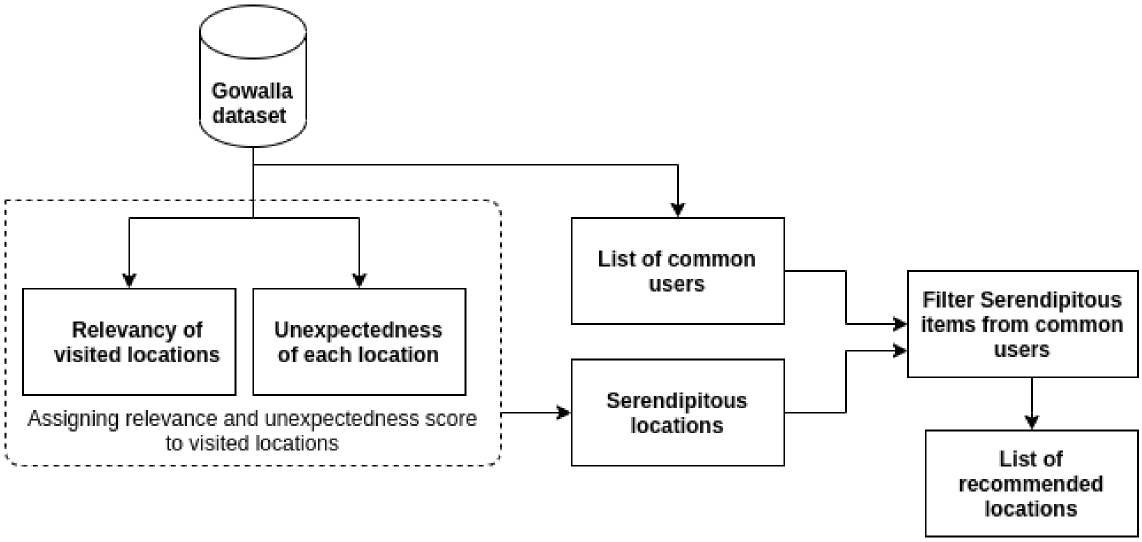

Figure 8. *Approach: serendipitous POI based recommender*

The general idea of the algorithm is to recommend those serendipitous locations to the user which are not in the popular category but are relevant to the interest of the user. These locations can be intimidating or interesting for users whom they are not going to find them otherwise.

4.2.2 Algorithm

Our approach[1] consist of four steps, in the first step we find all the relevant items to users. In a traditional recommender system, we use the rating to score each item, provided by the user and then mark each item as relevant to the user.

In a recommender system, relevancy is defined as a mechanism related to the interest of the user in a certain item, which can be modeled as a binary concept. An item can be

[1]https://github.com/NadeemMaqbool/seredipitous-recommender-systems

assigned a binary number if a user finds an item either interesting or not. So in a traditional recommender system, we can say an item i is said to be relevant to a user u if the user u has given the rating to that item, which is above the average rating provided by the user u on all the items. [18]

In POI recommender we don't have the user rating for locations, rather we use the frequency of check-ins as the rating parameter. Higher is the frequency of check-ins, the greater is the rating for the location.

So we can derive the definition of relevancy for the POI recommender system as a location l is relevant to a user u if the frequency of check-ins of a user u to a location l is above the average number of frequency of check-ins for all the check-ins. To find the relevancy for each visited location of users, we developed a user-location matrix where we have assigned each location a binary value depending on the user check-in frequency of that location. [18]

$$\text{Relevancy@N} = \frac{\sum_{i \in L} R(i)}{\text{N}}$$

$$\texttt{R(i)} = 1 \text{ if } i \text{ is relevant and 0 otherwise } \in$$

According to metrics above, L is size of subset and relevance of L is ratio of the size of subset of L divided by the L.

In the second phase, we find if a location is a popular item or not as the unexpectedness of an item is indirectly proportional to popularity. Unlike relevancy, unexpectedness can be defined independently of a user.

