UNIVERSITI TEKNIKAL MALAYSIA MELAKA

Faculty of Information and Communication Technology

A COLLABOTIVE FILTERING RECOMMENDER SYSTEM FOR INFREQUENTLY PURCHASED PRODUCT USING SLOPE-ONE ALGORITHM AND ASSOCIATION RULE MINING

Nur Azleen Binti Zolhani

Master of Computer Science (Database Technology)

2015

A COLLABOTIVE FILTERING RECOMMENDER SYSTEM FOR INFREQUENTLY PURCHASED PRODUCT USING SLOPE-ONE ALGORITHM AND ASSOCIATION RULE MINING

NUR AZLEEN BINTI ZOLHANI

A thesis submitted
in fulfillment of the requirements for the degree of Master of Computer Science
(Database Technology)

Faculty of Information and Communication Technology

UNIVERSITI TEKNIKAL MALAYSIA MELAKA

2015

DECLARATION

I declare that this thesis entitled “A Collaborative Filtering Recommender System for Infrequently Purchased Product using Slope-One Algorithm and Association Rule Mining” is the result of my own research except as cited in the references. The thesis has not been accepted for any degree and is not concurrently submitted in candidature of any other degree.

Signature : ....

Name : Nur Azleen Binti Zolhani

Date : ....

APPROVAL

I hereby declare that I have read this thesis and in my opinion this thesis is sufficient in terms of scope and quality for the award of Master of Computer Science (Database Technology)

Signature : ....

Supervisor’s Name : Dr. Noraswaliza Binti Abdullah

Date : ....

DEDICATION

With all due respect, I dedicate this dissertation done by myself to both of my parents

"Zolhani bin Shaari" and "Sofiah binti Hashim"

who are the pillar of my strength and

who are always praying for my success and well achievement.

Nevertheless, I dedicate this dissertation to my younger brother

"Nur Asyraf bin Zolhani"

who are always there for me to unwind my stress and my worries.

Lastly, I dedicate this dissertation to "Saifullah Aizzat bin Abdullah"

who has undying faith in me and who always brushes off my doubts.

ABSTRACT

Nowadays, tourism industry are actively being utilised in generating a state or country income. In order to attract tourist from all over places, information conveyance is important. Traditionally, people travels to certain places based on oral recommendation by families and friends. Now, people tends to go travel based on reviews that are read from blogs and websites. But, this leads to overflow of unfiltered information. In order to effectively recommending places to travel for tourist, recommendation engine are being developed. Most recommendation engine has suffice information to make recommendation for example Amazon.com recommendation and Google.com recommendation. Meanwhile, in tourism it is quite challenging in making recommendation because hotels are occasionally being booked or purchased by consumer. This is due to the fact that travelling are expensive and time consuming. This project implement the collaborative filtering using slope-one algorithm and also implement association rule mining in recommending hotels for tourist. This recommender system uses slope-one algorithm whereby it accumulate and takes into account of the difference in popularity. The objective of this project to study different types of recommendation techniques for infrequently purchased products and to investigate technique and dataset that are suitable to implement in recommending infrequently purchased products. As a conclusion, this collaborative filtering recommendation system will help user in decision making. Further research on other approaches in implementing recommender system in tourism domain can help in information delivery.

