Online Product Recommendation System

Submitted By

Nikita C. Sarang

15MCEC22

DEPARTMENT OF COMPUTER ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2017

Online Product Recommendation System

Major Project

Submitted in partial fulfillment of the requirements

for the degree of

Master of Technology in Computer Engineering

Submitted By

Nikita C. Sarang

(15MCEC22)

Guided By

Prof. Jaladhi Vyas

DEPARTMENT OF COMPUTER ENGINEERING

INSTITUTE OF TECHNOLOGY

NIRMA UNIVERSITY

AHMEDABAD-382481

May 2017

Certificate

This is to certify that the major project entitled **"Online Product Recommendation System"** submitted by **Nikita C. Sarang (Roll No: 15MCEC22)**, towards the partial fulfillment of the requirements for the award of degree of Master of Technology in Computer Engineering of Nirma University, Ahmedabad, is the record of work carried out by him under my supervision and guidance. In my opinion, the submitted work has reached a level required for being accepted for examination. The results embodied in this major project part-I, to the best of my knowledge, haven't been submitted to any other university or institution for award of any degree or diploma.

Prof. Jaladhi Vyas
Guide & Associate Professor,
CE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr. Priyanka Sharma
Associate Professor,
Coordinator M.Tech - CE
Institute of Technology,
Nirma University, Ahmedabad

Dr. Sanjay Garg
Professor and Head,
CE Department,
Institute of Technology,
Nirma University, Ahmedabad.

Dr. Alka Mahajan
Director,
Institute of Technology,
Nirma University, Ahmedabad

Certificate

This to certify that **Miss. Nikita Sarang(15MCEC22)**, a student of M.Tech CSE, Institute of Technology, Nirma University, Ahmed- abad is working with this organization, MLveda since 26/05/2016 and carried out her thesis work titled **Online Product Recommendation System**. The results embodied in this project, to the best of our knowledge, havent been submitted to any other university or institution for award of any degree or diploma.

We wish her all the best for his bright future.

Please do not hesitate to contact us if you required any further information.

Mr. Amrish Patel
CEO,
Mlveda,
Ahmedabad.

Statement of Originality

I, **Nikita C. Sarang**, Roll. No. **15MCEC22**, give undertaking that the Major Project entitled "**Online Product Recommendation System**" submitted by me, towards the partial fulfillment of the requirements for the degree of Master of Technology in **Computer Engineering** of Institute of Technology, Nirma University, Ahmedabad, contains no material that has been awarded for any degree or diploma in any university or school in any territory to the best of my knowledge. It is the original work carried out by me and I give assurance that no attempt of plagiarism has been made.It contains no material that is previously published or written, except where reference has been made. I understand that in the event of any similarity found subsequently with any published work or any dissertation work elsewhere; it will result in severe disciplinary action.

Signature of Student
Date:
Place:

Endorsed by
Prof. Jaladhi Vyas
(Signature of Guide)

Acknowledgements

It gives me immense pleasure in expressing thanks and profound gratitude to **Prof. Jaladhi Vyas**, Associate Professor, Computer Department, Institute of Technology, Nirma University, Ahmedabad for his valuable guidance and continual encouragement throughout this work. The appreciation and continual support he has imparted has been a great motivation to me in reaching a higher goal. His guidance has triggered and nourished my intellectual maturity that I will benefit from, for a long time to come.

It gives me an immense pleasure to thank **Dr. Sanjay Garg**, Hon'ble Head of Computer Engineering Department, Institute of Technology, Nirma University, Ahmedabad for his kind support and providing basic infrastructure and healthy research environment.

A special thank you is expressed wholeheartedly to **Dr. Alka Mahajan**, Hon'ble Director, Institute of Technology, Nirma University, Ahmedabad for the unmentionable motivation he has extended throughout course of this work.

I would also thank the Institution, all faculty members of Computer Engineering Department, Nirma University, Ahmedabad for their special attention and suggestions towards the project work.

- Nikita C. Sarang
15MCEC22

Abstract

In today's digital world, day to day number of customer drastically increases in online shopping. There are a huge amount of customer's information and product's information available to maintain. This information is very useful to increases sales revenue. The system is required which has filter to filter out these information and efficiently provide relevant recommendation in order to reduce the problem of information overload and increased Internet traffic. Online recommendation systems act as a virtual agent which help user to take up right product from the abundant amount of products purchasable on the e-commerce site by providing an effective recommendation. Reviews or ratings provided by user, used to build up product profile and user navigation is used to build up user profiles, both are used for recommending products that best matches with the user's interest. Recommendation systems identify recommendations automatically for individual buyers based on past purchases and searches, product rating and on other users behavior. This paper includes advantages and limitations of recommendation system and detail description of all techniques which are used for recommendation with its pros and cons. User-based and Item-based techniques are very popular for recommendation, we have discuss its algorithm with complexity analysis and quantitative analysis.

Abbreviations

RS	Recommender System.
MAE	Mean Absolute Error

–

Contents

List of Figures

Chapter 1

Introduction

1.1 Online Product Recommendation System

In today's digital world, day to day amount of available digital information increases wildly and the use of the Internet is going to be uncontrollable. In future, this will lead to the problem of information overload and increase Internet traffic. To avoid this situation we should have efficient information retrieval system which will automatically find relevant content and deliver to the user. On online store, there are crores of products available. Some products are similar but not exactly similar in term of their property and quality. At this point, a user gets confused in choosing the correct product from its large amount variety.

Information retrieval system, such as Google have partially solved this issue but it will give suggestions based on keyword used in the query because Google is a keyword based search engine but personalization and prioritization have been missing. This digital world needs an intelligent system which gives an experience of personalization and prioritization. This has increased the demand for recommendation systems. Recommendation system filters information from a large amount of information by a user's interest and preferences and then it will provide a meaningful recommendation. This leads to reducing jam of information overload and unnecessary Internet traffic too. For an example on Netflix user provide ratings to movies. Such data used to build up product profile. Then find most similar user or the item for the recommendation. Amazon records users search history based on that it will find the most similar the item for the recommendation.

In today's e-commerce world recommendation system act as virtual intelligent agent on online store by providing personalization to the user. Using recommendation system user can take up right product from the abundant amount of products purchasable on the e-commerce site [1]. Based on the user's profile or product profile RS has the power to predict whether a particular user would prefer a product or not. RS are advantageous to both business providers and users. They reduce successive search costs of finding products on an online shopping environment. RS support users by allowing them to move beyond catalog searches.

1.2 Benefits of Recommendation System

- **Better understand user's interest** RS analyzes the user's search history from which it finds user's interest and accordingly provides meaningful recommendations.

- **Engage user for a long time** RS provides personalization and prioritization to the user. They will quickly find a product without having to perform a successive search. So users spend more time on site as they served according to their interest by RS.