To find the unexpectedness we first use popularity criteria and then average rating. Popularity is defined as the ratio between users who rated item i and the sum of all the users. Derived from these criteria, item i is considered unexpected if the popularity score of item i is below the average popularity which was computed for all items. [18]

$$\text{Unexpectedness@N} = \frac{\sum_{i \in L} U(i)}{\text{N}}$$

$$\texttt{U(i)} = 1 \text{ if } i \text{ is unexpected and 0 otherwise}$$

We give each item a rating score according to the criteria derived from the above equation. If the score is greater than the average value we assign 0 for unexpectedness otherwise 1.

An item is serendipitous if it is relevant and unexpected at the same time. To recommend serendipitous locations to users we find similar users for each user from the dataset. Users are considered similar if they have visited at least a certain number of similar locations over time. From similar users, we took only serendipitous locations that the user haven't visited before and we added into recommended locations.

As it is define below, serendipity is the ratio between the size of subset of L who have the serendipitous items, i.e. they are relevant and unexpected at the same time to the size of L. [18]

$$\text{Serendipity@N} = \frac{\sum_{i \in L} S(i)}{\text{N}}$$

$$\texttt{S(i)} = 1 \text{ if } i \text{ is serendipitous and 0 otherwise}$$

4.3 Conclusion

In this section, we have given a detailed overview of our approach to the serendipitous algorithm. We have discussed the basic idea behind the approach and then we have also narrated the flow of the algorithm.

5. Experimental Evaluation

5.1 Introduction

In this section, we have discussed the experimental evaluation of the results. We have used different metrics to compare our results with baseline algorithms.

5.2 Evaluation Protocol

This section aims to validate the metrics results for top-k recommended locations. We aimed to propose a POI recommender algorithm that recommends serendipitous locations to common users based on their surprising element, which we extract from the fusion of relevancy and unexpectedness. Our objective was to find out how many locations from the test set are matched in the POI recommendation. We have used three metrics to validate our results which are Precision and Recall and F-measure. These metrics are commonly used to measure the performance of POI recommendations. We used Precision@N as a ratio between POI matched to the N recommended POI and Recall@N as a ratio between matched POI to count of records in the test set. [12]

5.2.1 Experimental Setup

To perform these operations, we have used a couple of machines to get the results. For local setup, we have used an Intel machine Core i5-8250U CPU @ 1.6GHz (8 CPUs), 1.8GHz, 16GB RAM. For some scenarios, where we needed more processing power to execute multiple baseline algorithms simultaneously, we have also used a university cloud server. We have used the Anaconda environment for development purposes, Jupyter Notebook as an IDE, and Python 3 as the programming language.

5.2.2 Datasets Description

We have used a publicly available real check-in dataset, crawled from Gowalla and the statistics of the datasets are available in Table 1.

Table 1. *Gowalla dataset Statistics*

Nr	**Title**	**count**
1	Number of Users	100
2	Number of locations	6436
3	Number of social links	606

We have split the Gowalla dataset into training and test non-overlapping sets. The training set contains 80% of the data whereas the test set contains 20% of the remaining proportion. we used the training set to train our model for our techniques which we described in the proposed approach and later we used the test set for prediction. In our experiment evaluation, we have used precision, recall, and f-measure over top-k recommendations on user groups of four types 5, 10, and 20.

To find similar users we have set a threshold of 5 similar locations, if two users have visited at least 5 similar locations they will be considered similar to each other.

5.2.3 Evaluation Metrics

We choose three metrics to evaluate our methods: Precision, Recall, and F-measure.To further explain the results we have plot precision and recall and f-measure in to bar charts.

Precision is the ratio between the number of relevant locations over the total recommended location as shown in the equation below:

$$\text{Precision} = \frac{\text{number of relevant locations}}{\text{total recommended locations}}$$

The second metric we used is Recall, the statistical definition of recall is count of true positive divided by the sum of the true positive and true negative. In POI can say the recall is the ratio between the number of relevant locations over the total relevant locations as shown below in equations:

$$\text{Recall} = \frac{\text{number of top-k relevant locations}}{\text{total relevant locations}}$$

Whereas F-Measure combines both recall and precision in a way that it gives a unified measure that possessed both properties.

$$\text{F-measure} = \frac{2 * \text{Precision} * \text{recall}}{\text{Precision} + \text{Recall}}$$

5.2.4 Comparison with competitor models

Overall comparison of our results with some of the baseline algorithms is concerning top-k recommendations on Gowalla datasets.

