MUSIC RECOMMENDER USING COLLABORATIVE AND CONTENT BASED FILTERING

Project Report

By

Abhimanyu Olla

ABSTRACT

Recommender systems are an important part of the information and e-commerce ecosystem. They represent a powerful method for enabling users to filter through large information and product spaces. Nearly two decades of research on collaborative filtering have led to a varied set of algorithms and a rich collection of tools for evaluating their performance. Research in the field is moving in the direction of a richer understanding of how recommender technology may be embedded in specific domains.

The differing personalities exhibited by different recommender algorithms show that recommendation is not a one-size fits-all problem. Here is the situation where personalised recommendation using collaborative filtering comes into play. We apply this collaborative filtering technique for suggesting personalised songs to user on the basis of their taste of listening we suggest user with songs on their listening of genres, artist and various other segmentations. At last we also present a statistical comparison between our model and the preexisting popularity-based model to show how our model suggest more precise user-friendly recommendation than suggesting common songs on the basis of popularity.

INTRODUCTION

This document contains all requirement specification architecture for Music Playlist Auto Recommendation Project. After giving information about the definition of the project at the beginning part of the document, we will give complete description for overview and list the requirements which meet the needs of the Project roughly.

1.1 Problem Definition

The aim is to create a Music Playlist Creator Web application which not only helps to download and create the playlist, but also keeps a track on listener's choice of listening and suggests and recommends the listener with the genre and the music he may like. Artist will also have an option to create account and manage their uploads.

Objective

1. Provide listener to download and create their own music playlist and also manage it.
2. Allow the artist to upload their music and manage their album release by giving reminder.
3. Recommend Listeners with their choice of music by tracking their frequent listening.
4. Choice for artists to plan their calendar and set reminder for timely uploads
5. Suggest artists with the choice and genre of music to create as per the trend.
6. Listeners will be given a choice to rate the songs and a final listener's average ratings will be displayed

LITERATURE SURVEY

After that e-commerce gained popularity of selling products and the development of new generation mobile phones and innovative inventions like tablets, shopping become easier than before. Therefore, new kind of advertising techniques came up. Recommendation systems are one of the most popular techniques nowadays. They are useful for both the company and the user, because they increasing product sales and reducing the time spend on shopping. However, they have some problems because of huge data. Main problems are being unable to get really relevant results and being unable to get results in reasonable time.

Up to now there are a lot of methods developed to resolve the problems stated above such as collaborative filtering, content-based filtering. However, they have some weaknesses. For example, collaborative filtering has the problem called cold start and means that recommendation system cannot produce any suggestion or recommendations. This problem occurs when items are provided in the system but there are few customers and few or no rankings. And the other example for content-based filtering, if the content is in lack of enough information to distinguish the items precisely, the recommendation cannot be precisely done. On the other hand, our system should find the most accurate recommendations.

Our potential users that we desire to help their problems are companies that use internet utilities to sell their products. And by proposing our product to these companies, we will reach to users of the websites that companies use. Therefore, we will reach to both customers and companies with our system.

Type	Percentage	Features
Savants	7	Everything in life seems to be tied up with music. Their musical knowledge is very extensive.
Enthusiasts	21	Music is a key part of life but is also balanced by other interests.
Casuals	32	Music plays a welcome role, but other things are far more important.
Indiﬀrents	40	They would not lose much sleep if music ceased to exist, they are a predominant type of listeners of the whole population.

Table 1. User Listening Experience Data Categorization

1. Process and phases of recommendation system:

a) Information collection phase

This collects relevant information of users to generate a user profile or model for the prediction tasks including user's attribute, behaviours or content of the resources the user accesses. A recommendation agent cannot function accurately until the user profile/model has been well constructed. The system needs to know as much as possible from the user in order to provide reasonable recommendation right from the onset. Recommender systems rely on different types of input such as the most convenient high-quality explicit feedback, which includes explicit input by users regarding their interest in item or implicit feedback by inferring user preferences indirectly through observing user behaviour. Hybrid feedback can also be obtained through the combination of both explicit and implicit feedback. In E-learning platform, a user profile is a collection of personal information associated with a specific user. This information includes cognitive skills, intellectual abilities, learning styles, interest, preferences and interaction with the system. The user profile is normally used to retrieve the needed information to build up a model of the user.

b) Explicit feedback

The system normally prompts the user through the system interface to provide ratings for items in order to construct and improve his model. The accuracy of recommendation depends on the quantity of ratings provided by the user. The only shortcoming of this method is, it requires effort from the users and also, users are not always ready to supply enough information. Despite the fact that explicit feedback requires more effort from user, it is still seen as providing more reliable data, since it does not involve extracting preferences from actions, and it also provides transparency into the recommendation process that results in a slightly higher perceived recommendation quality and more confidence in the recommendations.