- **Increases number of customers/users** RS provides personalization the experience to the user by recommending products of their interest. This shows their users what they value as an individual. Therefore users easily find what they want.

- **Increases average order value** RS provides a dynamic experience to the user by personalization. Hence user will spend more time site. RS helps the user a to purchase correct product from thousands of products. This will stop escaping user from the site. By recommending a relevant product to the user, a user will quickly buy products which he/she wants.

- **Increases the number of items per order** RS will find similar products for the user and also show what other users have bought. This way RS increases the number of items per order. When the customer has found options that meet his interest, he is preferably likely to purchase items.

- **Increases in sell by selling more neighborhood items** A RS should be able to recommend products that are less popular, the user may like to buy them. This leads to increases in sell and revenue.

- **Increase the user satisfaction** User's satisfaction is a major parameter in business. Traditionally they have to perform search after search to find a product of their interest. They will get exhausted by a querying again and again. By providing personalized experience to the user, a user feels that they valued as an individual. By providing a correct recommendation to the user, a user will become more engage and quickly find a product of their interest. So they will enjoy the system.

- **Valued each user** When old user visits website, treats him as loyal customer by providing personalized content. RS should be able to recommend products to the old customer based on his past purchases, browsing history and rating. As user spend more time on site system will be able to build a strong user profile and item profile, this will increases an efficiency of RS.

1.3 Limitations of Existing Recommendation System

- **Scalability** An E-commerce site has huge amounts of data, millions of customers and millions of distinct products [2]. User have to make lots of searches to find required product and there might be lots of user active at the same time and they frequently make a search. This will slow down the system.

- **Cold start** When a new user visits an online store, he/she has a limited amount of search history or product ratings. So his/her profile is almost empty. So it will difficult for a system to find out his/her taste. Because of this limited information system will recommend an irrelevant product. So this system suffers from cold start problem. This issue is also known as new user problem or new item problem.

- **Sparsity** On online store, a user has rated very few item or if a user is new to the system he has not rated any item yet. At this point, the system has no information about user So for the system, it is pretty difficult to determine his taste. Collaborative filtering used this information find neighborhoods of users using their profiles. If the system has less knowledge of the user, he/she could be related to the wrong neighborhood. Sparsity is the problem of lack of information.

- **Presence synonymy** Synonymy means the same kind of items may have different names. Most of RS treats these items separately so the performance of RS decreases.

- **Shilling attacks** There can be a case where people may give more rating for their own product and give bad ratings for their opponent. RS should be able to protect from this kind of anomaly.

1.4 Objective

Provide correct live recommendation on e-commerce site which makes all products on hand, which active user wants to purchase.

Chapter 2

Literature Survey

2.1 Related Work

Tapestry was very first recommendation system invented by David Goldberg, David Nichols, Brian M. Oki, and Douglas 7 rr in 1992 [3].This is mail filtering system which uses recommendation system as a filter. It is based on collaborative filtering. The user will read an electronic document and rated them (like or dislike). This data is used by a recommendation system to predict rating of the new document for other users which might be like by them. In this system, security is missing feature because integration of private mail with public information need a strong security.

Ringo is personalized music recommendation system developed by Upendra Shardanand and Pattie Maes in 1995 [4]. This recommendation system based on user-based collaborative filtering technique for the recommendation. This system had issues when a number of user increases and system need to narrow down domain for more efficiency.

Video recommendation system designed by Will Hill, Larry Stead, Mark Rosenstein and George Furnas in 1995 [5]. This recommendation system used user based collaborative filtering technique for the recommendation. RS finds most similar user to the active user and then recommend videoes which were liked by the most similar user.

In 2001, Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl designed recommendation system by using item-item collaborative filtering technique [6]. Their results prove that item-item collaborative filtering algorithm performs better than user-user collaborative filtering algorithm.

In 2003, Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl designed recommendation system which is used by amazon [2]. This recommendation system by using item-item collaborative filtering technique. This online recommendation system is able to react quickly when user preferences change and make relative recommendations. In future, this system can be extended recommendation for targeted marketing, online and offline.

Infrequent Purchased Product Recommendation System designed by Noraswaliza Abdullah, Yue Xu, Shlomo Geva, Jinghong Chen in 2010 [1]. This system used adaptive collaborative filtering technique for recommendation [1]. User's opinion or reviews are collected to build product profile and user's navigation data is used to build up the user profile, these both profiles are used for predicting recommendations. This system uses a round robin algorithm to select products for the recommendation from different product groups generated by the expanded queries. To improve an efficiency of the system, there is need of more advanced fusion techniques to select products from these groups.

Online Product Prediction and Recommendation using probability graphical model and collaborative filtering were proposed by S. S. Thakur, J. K. Sing in 2011 [7]. This paper presents an approach that used Bayesian belief networks (BBN) and the probabilities of inter-dependent events [7]. Nearest-neighbor collaborative filtering provides a meaningful recommendation for the active user. In future system needs to improve to increase accuracy, a system needs improvement to deal with a very large database for the real-time recommendation.

Product Recommendation Based on Search Keywords proposed by Jiawei Yao, Jiajun Yao, Rui Yang, Zhenyu Chen in 2012 [8]. They propose a new approach for active users based on the search keywords. To describe correlation between keywords and product's attributes are represented in a graph [8]. All these data will be utilized to predict recommendation of an active user. In future to improve the performance of RS, consider more attributes of the product to improve recommendations.

User Profile based Product Recommendation on Android Platform proposed by Nor Aniza Noor Amran, Norliza Zaini, Mustaffa Samad in 2014 [9]. They used user based collaborative filtering technique to predict preferences of an active user[9]. In this technique, system will consider user's attribute like gender, religion and medical condition [9].

Recommendation systems Principles, methods, and evaluation proposed by F.O. Isinkaye, Y.O. Folajimi, B.A. Ojokoh in 2015 [10]. This paper describes the different techniques and pros/cons of different prediction techniques in RS [10].

In 2008 A. Felfernig, R. Burke proposed Constraint-based recommendation Systems [11]. In some cases, RS failed to deliver recommendation [11]. This situation can be handled by constrained- based recommendation system where specific product properties as well as the correlation between customer interest and products are considered in the form of constraints and according to those constraints recommendations are given.

In 2016 Mehdi Elahi, Francesco Ricci, Neil Rubens performs a survey on active learning in collaborative filtering recommendation systems [12]. These works include a survey of two different techniques personalization and hybridization. Where personalization is a technique which recommends items to users according to their interest and hybridization is the technique which uses active learning to recommend items to users according to their interest by a single criterion or multiple criteria [12].

In 2016 Young-Duk Seo, Young-Gab Kim, Euijong Lee, Doo-Kwon Baik introduced Personalized recommendation system based on friendship strength in social network services. Personalized recommendation system based on collaborative filtering technique to recommend topics or interest according to user's taste. This paper introduces a proper method to calculate closeness between users on social network service by considering the various detail of user like the social circle, friendship strength and that closeness factor used to recommend user with their interest or topic [13].