As you can see in graph figure 9, which shows average precision for the top-5, 10, 20 recommendations for our proposed approach and the other baseline algorithms. The average precision for the serendipitous algorithm is the best among all the 5 models.

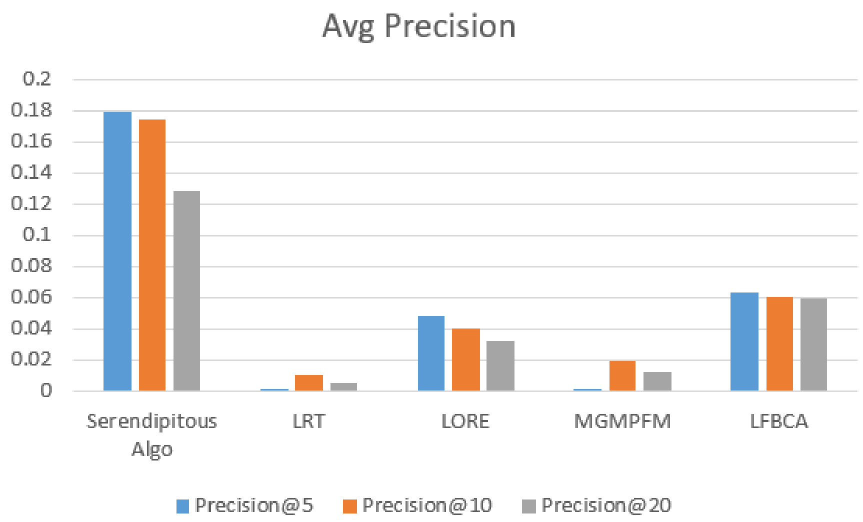

Figure 9. *Average Precision*

Figure 10 shows average recall for the top-k recommendations of baseline and our proposed approach. Recall tells us about the ability of an model to find the relevant cases from given dataset. The average recall for top-20 recommendations is quite impressive as compared to other baseline algorithms whereas average recall value for top-5 and top-10 recommendations, the results are very close to LFBCA. we can drive from the results as the number of recommendations increases the recall value for serendipitous algorithm increases vary rapidly.

From data science concepts, in order to achieve maximum results in one metric we need to trade-off in the other. so when want to achieve maximum precision we are eventually decreasing the recall and vice versa.

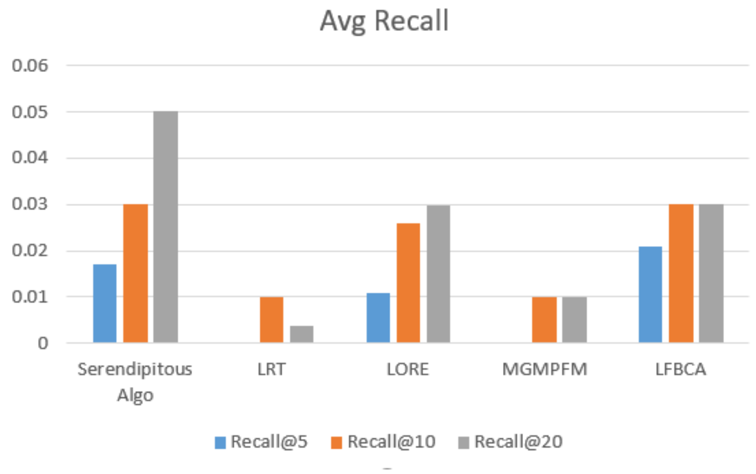

Figure 10. *Average Precision*

By formal definition F-measure is the harmonic mean of precision and recall. It is used to reduce the peak values of both the metrics and it is the most commonly used metric. In order to find the optimal results from precision and recall we use another metric which is the blend of precision and recall. It helps to balance out both the concerns of precision and recall in a single score.

From the figure 11 we can see the serendipitous algorithm had comparatively better f-measure value for almost all the top recommendations.

