ITEM-BASED COLLABORATIVE FILTERING TECHNIQUE FOR MOVIE RECOMMENDER BASED ON USER PREFERENCES

WONG WAI LEONG

BACHELOR OF COMPUTER SCIENCE

UNIVERSITI MALAYSIA PAHANG

Universiti
Malaysia
PAHANG
Engineering • Technology • Creativity

SUPERVISOR’S DECLARATION

I hereby declare that I have checked this thesis and in my opinion, this thesis is adequate in terms of scope and quality for the award of the degree of Bachelor of Computer Science in Software Engineering.

____

(Supervisor’s Signature)

Full Name : Dr Awanis binti Romli

Position :

Date :

Universiti
Malaysia
PAHANG
Engineering • Technology • Creativity

STUDENT’S DECLARATION

I hereby declare that the work in this thesis is based on my original work except for quotations and citations which have been duly acknowledged. I also declare that it has not been previously or concurrently submitted for any other degree at Universiti Malaysia Pahang or any other institutions.

____

(Student’s Signature)

Full Name : WONG WAI LEONG

ID Number : CB15027

Date :

ITEM-BASED COLLABORATIVE FILTERING TECHNIQUE FOR MOVIE RECOMMENDER BASED ON USER PREFERENCE

WONG WAI LEONG

Thesis submitted in fulfillment of the requirements
for the award of the degree of
Bachelor of Computer Science in Software Engineering

Faculty of Computer Systems & Software Engineering

UNIVERSITI MALAYSIA PAHANG

DECEMBER 2018

ACKNOWLEDGEMENTS

I am deeply grateful to my supervisor, Dr Awanis binti Romli who provided her expertise opinion and motivation to help me along the way to complete my final year project.

I would like to give my appreciation to those lecturers for sharing their knowledge during the course of research. Lastly, I want to express my gratitude to all my friends who willing to help me when I face problem during the process of carry out this research.

In additional, I would like to say thank to my parents and sisters who always support me on this journey.

ABSTRAK

Dengan kedatangan web 2.0, kebanyakan orang mula melakukan perkara seperti pembelian product, tempahan barangan melalui internet. Dengan wujudnya internet, ia membawa senang kepada kehidupan kita. Walaubagaimanapun, terdapat beberapa masalah wujud seperti pilihan yang banyak boleh dipilih di internet menyebabkan pengguna mengalami masalah dilema. Di samping itu, dengan bantuan sistem rekomen, pengguna dapat melihat dan menikmati filem mengikut keutamaan mereka dan mengurangkan masa untuk mencari filem yang sesuai. Oleh itu, teknik penapisan kolaborasi akan digunakan dalam pemilihan filem untuk menganalisis hubungan antara pengguna dan filem berdasarkan rating dan penandaan yang diberikan oleh pengguna lain. Teknik penapisan kolaborasi berasaskan item sesuai digunakan untuk rekomen filem sambil menjana cadangan kepada pengguna kerana teknik ini mengira kesamaan antara filem dan filem. Skema yang dicadangkan akan dilaksanakan dalam Anaconda dan mengunakan python. Keputusan akan dinilai pada akhir penyelidikan. Keputusan ahkir bagi penyelidikan ini adalah filem teratas 10 akan dihasilkan untuk pengguna berdasarkan rating yang dihasilkan olen pengguna lain. Bagi pengguna yang baharu, dia diperlukan memasukkan jenis filem yang mereka suka. Selepas itu, senarai baru filem teratas 10 akan dihasilkan berdasarkan keutamaan pengguna tertentu.

ABSTRACT

With the advent of web 2.0, most of the people start using internet to do the thing for example purchase product through online, online booking stuff and we can say that with internet, it bring convenient to our living. Although internet bring so much convenient for us but there are some problem occur when there are too many choices can be choose on internet. This problem will caused dilemma happened on user. Besides, with the help of the recommender system, user can noticed and enjoyed the movie according to their preference and time consuming while searching the movie will be reduced. Therefore, item-based collaborative filtering technique is apply on online movie in order to analyses the relationship between user and movie based on the rating and tagging provided by user. Item-based collaborative filtering technique is suitable used in movie recommender while generate recommendation to user because it calculate the similarity between movie and movie. The proposed schema will be implement in python language on Anaconda and result will be evaluated in the end of research. The final result of this research is top-10 movies will be generated for user based on the highest rating provided by the previous user. For new user, he/she can enter the type of the movie they may like in the column. After that, another top-10 movies list will be generated for the new user based on the specific preference of the particular user.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