In 2017 Maryam Khanian Najafabadi, Mohd Naz'ri Mahrin, Suriayati Chuprat, Haslina Md Sarkan proposed recommendation system to improve the accuracy of collaborative filtering using clustering and association rules mining on implicit data. They develop a system to eliminate the sparity problem of conventional CF technique. They use association rules mining to process massive data. Instead of counting total purchases made, it captures the multiple purchases per transaction in association rules. To reduce data size they use clustering as a dimensionality reduction technique. Then based on feature similarity between every pair of items are computed in order to make the prediction [14].

Chapter 3

Recommendation process

3.1 Phases of recommendation process

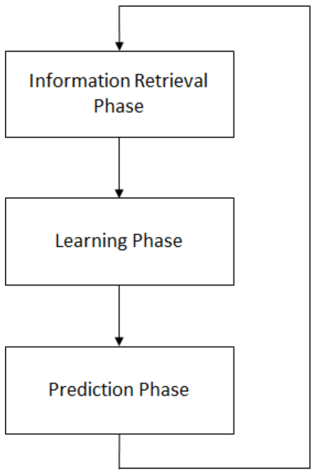


Figure 3.1: Phases of recommendation process

- **Information retrieval phase** In this phase, a user profile is build up by collecting user's information. This information can be collected from user's navigation, user's attribute, user's purchasing activity, or user's favorite item profile (Wish List). RS works efficiently if and only if user profile builds strongly. This information can be collected in two ways. Like explicit feedback, in which user have to provide feedback (rating) explicitly or implicit feedback by inferring user preferences indirectly through search history, previous purchases or from the wish list. Hybrid feedback can also be useful by combining both explicit and implicit feedback and also overcome issues of each other.

- **Learning phase** It applies a learning algorithm to predict recommendations by using user's profile or item profile collected in information retrieval phase.

- **Prediction/recommendation phase** System predict/recommend what kind of items the user may like based on information collected in the previous phase.

Chapter 4

Recommendation Techniques

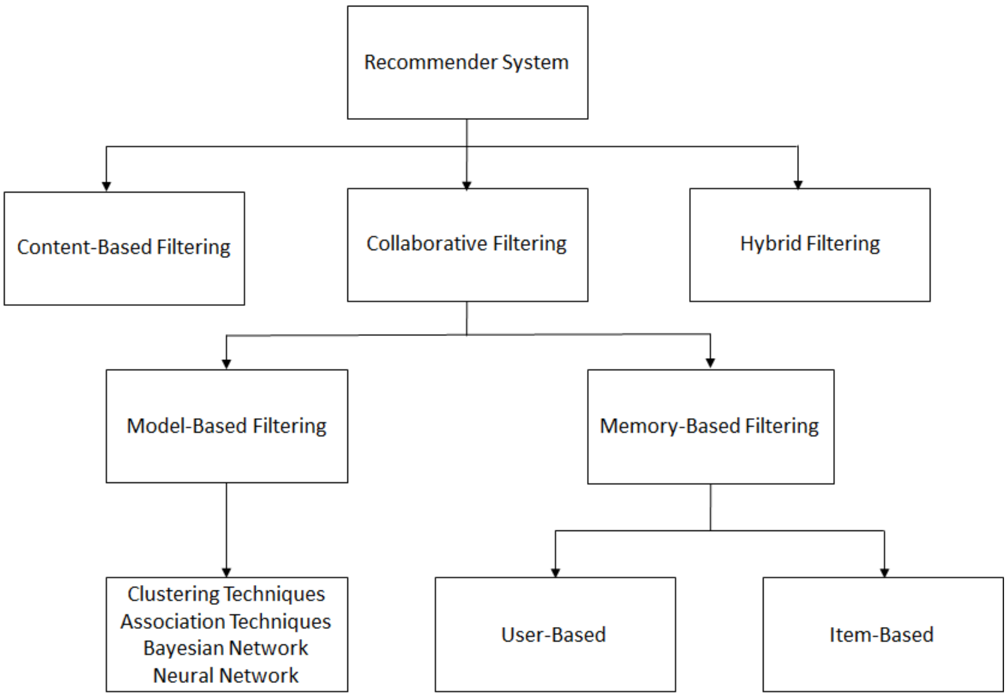


Figure 4.1: Recommendation Techniques

4.1 Content- based Filtering technique

In content-based filtering attributes of items play a significant role for the recommendation. It is a domain-dependent algorithm [10]. Content -based filtering technique analyze attributes of an item which were purchased or liked by the user in past and then select items for the recommendation which are similar with user liking [15]. Items that are mostly related to the positively rated items are recommended to the user [10]. Recommendations are based on the content of items rather on other users opinion.

The system creates a user profile which contains the description of items that user prefers. For example, the user previously liked three movies which are Titanic, Minority Report, Avatar based on this information system will infer that user mostly like sci-fi movie more than other so it will recommend Inception which is sci-fi kind of movie.

- Advantages of Content-based filtering technique are as follow
  - There is no need for data of other users because it is based on attributes of an item.
  - There is no cold start problem or sparity problem.
  - The system is able to recommend users with unique taste because it will not consider what other users like. It will only focus on target user only.
  - The system is able to recommend new and unpopular items.
- Limitations of Content-based filtering technique are as follow
  - The system is unable to use quality judgment from other users. For an example, other users rated one product very badly but there are chances that this product will be recommended to target user which is not so good.
  - Finding appropriate attribute of an item is difficult.
  - If user has rated or purchased more products than it is difficult to query on all item. In such a case algorithm needs to use subset or summarized data which will degrades recommendation quality.

4.2 Collaborative Filtering technique

Collaborative Filtering technique Collaborative technique finds a subgroup in which users have similar interest and preferences to the active user and recommendation will be given from this subgroup.

Collaborative Filtering systems collect user feedback. The user gives feedback in the form of ratings for items. Then system finds similarities in rating among several users to determine to recommend an item [15].

This technique builds a user-item matrix of preferences for items by users [10]. After that, it will finds similarities between users profiles. Such users create a subgroup called a neighborhood. A system will offer a recommendation of those items that were already positively rated by users in his neighborhood [10]. As this technique used widely due to its efficiency, simplicity, we will discuss in brief.

- Basic Assumption made in this technique are as follows
  - A user with similar taste has common preferences.
  - The system has enough data about preferences.

Categories of collaborative filtering

1. Memory-based technique or (neighbourhood-based)
2. Model-based technique

4.2.1 Memory-based technique or (neighbourhood-based)

Memory based collaborative filtering technique takes rating matrix as a input followed by similarity calculation, neighborhood selection, rating prediction of item for active user, ranking of most similar items and as a result of the entire process it provides top-N recommendations to active user. The entire procedure of memory based technique depicted in Fig 4.2.