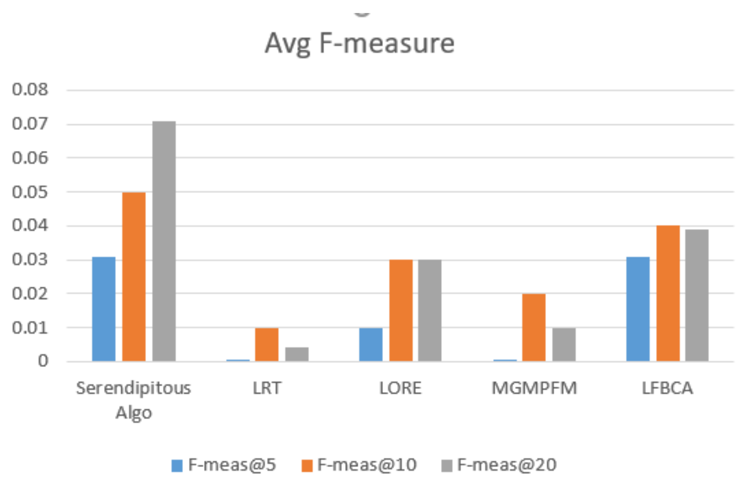

Figure 11. *Average F-measure*

6. Conclusion and Future Work

6.1 Summary

We have proposed an approach to overcome the issue of overspecialization in the POI-based recommender system. we have discussed serendipity and how to relate serendipity in the POI recommender system. we have explained how we can use check-in data to create user-location metrics to recommend new interesting places to users. we try to back up our approach with different evaluation metrics and try to show how our results are better than some of the baseline models.

6.2 Outlook

To further enhance our approach, we can infuse more features or user information to get better results. We can further work on the time efficiency of the algorithm so that we can handle big datasets with ease. In order to verify the results in different dimensions, we can also use some other type of evaluation metrics, for example, Mean Reciprocal Rank, which tries to measure Where to locate the first relevant item. Normalized Distance-based Performance Measure which compares the order of ranking of two lists.

Bibliography

[1] Jinghua Zhu et al. “Friend and POI recommendation based on social trust cluster in location-based social networks”. In: *Eurasip Journal on Wireless Communications and Networking* 2019 (Apr. 2019). DOI: 10.1186/s13638-019-1388-2.

[2] Shenglin Zhao, Irwin King, and Michael Lyu. *A Survey of Point-of-interest Recommendation in Location-based Social Networks*. July 2016.

[3] Xin Technology. “Challenges in recommender systems : scalability, privacy, and structured recommendations”. PhD thesis. Jan. 2015.

[4] Shudong Liu. “User Modeling for Point-of-Interest Recommendations in Location-Based Social Networks: The State of the Art”. In: *Mobile Information Systems* 2018 (Dec. 2017). DOI: 10.1155/2018/7807461.

[5] Denis Kotkov, Jari Veijalainen, and Shuaiqiang Wang. “Challenges of serendipity in recommender systems”. In: *WEBIST 2016: Proceedings of the 12th International conference on web information systems and technologies. Volume 2, ISBN 978-989-758-186-1*. SCITEPRESS. 2016.

[6] Enrico Palumbo et al. “Predicting Your Next Stop-over from Location-based Social Network Data with Recurrent Neural Networks”. In: *Proceedings of the 2nd Workshop on Recommenders in Tourism co-located with 11th ACM Conference on Recommender Systems (RecSys 2017), Como, Italy, August 27, 2017*. Ed. by Julia Neidhardt et al. Vol. 1906. CEUR Workshop Proceedings. CEUR-WS.org, 2017, pp. 1–8. URL: http://ceur-ws.org/Vol-1906/paper1.pdf.

[7] Diana Nurbakova et al. “Itinerary Recommendation for Cruises: User Study”. In: *RecTour 2017: 2nd Workshop on Recommenders in Tourism*. Ed. by Julia Neidhardt et al. Vol. Vol-1906. CEUR Workshop Proceedings urn:nbn:de:0074-1906-6. Como, Italy, Aug. 2017, pp. 31–34. URL: https://hal.archives-ouvertes.fr/hal-01577228.