ABSTRAK

Pada masa kini, penglibatan secara aktif industri pelancongan dalam menjana pendapatan negeri atau negara semakin meningkat. Dalam usaha menarik pelancong dari seluruh tempat, penyampaian maklumat adalah penting. Secara tradisinya, pelancong akan bercuti ke tempat tertentu berdasarkan cadangan lisan oleh keluarga dan rakan-rakan. Sekarang, pelancong lebih cenderung untuk bercuti berdasarkan ulasan yang dibaca dari blog dan laman web. Tetapi, ini membawa kepada limpahan maklumat yang tidak ditapis. Untuk mengesyorkan tempat-tempat menarik untuk pelancong, enjin cadangan (*recommendation engine*) telah dibangunkan. Kebanyakan enjin cadangan mempunyai maklumat yang memadai untuk membuat saranan sebagai contoh Amazon.com dan Google.com. Sementara itu, dalam industri pelancongan ianya agak mencabar untuk membuat cadangan contohnya seperti cadangan hotel akan ditempah oleh pengguna dalam masa berkala. Ini adalah disebabkan oleh aktiviti melancong atau bercuti melibatkan bajet yang tinggi dan memakan masa. Projek ini melaksanakan pembangunan sistem cadangan (*recommendation system*) atas perbezaan populariti antara item dalam mengesyorkan hotel untuk pelancong. Sistem cadangan ini menggunakan *slope-one algorithm* di mana ia mengumpul dan mengambil kira perbezaan dalam populariti. Objektif projek ini adalah untuk menyiasat algoritma yang sesuai yang sesuai dengan tingkah laku pengguna dalam membeli item jarang dibeli. Kesimpulannya, Kedudukan item ke item sistem cadangan ini akan membantu pengguna dalam membuat keputusan. Penyelidikan lanjut mengenai pendekatan yang lain dalam melaksanakan sistem cadangan (*recommender system*) dalam domain pelancongan boleh membantu dalam penyampaian maklumat.

ACKNOWLEDGEMENT

In the name of Allah SWT, the most Gracious and most Merciful. Alhamdulillah, all praises to Allah for His blessing; and the strength granted to me physically and mentally.

First and foremost, I would like to express my appreciation towards my supervisor Dr. Noraswaliza binti Abdullah for her assistances and guidance in completing this project. Without her guidance, I am unable to finish this project successful. Also, I would like to thank all panels, who had given me suggestion in improving my project.

Next, I would like to thank my fellow friends especially Nor Rabbiatun binti Mohamed Nazri for her advice and also the resources that she had shared with me in helping me completing this project and dissertation. Also, I would like to express my gratitude towards Aiza Syahida binti Zakaria for answering most of my uncertainties regarding my project and dissertation despite being busy working. Also, thank you to Mohd. Zikre bin Ahmad Puad for his technical support in installing most of the software that I used while completing this project.

Not forgetting, I would like to thank my siblings Nur Nazreen bin Zolhani and Nur Hariz bin Zolhani who are responsible in taking care of my five cats at home, which are my bundle of joy.

Last but not least, I would like to express my greatest gratitude towards both of my parents for their undying support. Thank you for always praying for my success and listening to my doubts and worries while I am completing my master degree. I know that I am disorganised and confused, but thank you for not neglecting me despite the fact that you are sick and are physically unable to be with me while I am completing my master degree.

TABLE OF CONTENT

CHAPTER 4

CHAPTER 5

CHAPTER 6

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICE

LIST OF ABBREVIATION

RS	-	Recommender System
CBF	-	Content Based Filtering
CF	-	Collaborative Filtering
KBF	-	Knowledge Based Filtering
PL/SQL	-	Procedural Language/Structural Query Language
JSON	-	JavaScript Object Notation
RestAPI	-	Representational State Transfer API
RMSE	-	Root Mean Square Error
MAE	-	Mean Absolute Error

CHAPTER 1

INTRODUCTION

Tourism industry has become a backbone for supporting a country or a city economic status and solely depends on the uniqueness of place to generate incomes. This is mainly due to tourists that seek for serenity and clarity from vacations as a brief hiatus from the hustle and bustle of their daily routine. It is important that the uniqueness of place offer by different countries or cities gain recognition and this depends on the information describing that particular places. Information that encapsulate the places for tourist attractions commonly are the accommodation in terms of places to stay, comfortability, prices, mobility and most importantly, uniqueness. This information varies to different users as a person interest is a subjective matter and to serve this purpose, the information delivers must heed the user preference. Thus, tourist normally search for recommendations to satisfy their needs from web pages.