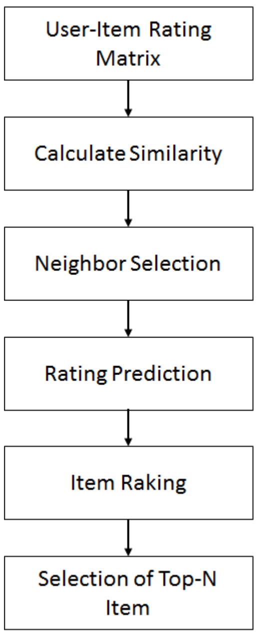


Figure 4.2: Procedure of Memory based collaborative filtering technique

Categories of Memory-based technique

1. User-based collaborative filtering

2. Item-based collaborative filtering

User-based collaborative filtering

Users \ Items	1	2	...	i	...	m
1	5	3		1	2	
2		2				4
:			5			
u	3	4		2	1	
:					4	
n			3	2		
a	3	5		?	1	

Figure 4.3: User-item matrix

This technique uses user-item rating matrix where each column represents item and each row represents a user. In user-item matrix each entry $r_{u,i}$ represents a rating of user u for item i. An objective of the system is to predict the missing rating $r_{a,i}$ for the active user a. To find similar user, system calculates a similarity between users. The system will compare ratings of other users with an active user on the same item. By averaging the ratings of the item by users similar to the active user, missing rating can be predicted for the item. The required algorithm and steps with description are given as below.

Algorithm 1 User based Algorithm

Data:User-Item Rating Matrix

Result:Top N most similar items

Initialization

for `each item` i `that active user` a `has no rating` **do**

  for `each user` u `that has a rating for item` i **do**

    Calculate similarity $W_{a,u}$ between active user a and u

  Select top K users with highest similarity as neighbourhood

  Calculate $P_{a,i}$ by weighted average of deviations from the neighbor's mean

Sort items in decreasing order of $P_{a,i}$

return Top N most similar items

1. All users are assigned weight according to similarity with the active user. Here to measure the similarity between the rating of two users, we used Pearson correlation coefficient. Pearson correlation coefficient can be defined as below.

$$\mathcal{W}_{a,u} = \frac{\sum_{i\in\mathcal{I}} \left(r_{a,i} - \bar{r}_a\right)\left(r_{u,i} - \bar{r}_u\right)}{\sqrt{\sum_{i\in\mathcal{I}} \left(r_{a,i} - \bar{r}_a\right)^2 \sum_{i\in\mathcal{I}} \left(r_{u,i} - \bar{r}_u\right)^2}} \tag{4.1}$$

   Where $W_{a,u}$ is a measure of similarity between the user u and the active user a, I is the set of items rated by user u and active user a, $r_{u,i}$ is the rating given to item i by user u, and $\bar{r_u}$ is the mean rating given by user u.

2. Top K users are selected who have the highest similarity with the active user. This top K selected user make subgroup called a neighborhood.

3. From a weighted combination of the selected neighbors prediction of rating will be computed [15].

$$\mathcal{P}_{a,i} = \bar{r}_a + \frac{\sum_{u\in\mathcal{K}} \left(r_{u,i} - \bar{r}_u\right) \times \mathcal{W}_{a,u}}{\sum_{u\in K} \mathcal{W}_{a,u}} \tag{4.2}$$

   where $P_{a,i}$ is a prediction for the active user a for item i, $W_{a,u}$ is similarity between active user a and other users u who belongs to the same neighborhood K.

Rather than using Pearson correlation coefficient to measure similarity we can use Cosine similarity coefficient which can be calculated using equation given below.

$$\begin{aligned}\mathcal{W}_{a,u} &= \cos\left(\vec{r}_a, \vec{r}_u\right) \\ &= \frac{\vec{r}_a . \vec{r}_u}{||\vec{r}_a||_2 \times ||\vec{r}_u||_2} \\ &= \frac{\sum_{i=1}^{m} r_{a,i} r_{u,i}}{\sqrt{\sum_{i=1}^{m} r_{a,i}^2}\sqrt{\sum_{i=1}^{m} r_{u,i}^2}}\end{aligned} \tag{4.3}$$

When computing cosine similarity, one cannot have negative ratings, and unrated Items are treated as having a rating of zero. In such a cases Pearson correlation generally performs well [15].

- Complexity analysis of User-based collaborative filtering are as follow

  Consider M is number of users and N is number of items. User-based CF technique examines M number of user and N numberof ietms for each user therefore complexity of user-based CF is O(MN)[2].

- Advantages of User-based collaborative filtering are as follow
  - No knowledge about item features needed.

- Limitations of User-based collaborative filtering are as follow
  - New user suffers from cold start problem because a new user has not sufficient data to build user profile thus RS can't recommend items efficiently.
  - New item suffers from cold start problem because items with few rating cannot be easily recommended.
  - Similarity matrix can not be computed offline. Because every time system needs to calculate active user's neighborhood. This online computation are very expensive in terms of time.

Item-based collaborative filtering

This technique uses user-item rating matrix. Instead of calculating a similarity between users system will calculate a similarity between items. It builds a model of items that are highly correlated. Then it selects the most similar items. In this approach similarities are calculated offline using Pearson correlation [15].