[8] Quan Yuan et al. “Time-aware point-of-interest recommendation”. In: *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval*. 2013, pp. 363–372.

[9] Yiding Liu et al. "An Experimental Evaluation of Point-of-interest Recommendation in Location-based Social Networks". In: *Proc. VLDB Endow.* 10.10 (2017), pp. 1010–1021. DOI: 10.14778/3115404.3115407. URL: http://www.vldb.org/pvldb/vol10/p1010-liu.pdf.

[10] Jia-Dong Zhang and Chi-Yin Chow. "Exploiting Sequential Influence for Personalized Location-Based Recommendation Systems". In: *Encyclopedia of GIS*. Ed. by Shashi Shekhar, Hui Xiong, and Xun Zhou. Cham: Springer International Publishing, 2017, pp. 565–570. ISBN: 978-3-319-17885-1. DOI: 10.1007/978-3-319-17885-1_1582. URL: https://doi.org/10.1007/978-3-319-17885-1_1582.

[11] Huayu Li et al. "Point-of-Interest Recommendations: Learning Potential Check-ins from Friends". In: Aug. 2016, pp. 975–984. DOI: 10.1145/2939672.2939767.

[12] Chen Cheng et al. "Fused Matrix Factorization with Geographical and Social Influence in Location-Based Social Networks." In: *Aaai*. Vol. 12. 2012, pp. 17–23.

[13] Huiji Gao et al. "Exploring Temporal Effects for Location Recommendation on Location-Based Social Networks". In: *Proceedings of the 7th ACM Conference on Recommender Systems*. 2013, pp. 93–100. ISBN: 9781450324090. URL: https://doi.org/10.1145/2507157.2507182.

[14] Hao Wang, Manolis Terrovitis, and Nikos Mamoulis. "Location recommendation in location-based social networks using user check-in data". In: *Proceedings of the 21st ACM SIGSPATIAL international conference on advances in geographic information systems*. 2013, pp. 374–383.

[15] Q. Guo et al. "A Survey on Knowledge Graph-Based Recommender Systems". In: *IEEE Transactions on Knowledge Data Engineering* 01 (Oct. 5555), pp. 1–1. ISSN: 1558-2191. DOI: 10.1109/TKDE.2020.3028705.

[16] B. Liu et al. "A General Geographical Probabilistic Factor Model for Point of Interest Recommendation". In: *IEEE Transactions on Knowledge and Data Engineering* 27.5 (2015), pp. 1167–1179. DOI: 10.1109/TKDE.2014.2362525.

[17] Panagiotis Adamopoulos and Alexander Tuzhilin. "On Unexpectedness in Recommender Systems: Or How to Better Expect the Unexpected". In: *ACM Transactions on Intelligent Systems and Technology* 5 (Jan. 2014), pp. 1–32. DOI: 10.1145/2559952.

[18] Marco de Gemmis et al. "An investigation on the serendipity problem in recommender systems". In: *Information Processing Management* 51.5 (2015), pp. 695–717. ISSN: 0306-4573. DOI: https://doi.org/10.1016/j.ipm.

2015.06.008. URL: http://www.sciencedirect.com/science/article/pii/S0306457315000837.

Appendices

Appendix 1 - Non-exclusive licence for reproduction and publication of a graduation thesis

I Muhammad Nadeem

1. Grant Tallinn University of Technology free licence (non-exclusive licence) for my thesis “POI-based Serendipitous Recommender Algorithm”, supervised by Prof Dr Sadok Ben Yahia.
   1.1. to be reproduced for the purposes of preservation and electronic publication of the graduation thesis, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright;
   1.2. to be published via the web of Tallinn University of Technology, incl. to be entered in the digital collection of the library of Tallinn University of Technology until expiry of the term of copyright.
2. I am aware that the author also retains the rights specified in clause 1 of the non-exclusive licence.
3. I confirm that granting the non-exclusive licence does not infringe other persons' intellectual property rights, the rights arising from the Personal Data Protection Act or rights arising from other legislation.

05.01.2021