A recommender system (RS) is a computer application that helps user in making decisions which items to choose from a vast amount of information available on the internet. A user can be indecisive due to information overload on the Web especially tourism information as nowadays tourism industry has become a competitive business. Many commercial recommender systems have been developed for frequently purchased products such as books, movies and songs, as a large amount of ratings data available to be utilised by a recommender system. However, for infrequently purchased items like tourism services, the information delivers may be biased as a large amount of rating data of places to be visited by the tourist are unavailable. This situation may be caused by some users are not frequently purchase items in tourism industry. In addition, users are unwilling to input ratings data and reviews data. Furthermore, the current research focus on frequently purchased products as the dataset is publicly available.

The objectives of this study are to study different types of recommendation techniques for infrequently purchased products and to investigate technique and dataset that are suitable to implement in recommending infrequently purchased products such as Content-based recommendation technique, Collaborative-Filtering recommendation technique, Hybrid recommendation technique, and lastly Knowledge-based recommendation technique. Next objectives are to develop a prototype in recommending hotels to user using user's action view and user rating data. By conducting this study, a recommendation technique can be applied in various domain including for infrequently purchased products and the last objective is to evaluate the proposed recommendation techniques using user testing evaluation

1.1 Background of study

According to the Oxford Dictionaries website, recommendation is a behaviour of providing a proposition either by a person or an authoritative organizations into subliminally influencing another person into taking the most beneficial options. Whereas a recommender system is a program or application that gives it user a counsel which subconsciously dominates the user decisions. Recommender systems are embedded with corresponding techniques that implied to certain algorithms which mostly are from a field of study known as data mining.

Each recommender system is different from one another as each of it may implement different recommendation techniques. These recommendation techniques can be divided into two categories, which are traditional recommendation techniques and modern recommendation techniques (Wanaskar et al. 2013). The traditional recommendation techniques can be categorized as follows which are, Content-Based recommendation technique, Collaborative Filtering recommendation technique, Hybrid recommendation technique and lastly Knowledge-Based recommendation technique. Meanwhile, the modern recommendation techniques can be divided into four which are Context-Aware recommendation technique, Semantic-Based recommendation technique, Peer-to-Peer recommendation technique and lastly Matrix Factorization recommendation technique.

Whereas according to Bobadilla et al. (2013), there are four filtering algorithm, namely content-based filtering, collaborative filtering, demographic filtering, and lastly hybrid filtering. The content-based filtering is recommendation filtering based on the active user purchasing history. Meanwhile, the collaborative filtering is recommendation based on other user's preferences that similar to the active user likings. The demographic filtering concerned on the likings based on demographic similarity, for example age or gender. Lastly, hybrid filtering is recommendation based on combining two filtering algorithms, for example combination between content-based filtering and collaborative filtering or combination between demographic filtering and collaborative filtering. Another filtering algorithm is the rating-based item-to-item algorithms. According to Lemire & Mcgrath (2013) this which is the slope-one algorithm involve in predicting the similarity differences in item rather than user similarity.

Nowadays, the current research focuses on frequently purchased products such as books, movies, and songs because this product are inexpensive thus it is frequently purchased by consumer. Also, for frequently purchased product the accumulation of rating data and reviews are high in number. Thus, making recommendation for frequently purchased product more efficient. Recently, researchers are exploiting knowledge-based recommendation techniques for infrequently purchased products. However, this technique is more applicable in case-based reasoning, meaning it requires a profound knowledge of the product domain in making recommendation to users.

1.2 Problem statement

Due to tourism information overflow, tourists are being surrounds by invalid information that indirectly hinders the legitimate tourism information from being delivered. This situation can be troublesome to tourists as traveling does cost a fair amount of money and a right decision can help them save up and thus helping them in making travel expenditure beforehand. This study is motivated by several problems which are:

i. Knowledge-based technique in recommender system for infrequently purchased products or items require extensive knowledge reengineering process.

ii. Most of rating data for infrequently purchased products are unavailable because tourists rarely stayed or rarely check-in to the hotel

iii. The current recommender system for infrequently purchased products or items requires substantial user involvement. However, tourists are unwillingly to provide input such as reviews or ratings which are important in providing recommendations.