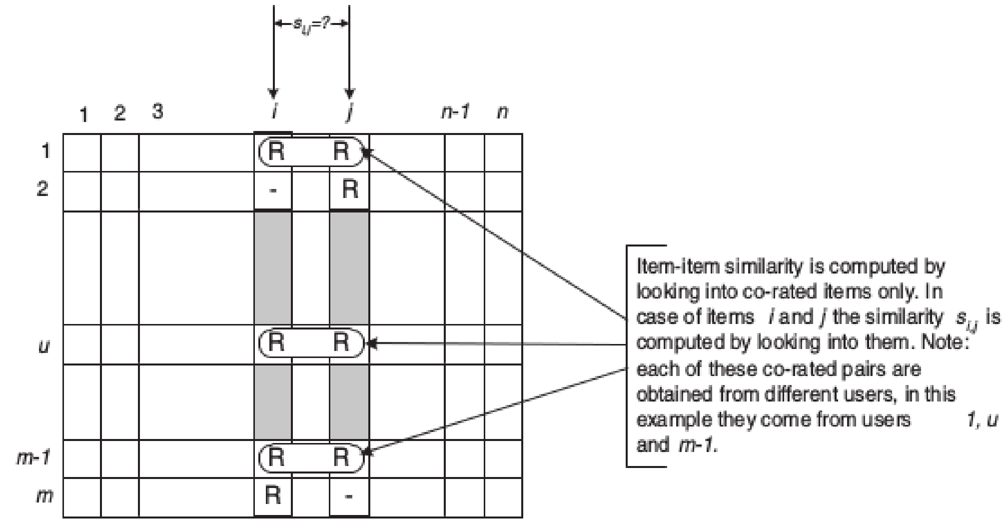


Figure 4.4: Item-based collaborative filtering

Algorithm 2 Item-based Algorithm

Data:User-Item Rating Matrix

Result:Top N most similar items

Initialization

for `each item` i `not rated by active user` a **do**

  for `each item` j `rated by user` a **do**

    Calculate similarity $W_{i,j}$ between item i and j

Select top K item with highest similarity as neighbourhood

Calculate $P_{a,i}$ by weighted average of deviations from the neighbor's mean

Sort items in decreasing order of $P_{a,i}$

return Top N most similar items

$$\mathcal{W}_{i,j} = \frac{\sum_{u\in\mathcal{U}}\left(r_{u,i}-\bar{r}_i\right)\left(r_{u,j}-\bar{r}_j\right)}{\sqrt{\sum_{u\in\mathcal{U}}\left(r_{u,i}-\bar{r}_i\right)^2\sum_{u\in\mathcal{U}}\left(r_{u,j}-\bar{r}_j\right)^2}} \tag{4.4}$$

where U is a set of all users who have rated both items i and j, $r_{u,i}$ is rating of user u on item i, $\bar{r}_i$ is the average rating of the ith item across users.

An objective of the system is to predict a rating of item i for active user a. This can be calculated by the following equation.

$$\mathcal{P}_{a,i} = \frac{\sum_{j \in \mathcal{K}} r_{a,j} \mathcal{W}_{i,j}}{\sum_{j \in K} |\mathcal{W}_{i,j}|} (4.5)$$

where K is subgroup of k most similar items rated by an active user a.

- Complexity analysis of Item-based collaborative filtering are as follow

  Consider M is number of users and N is number of items.After analyzing above algorithm,in worst case it will take O(MN) of processing time [2].

- Advantages of Item-based collaborative filtering are as follow

  – No knowledge about item features needed.

  – Expensive similarity matrix can be calculated offline. Online function will be only finding most similar item which will match active user's preference

- Limitations of Item-based collaborative filtering are as follow

  – New user suffers from cold start problem because a new user has not sufficient data to build user profile thus RS can't recommend items efficiently.

  – New item suffers from cold start problem because items with few rating cannot be easily recommended.

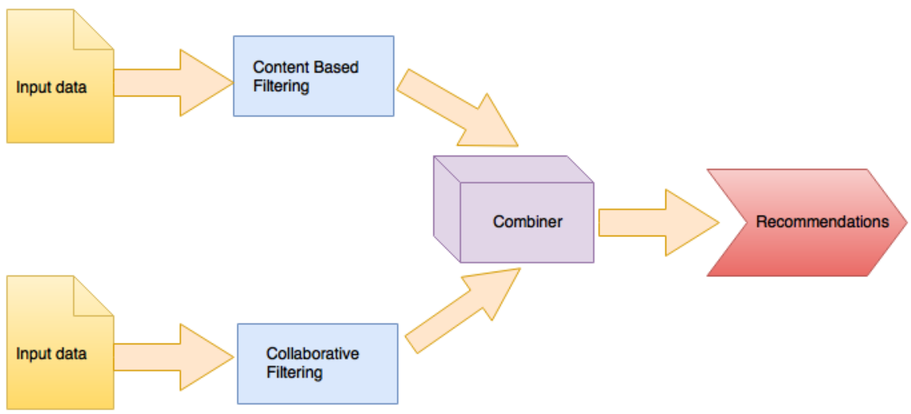
4.2.2 Model-based technique

In Model-based technique, feedback is collected from the user in form of ratings and use it to build a model. Using any machine learning algorithm model can be built. This model is build using extracting some information from the huge dataset, it can be of user information or item information. Then this model is trained using any machine learning algorithm and some common pattern can be captured from this huge dataset. Then from that pattern system can quickly recommend a set of items to an active user. Examples of these techniques include Regression, Clustering, Decision Tree, Artificial Neural network and Bayesian Classifiers.

- Advantages of model-based collaborative filtering are as follow
  - Clustering can be perform offline because model-based will just compare with controlled number of cluster not with all user.
- Limitations of model-based collaborative filtering are as follow
  - In clustering number of user group to gather to make a cluster or segment, then match active user to a cluster, and then it will consider all user in cluster as most similar user for providing recommendations.Because the similar user that the cluster find are not the most similar user, which will degrades recommendation quality.
  - To improve recommendation quality by finding fine grained cluster, but then online classification of user becomes more expensive.

4.3 Hybrid Filtering technique

Hybrid filtering technique improves an efficiency of RS because it uses both content-based filtering technique and collaborative filtering technique.



- Different approaches for implementation
  - Individual implementation of content-based filtering and collaborative filtering and then combining the result
  - Implement some content-based filtering in collaborative approach
  - Implement some collaborative filtering in content-based approach
  - Creating a unified RS that brings together both approaches

- Advantages of hybrid filtering are as follow
  - By combining above two approaches limitations of pure RS can be overcome.
- Limitations of hybrid filtering are as follow
  - It is very complex and combination of two algorithm and depending on weight assign to algorithm result will be combine but it is difficult to decide how much weight should be assign to both algorithm. A weight can be very from user to user so weight need to calculated online which will take some more time.
  - Determining correct features of item for recommendation is difficult task because of this limitation of content based filtering, quality of recommendation will degrades.

Chapter 5

Tools and Technology

5.1 Introduction of Apache Mahout

Apache Mahout is a developed by Apache Software Foundation. It used to implements various algorithms of collaborative filtering likes User-Based Collaborative Filtering, Item-Based Collaborative Filtering. It also used to implement Classification algorithms, Clustering algorithms and Dimentionality reduction algorithms. We have used java dependency of apache-core 0.8.