TF	Term frequency
IDF	Inverse document frequency
HDFS	Hadoop File System
ANN	Artificial Neural Network
MAE	Mean Absolute Error
CF	Collaborative filtering
CBF	Content-based filtering

CHAPTER 1

INTRODUCTION

1.1 Introduction

Recommender system are defined as decision making strategy for user under complex information environment and help user discover item that they may like interest (Anastasiu, Christakopoulou, Smith, Sharma, & Karypis, 2016).Nowadays, recommender systems are ubiquitous anywhere and even change our lives in many ways for example E-commerce, education, and social media. Recommender system are information filtering systems that deal with the problem of information overload by filtering important information fragment out of large amount of dynamically generated information according to user to provide them a personalised content (Isinkaye, Folajimi, & Ojokoh, 2015).

As Big Data is the driving force behind recommender system, In the Big Data, there are a lot of user information such as browsing history, rating and feedback for recommender system generate relevant and useful information to user. Recommender system cannot make recommendation accurate if there is insufficient data (Pal, 2015).

A typical recommender system works in well-defined and logical. There are three phases of how recommender system work which are data collection, rating and filtering (Pal, 2015).When user is surfing a website and read information by clicking on a link, an event such as Ajax event could be fired. Different technology may make change on different event. Usually data will be stored in database such as NoSQL database. When user logged in, details will be extracted from system cookies or HTTP. User details will be captured and read in layman‘s language, for example the data could be read something like "User X clicked Video Z once" and data will stored for future recommendation.

Rating is one of the important sense to reflect a user on their feeling toward the product. Action taken by user is very important information for recommender system.

Recommender system will assign implicit based on action taken by user. For example, user A watched movie X and movie Y, and user A give high rating to both movie. So, when user B does watch on movie X, then the system will suggest movie B which data collected from user A.

Recommender system use three types of filtering to filter product based on rating and data gained from other users. Example of filtering technique are collaborative filtering technique, content-based filtering technique and hybrid filtering technique. In collaborative filtering, user choice will compared with other user before suggest any recommendation. For example if user A watch movie X,Y and user B watch movie X,Y,Z, user A will be recommend movie Z too because both user have a lot of common behaviour.

There are reason why movie website concern about their recommender system which to help user to find the movie based on their perception and experience from other users. Recommender system will help user save their time from searching movie out of thousand type of movie list in a quality, efficient, and effective way.

1.2 Problem Statement

Table 1.1 shows the research problem with description and effect of these problem.

Table 1.1 Research problem

No	Problem	Description	Effect
1.	We may not know what we want	Users may not know what they really want since there are plenty of movie on the movie web.	There are many choices can be choose by users and it will caused them suffering from decidophobia.
2.	Lack of chance to explore interest movie that related to user preference	Users may not notice that there are others interested movie on the movie web which close to their preference.	User only look at the movie they want but actually there are others nice and interested movie they may interest too.
3.	Time consuming	Users may click many button to search the product information	Waste of time to looking for movie.

1.3 Objectives

The aim of this research is apply item-based collaborative filtering technique to generate a list a movie for user based on their preference. In order to achieve this aim, there are few objectives are necessary in this research such as:

i. To study the strength and weakness of the existing recommender filtering techniques and impact to the user.

ii. To apply the item-based collaborative filtering technique while making prediction on user profile based on their preference on finding movie.

iii. To evaluate item-based collaborative filtering technique in predicting user preference on finding movie.

REFERENCES

Adomavicius. (2005). *A survey of the state-of-the-art and possible extensions.*

Adomavicius, & Z. J. (2012). *Impact of data characteristics on recommender systems performance.*

Adomavicius, G. (2005). *Toward the next generation of recommender system.*

Adomavicius, G. (2005). *Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extension.*

Anastasiu, D. C., Christakopoulou, E., Smith, S., Sharma, M., & Karypis, G. (2016). *Big Data and Recommender Systems.*

B. Sarwar, G. Karypis, J. Konstan, & J. Reidl. (2001). *Item-based collaborative filtering recommendation algorithms.*

Big Data Appliance Software User's Guide. (n.d.). Retrieved from https://docs.oracle.com/cd/E65728_01/doc.43/e65665/GUID-69A231EA-0F4D-4460-84A5-5F8DF105A2EE.htm#BIGUG132

Bin, W. (2018). Python Implementation of Baseline Item-Based Collaborative Filtering.