c) Implicit feedback

The system automatically infers the user's preferences by monitoring the different actions of users such as the history of purchases, navigation history, and time spent on some web pages, links followed by the user, content of e-mail and button clicks among others. Implicit feedback reduces the burden on users by inferring their user's preferences from their behaviour with the system. The method though does not require effort from the user, but it is less accurate. Also, it has also been argued that implicit preference data might in actuality be more objective, as there is no bias arising from users responding in a socially desirable way and there are no self-image issues or any need for maintaining an image for others.

d) Hybrid feedback

The strengths of both implicit and explicit feedback can be combined in a hybrid system in order to minimize their weaknesses and get a best performing system. This can be achieved by using an implicit data as a check on explicit rating or allowing user to give explicit feedback only when he chooses to express explicit interest.

e) Learning phase

It applies a learning algorithm to filter and exploit the user's features from the feedback gathered in information collection phase.

f) Prediction/recommendation phase:

It recommends or predicts what kind of items the user may prefer. This can be made either directly based on the dataset collected in information collection phase which could be memory based or model based or through the system's observed activities of the user. Fig. 1 highlights the recommendation

2. Feedback System

Recommendation filtering techniques

The utilization of effective and exact recommendation procedures is essential for a framework that will give great and helpful recommendation to its individual clients. This clarifies the significance of understanding the highlights and possibilities of various recommendation strategies.

Fig. 1 Recommendation System

ARCHITECTURAL MODEL

The model uses several techniques to make it more precise. In this section I show you how the model works with some structural and data flow diagrams. In popularity based, since between users, common songs exist(repeated), so similarity between user is calculated and suggested.

But in my model which uses item personalisation so basically using item similarity based cooccurrence matrix (similarity on the basis of users who have listeners to both songs i and j), here I will have song by song matrix where each cell will have normalised value of the listen count dedicated to each user.

IMPLEMENTATION

I implemented the project on the Jupyter notebook platform. I basically compared the existing popularity-based model with my item personalization model and got the following result

Use the popularity model to make some predictions

```
In [13]: user_id = users[5]
         pm.recommend(user_id)
```

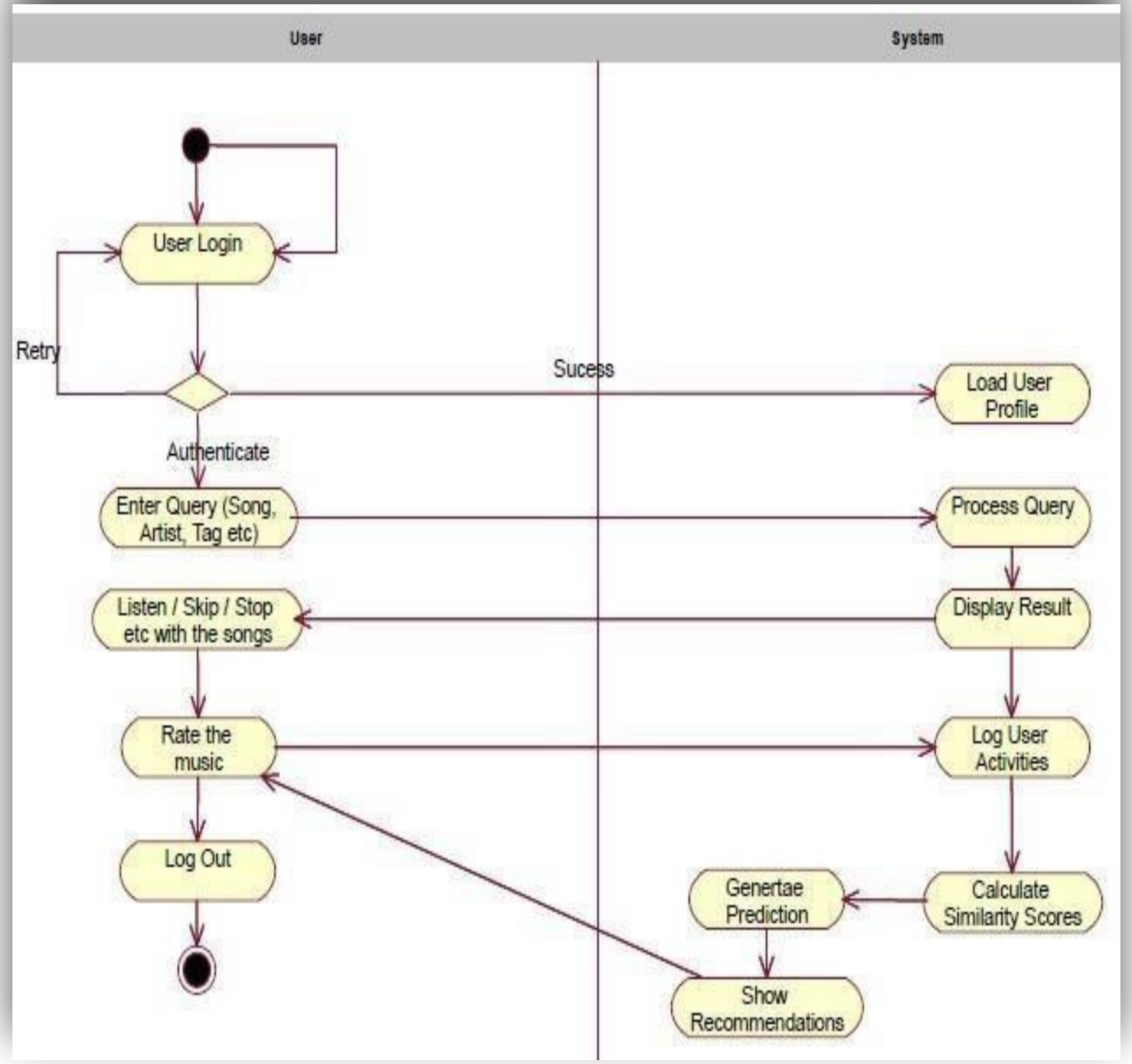
Out[13]:

	user_id	song	score	Rank
3194	4bd88bfb25263a75bbdd467e74018f4ae570e5df	Sehr kosmisch - Harmonia	37	1.0
4083	4bd88bfb25263a75bbdd467e74018f4ae570e5df	Undo - Björk	27	2.0
931	4bd88bfb25263a75bbdd467e74018f4ae570e5df	Dog Days Are Over (Radio Edit) - Florence + Th...	24	3.0
4443	4bd88bfb25263a75bbdd467e74018f4ae570e5df	You're The One - Dwight Yoakam	24	4.0
3034	4bd88bfb25263a75bbdd467e74018f4ae570e5df	Revelry - Kings Of Leon	21	5.0
3189	4bd88bfb25263a75bbdd467e74018f4ae570e5df	Secrets - OneRepublic	21	6.0
4112	4bd88bfb25263a75bbdd467e74018f4ae570e5df	Use Somebody - Kings Of Leon	21	7.0
1207	4bd88bfb25263a75bbdd467e74018f4ae570e5df	Fireflies - Charttraxx Karaoke	20	8.0
1577	4bd88bfb25263a75bbdd467e74018f4ae570e5df	Hey_ Soul Sister - Train	19	9.0
1626	4bd88bfb25263a75bbdd467e74018f4ae570e5df	Horn Concerto No. 4 in E flat K495: II. Romanc...	19	10.0

```
In [14]: #recommendation for other user
         user_id = users[8]
         pm.recommend(user_id)
         #as we see the recommendation are same for both the users, so there are no personalisat
```

Out[14]:

	user_id	song	score	Rank
3194	9bb911319fbc04f01755814cb5edb21df3d1a336	Sehr kosmisch - Harmonia	37	1.0
4083	9bb911319fbc04f01755814cb5edb21df3d1a336	Undo - Björk	27	2.0
931	9bb911319fbc04f01755814cb5edb21df3d1a336	Dog Days Are Over (Radio Edit) - Florence + Th...	24	3.0
4443	9bb911319fbc04f01755814cb5edb21df3d1a336	You're The One - Dwight Yoakam	24	4.0
3034	9bb911319fbc04f01755814cb5edb21df3d1a336	Revelry - Kings Of Leon	21	5.0
3189	9bb911319fbc04f01755814cb5edb21df3d1a336	Secrets - OneRepublic	21	6.0
4112	9bb911319fbc04f01755814cb5edb21df3d1a336	Use Somebody - Kings Of Leon	21	7.0
1207	9bb911319fbc04f01755814cb5edb21df3d1a336	Fireflies - Charttraxx Karaoke	20	8.0
1577	9bb911319fbc04f01755814cb5edb21df3d1a336	Hey_ Soul Sister - Train	19	9.0
1626	9bb911319fbc04f01755814cb5edb21df3d1a336	Horn Concerto No. 4 in E flat K495: II. Romanc...	19	10.0