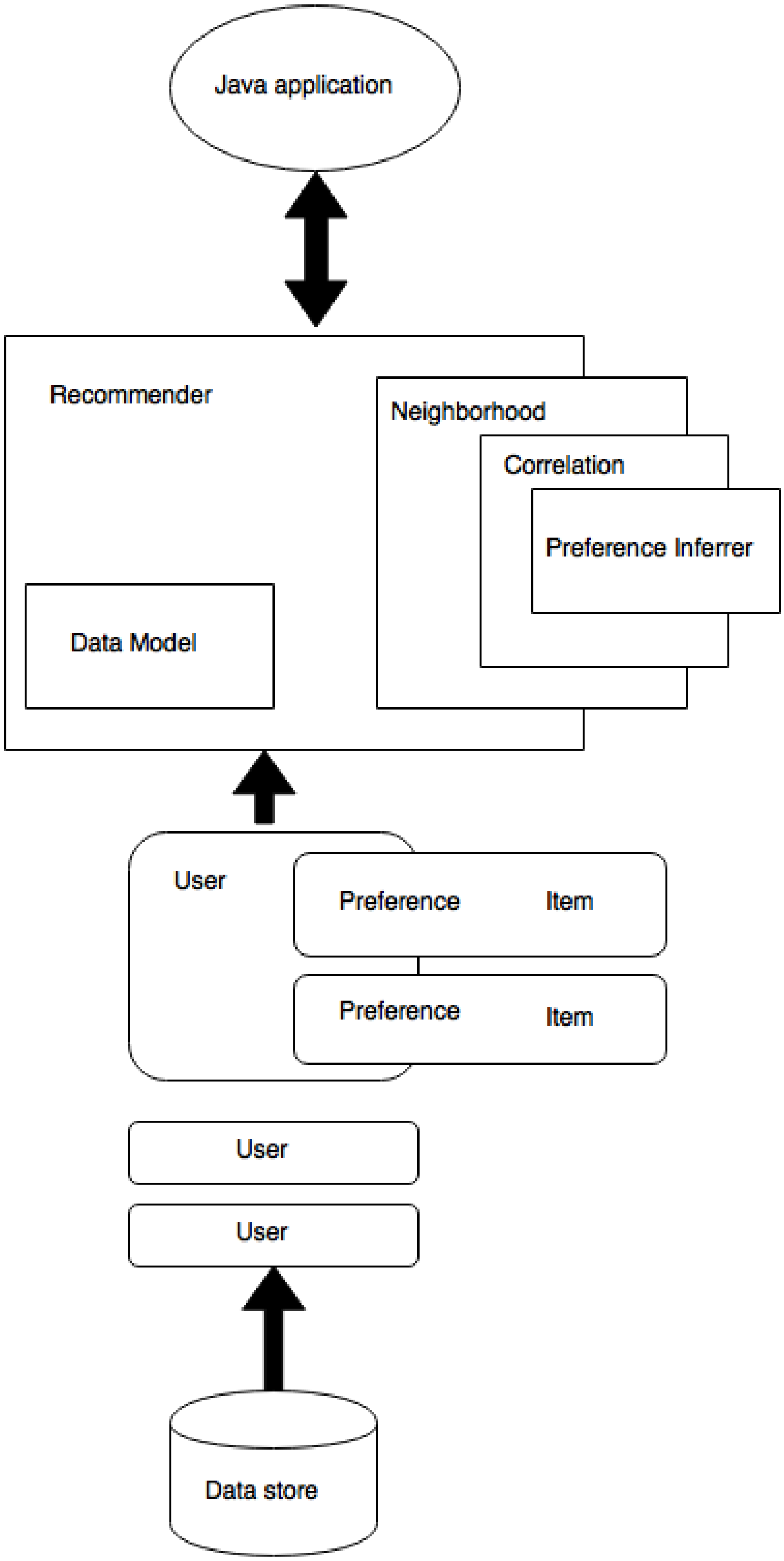


Figure 5.1: Architecture of Apache Mahout Recommender Engine

A Mahout-based collaborative filtering engine takes users' preferences for items ("tastes") and returns estimated preferences for other items.

Top-level packages define the Mahout interfaces to these key abstractions:

- Data Model
- User Similarity
- Item Similarity
- User Neighbourhood
- Recommender

Chapter 6

Experimental evaluation and results

6.1 Dataset

We have use latest dataset of Shopify. Which contain 1,00,000 ratings for 27,000 movies by 700 users.

6.2 Accuracy metrics

There are multiple ways to measure accuracy of prediction algorithm. MAE is used to measure how close predicted values to the actual value. MAE is defined as average difference between predicted rating p_i and actual rating a_i. The MAE is give by

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|p_i - a_i| \tag{6.1}$$

For better and accurate recommendations MAE should be as lower as possible. To calculate MAE, we have consider 80% data as a training data and remaining 20% data as testing data. MAE has been computed for item-based filtering technique and user-based filtering technique for different numbers of user in system.