1.3 Research Questions

Based on the problem statement above, there are several research questions that stimulates this study, which are:

i. Which of the recommendation techniques can assist in recommending the infrequently purchased products or items. For example, in tourism industry which hotel best fit the tourist interest and should be recommended?

ii. Which data from the information collected can be utilized to recommend infrequently purchased products without requiring extensive users' involvement?

iii. Which algorithm best suited with the chosen recommendation techniques to optimize recommendations of the infrequently purchased products or items.

1.4 Research Objective

Motivated by the problem statements and the research questions stated in the previous section, this study has three known objectives which are:

i. To study different types of recommendation techniques for infrequently purchased products and to investigate technique and dataset that are suitable to implement in recommending infrequently purchased products.

ii. To developed a prototype in recommending hotels to user using user's action view and user rating data.

iii. To evaluate the proposed recommendation techniques using user testing evaluation.

1.5 Research Scope and Limitation

This study requires a collection of data for infrequently purchased products or items in the tourism industry. Most of recommender systems are developed for recommending the frequently purchased products or items such as books, movies and songs. Tourism data can be considered as volatile as the tourism industry is a seasonal business, its information may varies within a year. Due to its vast amount of information, there are no known dataset in tourism industry being recorded that can be freely downloaded on the Internet to be analysed.

Besides, the information gathered may be exaggerated as the places with tourist attractions are competing in promoting their businesses regardless of the true nature of the place. To validate this information it is important to investigate the information of the rating data alongside the reviews and also taking into account the user or tourist preferences. Keep in mind that the effectiveness of recommendation techniques depends on the validity of the data.

The research scope of this study are:

i. Using a collaborative recommendation technique compose of slope-one algorithm for recommending the infrequently purchased products or items.

ii. Using association rules mining in recommending view, buy or rate action done by user while browsing through the webpages.

1.6 Significant and Research Contribution

This study result in a significant contribution towards recommending infrequently purchased product or item such as recommending tourist a hotel to stay according to their preferences. Hence, the tourists can have a reliable source of information. This project will develop a prototype and are being tested by dedicated users or participants.

1.7 Organization of the Thesis

This study is documented into dissertation that made up of five chapters and are structurally compose accordingly to as follows:

i. Chapter 1: Introduction

The first chapter of this dissertation is the introductory of the study which compromises of background of study that briefly explained the chronology of the study in terms of what does lead this study, lists of problem statements that can be identified, the objectives of the study which define how to solve the problems, the study research questions which motivated the study, scope and limitation of the study, and lastly the contribution of the study. This chapter provides enlightenment to reader of the overall concept of the study.

ii. Chapter 2: Literature Review

The second chapter of this dissertation is the literature review. Literature review is a documented review or critic of related work that similarly serve the same purpose as the study. In this dissertation, the literature review focuses on different recommender systems that uses several recommendation techniques respectively. This differences are then tabulated for better understanding. The purpose of literature review is to explore different kinds of techniques available for developing the best suited recommender systems for tourism industry.

iii. Chapter 3: Research Methodology

The third chapter of this dissertation is the methodologies used in conducting the research in order to satisfy the objectives of the study. This chapter consists of the type of research method, the chronology of the research design, proposed training framework, and lastly list of research tools use while conducting this research. This chapter gives reader the insight of how the study is conducted.

iv. Chapter 4: Experimental Result

The fourth chapter of this dissertation is the experimental result. This chapter is documented after the experiments has been carried out. This chapter consist of fact and figures that strengthen and defined the objectives of this study. The results are presented in manners of tables, graphs and other form of illustrated representation for the ease of reader to understand.

v. Chapter 5: Conclusion and Future Work

The fifth and last chapter of this dissertation is the conclusion and suggestion to future work. This chapter discuss on the conclusion in reference to the results from this study on whether this study has achieve its objective and also define its purpose or otherwise stated. On the other hand, this chapter provides several suggestions in order to improve this study for future works enhancement.

The next chapter of this dissertation is Chapter 2, Literature Review that elaborates on related works that similar in objective to this study.