Breese, J. S., & D. Heckerman. (1998). *Empirical analysis of predictive algorithms for collaborative filtering.*

C, P., & Li W. (2010). *Recommendation with topic analysis. In Computer Design and Applications.*

Claypool, M. (2001). *Implicit Interest Indicators.*

D, B. (1999). *hybrid user model for news story classification.*

DW, O., & Kim J. (2000). *Implicit feedback for recommender systems.*

Ernest, G. (2015). *Building a Recommendation Engine with Spark ML on Amazon EMR using Zeppelin*. Retrieved from https://aws.amazon.com/blogs/big-data/building-a-recommendation-engine-with-spark-ml-on-amazon-emr-using-zeppelin/

Felfernig, A. (2008). Constraint-based Recommender Systems: Technologies and Research Issues.

G.Adomavicius, & A., T. (2005). *Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions.*

Garcia S, L. J. (2015). Data Preprocessing in Data Mining. *Big data preprocessing: methods and prospects*.

HC, Y. (2002). *A personalized recommender system based on web usage mining and decision tree induction.*

Hosseini-Pozveh. (2009). *A multimedia approach for context-aware recommendation in mobile commerce.*

Isinkaye, F., Folajimi, Y., & Ojokoh, B. (2015). *Recommendation systems: Principles, methods and evaluation.*

J, B. (2009). *Empirical analysis of predictive algorithms for collaborative filtering.*

J, B., & Schwind C. (2012). *Learning with personalized recommender systems.*

JA, K., & Riedl J. (2012). *Recommender system (from algorithms to user experience).*

JA, M. (2004). *Data mining technique.*

JAIN, A. (2016). Build a Recommendation Engine in Python.

JB, S., Konstan J, & Riedl J. (1999). *Recommender system in e-commerce.*

JL.Herlocker. (2004). *Evaluating collaborative filtering recommender systems.*

Larose. (2005). *Discovering knowledge in data.*

Linden. (2003). *Amazon.com recommendation: item to item collaborative filtering.*

M.Shaw. (2001). *A Survey of Recommendation System in Electronic Commerce.*

McSherry. (2004). *Explaining the pros and cons of conclusions in CBR.*

Miranda. (1999). *Combining content- based and collaborative filters.*

Nkandu, J. (2014). Building and Evaluating an Adaptive Real-time Recommender System.

Pal, K. (2015). *How Big Data is used in Recommendation Systems to change our lives.*

Pan. (2010). *analysis. In Computer Design and Applications.*

R, B. (2002). *Hybrid recommender systems: survey and experiments. User Model User-adapted Interact.*

Regi, A. N. (2013). *A Survey on Recommendation Techniques .*

Resnickk, & Varian. (2009). *Recommender system.*

Rodriguez, G. (2018). Introduction to Recommender Systems in 2018.

Sadeghian, A. (2014). Recommender Systems in E-Commerce.

Salton, G. (1999). *Automatic text processing: the transformation, analysis.*

Sarwar, B. (2000). *Analysis of recommendation algorithms for e-commerce.*

Sarwar, B. (2001). *Item-based Collaborative Filtering Recommendation.*

Sarwar, B., & Geoge Karypis. (2010). *Analysis of recommendation algorithm for e-commerce.*

Sarwar, B., G. Karypis, & G. Konstan. (2001). *Item-based collaborative filtering recommendation algorithms.*

SC, G., & Lhuillier N. (2007). *Addressing uncertainty in implicit preferences.*

Schafer, B. (1999). *Recommender systems in ecommerce.*

Sivapalan, S. (2014). Recommender Systems in E-Commerce.

Smyth B. (2000). *personalized TV listings service for the digital TV age.*

Sun, Z., & N. Luo. (2010). *A new user-based collaborative filtering algorithm combining data-distribution.*

Szerovay, G. (2018). *Why you need Python environments and how to manage them with Conda*.

Team, D. (2016). *Apache Spark vs. Hadoop MapReduce*. Retrieved from https://data-flair.training/blogs/apache-spark-vs-hadoop-mapreduce/