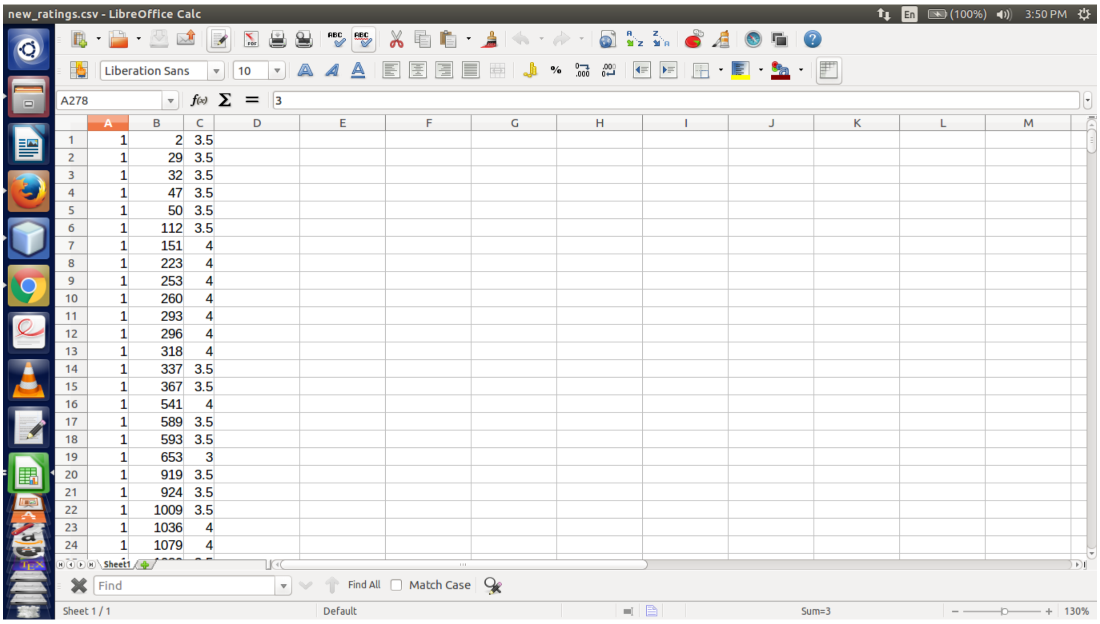


Figure 6.1: Input file (userID,ItemID,Ratings) to Recommender system

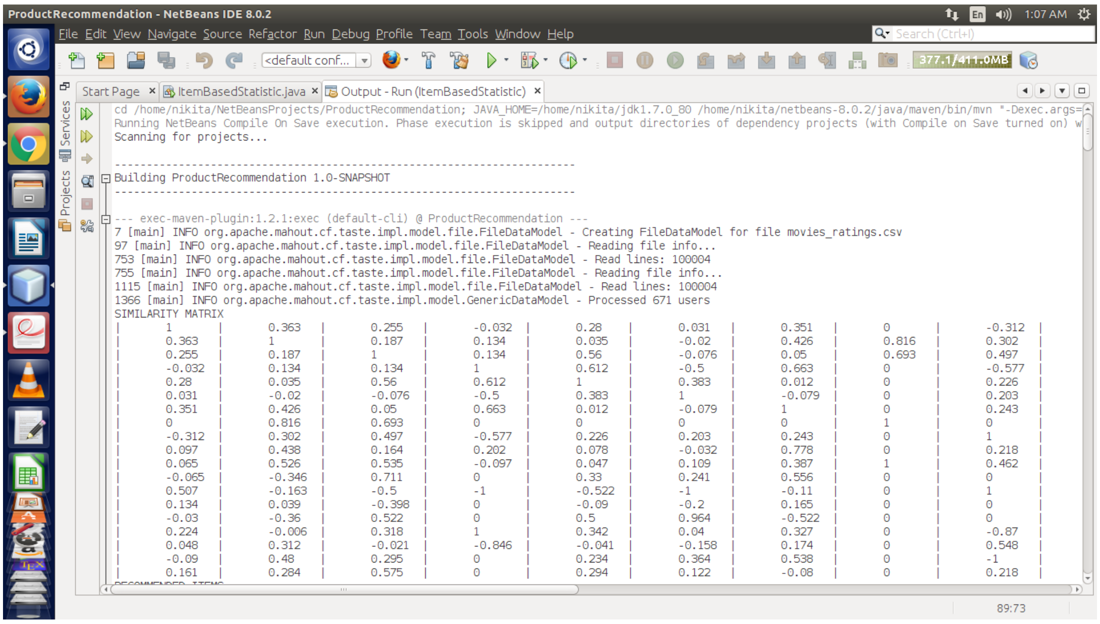


Figure 6.2: Similarity Matrix

Figure 6.2 This similarity matrix of user ID 114 calculated using pearson correlation similarity.

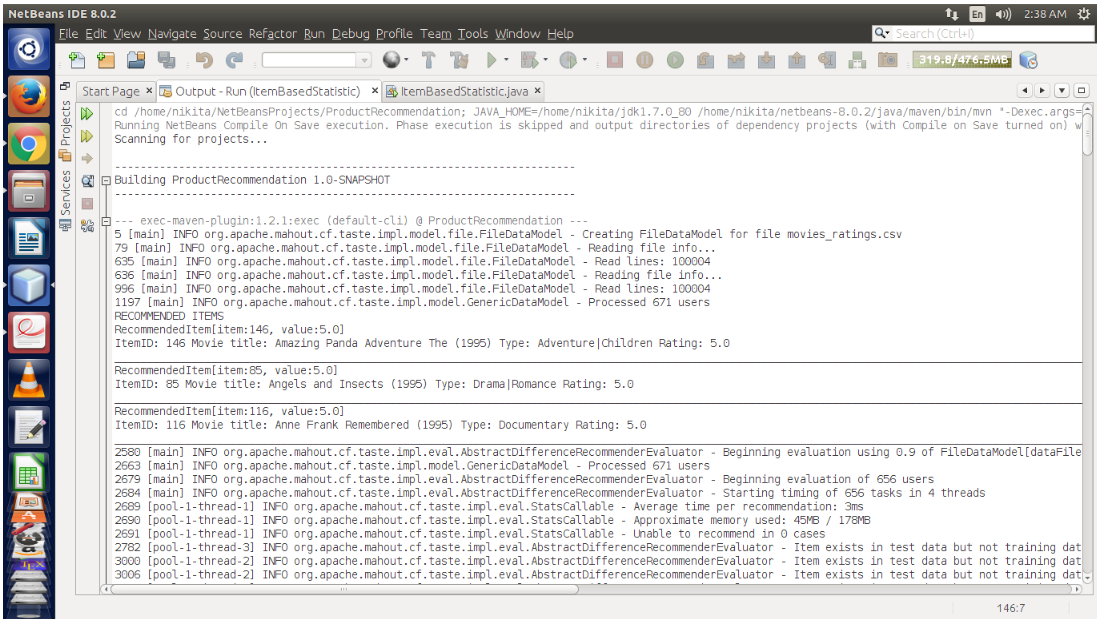


Figure 6.3: Recommended items to active user

Figure 6.4 shows top 3 recommended items which has highest rating (5 star).

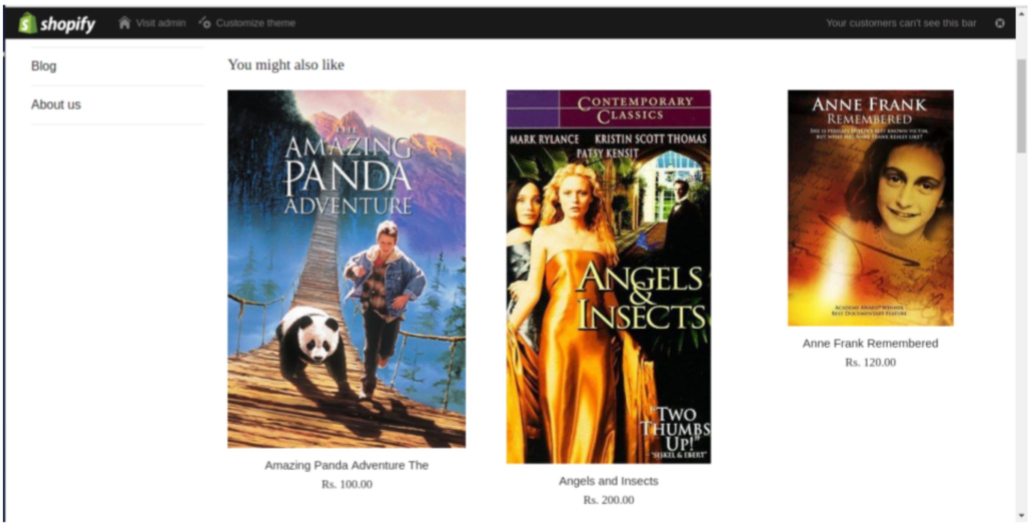


Figure 6.4: Live recommendation on e-commerce store

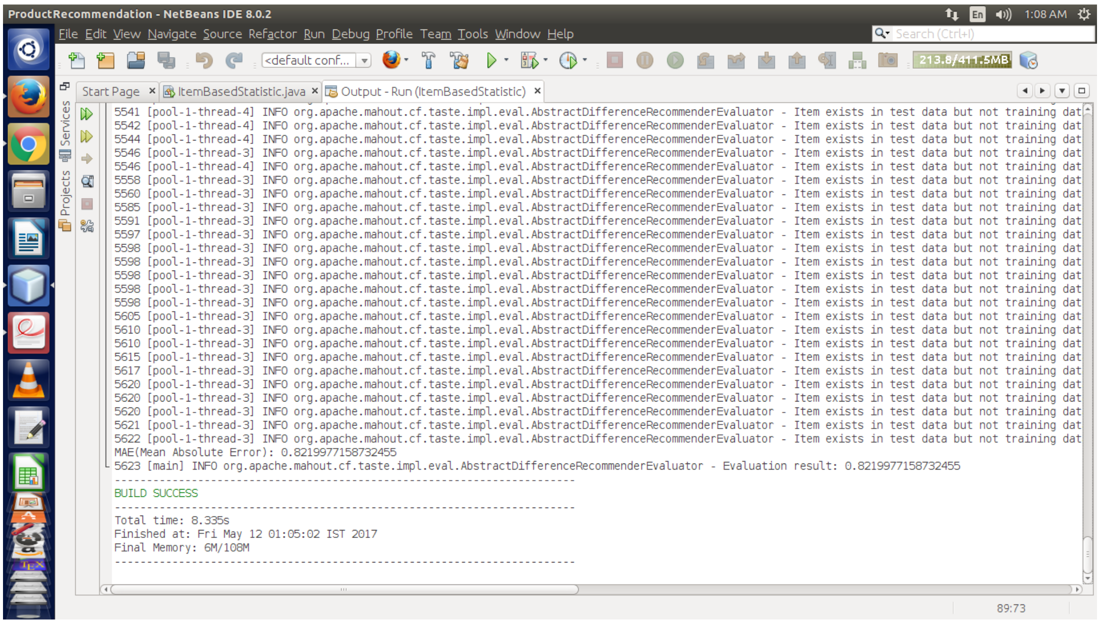


Figure 6.5: Mean absolute error

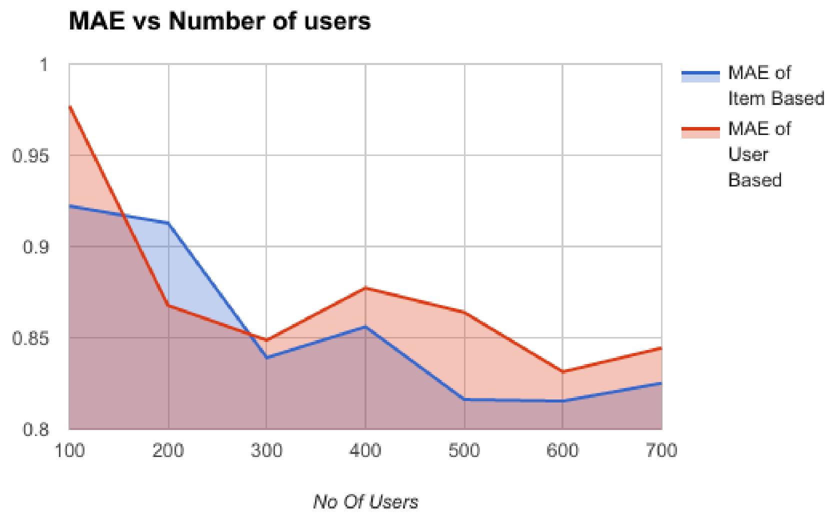


Figure 6.6: Comparison between user based and item based in terms of MAE and Number of user

Figure 6.6 is shows difference between user based and item based in terms of scalability. This results shows that after some point when number of user increases value of MAE also increases for user based where as in item based filtering technique when number of user increases MAE decreases. An item based filtering technique perform better than user based filtering technique in terms of scalability.

Chapter 7

Conclusion

There are so many RS have been introduced, that are based on content-based filtering, collaborative filtering and hybrid recommendation technique. Collaborative filtering technique is most popular recommendation techniques as it focused on user's interest instead of considering other factors. Quantitative analysis of both user-based filtering and item-based filtering shows that when number of user increases, performance of user-based filtering degrades and item-based filtering performs better than user based.

Chapter 8

Future work

To solve the scalability problem of user based filtering technique and to improve recommendation quality, system should use hydrid model. In hybrid model we can combine both user based filtering and content based filtering. So by applying content based after user based we can make constrain on attributes of item. Which provides better performance and reduce error rate when number of user increases.

Bibliography

[1] N. Abdullah, Y. Xu, S. Geva, and J. Chen, "Infrequent purchased product recommendation making based on user behaviour and opinions in e-commerce sites," pp. 1084–1091, 2010.

[2] G. Linden, B. Smith, and J. York, "Amazon. com recommendations: Item-to-item collaborative filtering," *IEEE Internet computing*, vol. 7, no. 1, pp. 76–80, 2003.

[3] D. Goldberg, D. Nichols, B. M. Oki, and D. Terry, "Using collaborative filtering to weave an information tapestry," *Communications of the ACM*, vol. 35, no. 12, pp. 61–70, 1992.

[4] U. Shardanand and P. Maes, "Social information filtering: algorithms for automating "word of mouth"," pp. 210–217, 1995.

[5] W. Hill, L. Stead, M. Rosenstein, and G. Furnas, "Recommending and evaluating choices in a virtual community of use," pp. 194–201, 1995.

[6] B. Sarwar, G. Karypis, J. Konstan, and J. Riedl, "Item-based collaborative filtering recommendation algorithms," pp. 285–295, 2001.

[7] S. Thakur and J. Sing, "Online product prediction and recommendation using probability graphical model and collaborative filtering: A new approach," pp. 151–156, 2011.

[8] J. Yao, J. Yao, R. Yang, and Z. Chen, "Product recommendation based on search keywords," pp. 67–70, 2012.

[9] N. A. N. Amran, N. Zaini, and M. Samad, "User profile based product recommendation on android platform," pp. 1–6, 2014.

[10] F. Isinkaye, Y. Folajimi, and B. Ojokoh, "Recommendation systems: Principles, methods and evaluation," *Egyptian Informatics Journal*, vol. 16, no. 3, pp. 261–273, 2015.

[11] A. Felfernig and R. Burke, "Constraint-based recommender systems: technologies and research issues," p. 3, 2008.

[12] M. Elahi, F. Ricci, and N. Rubens, "A survey of active learning in collaborative filtering recommender systems," *Computer Science Review*, vol. 20, pp. 29–50, 2016.

[13] Y.-D. Seo, Y.-G. Kim, E. Lee, and D.-K. Baik, "Personalized recommender system based on friendship strength in social network services," *Expert Systems with Applications*, vol. 69, pp. 135–148, 2017.

[14] M. K. Najafabadi, M. N. Mahrin, S. Chuprat, and H. M. Sarkan, "Improving the accuracy of collaborative filtering recommendations using clustering and association rules mining on implicit data," *Computers in Human Behavior*, vol. 67, pp. 113–128, 2017.

[15] P. Melville and V. Sindhwani, "Recommender systems," pp. 829–838, 2011.