Here in popularity-based prediction I get to know that the model produces similar results for all user irrespective of their choice of listening.

```
--------------------------------------------------------------------------------
Training data songs for the user userid: 4bd88bfb25263a75bbdd467e74018f4ae570e5df:
--------------------------------------------------------------------------------
Just Lose It - Eminem
Without Me - Eminem
16 Candles - The Crests
Speechless - Lady GaGa
Push It - Salt-N-Pepa
Ghosts 'n' Stuff (Original Instrumental Mix) - Deadmau5
Say My Name - Destiny's Child
My Dad's Gone Crazy - Eminem / Hailie Jade
The Real Slim Shady - Eminem
Somebody To Love - Justin Bieber
Forgive Me - Leona Lewis
Missing You - John Waite
Ya Nada Queda - Kudai
----------------------------------------------------------------------
Recommendation process going on:
----------------------------------------------------------------------
No. of unique songs for the user: 13
no. of unique songs in the training set: 4483
Non zero values in cooccurence_matrix :2097
```

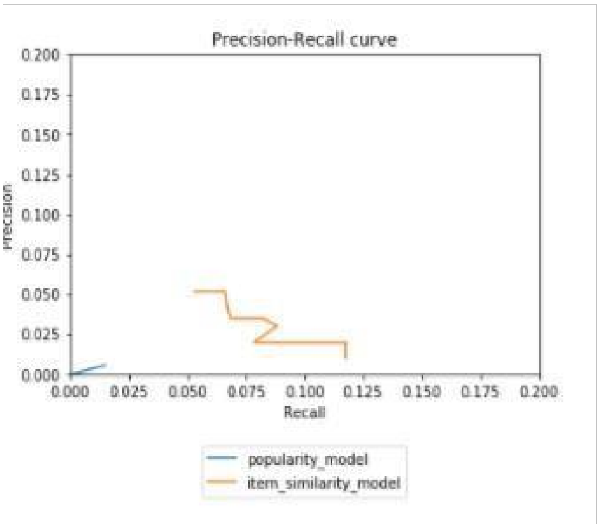
]:

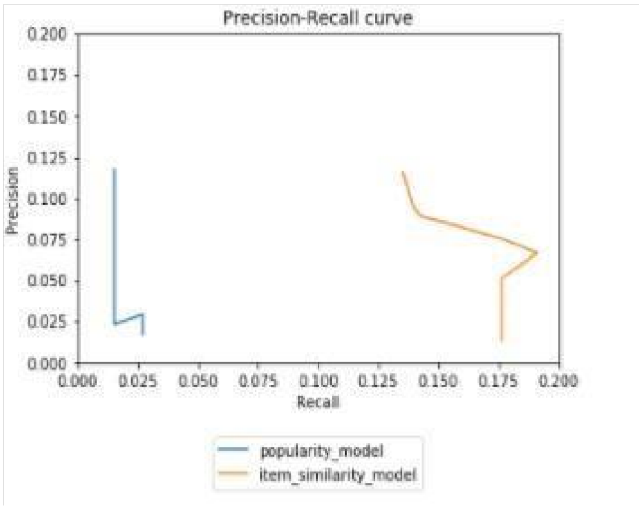
	user_id	song	score	rank
0	4bd88bfb25263a75bbdd467e74018f4ae570e5df	Superman - Eminem / Dina Rae	0.088692	1
1	4bd88bfb25263a75bbdd467e74018f4ae570e5df	Mockingbird - Eminem	0.067663	2
2	4bd88bfb25263a75bbdd467e74018f4ae570e5df	I'm Back - Eminem	0.065385	3
3	4bd88bfb25263a75bbdd467e74018f4ae570e5df	U Smile - Justin Bieber	0.064525	4
4	4bd88bfb25263a75bbdd467e74018f4ae570e5df	Here Without You - 3 Doors Down	0.062293	5
5	4bd88bfb25263a75bbdd467e74018f4ae570e5df	Hellbound - J-Black & Masta Ace	0.055769	6
6	4bd88bfb25263a75bbdd467e74018f4ae570e5df	The Seed (2.0) - The Roots / Cody Chestnutt	0.052564	7
7	4bd88bfb25263a75bbdd467e74018f4ae570e5df	I'm The One Who Understands (Edit Version) - War	0.052564	8
8	4bd88bfb25263a75bbdd467e74018f4ae570e5df	Falling - Iration	0.052564	9
9	4bd88bfb25263a75bbdd467e74018f4ae570e5df	Armed And Ready (2009 Digital Remaster) - The ...	0.052564	10

In this model user gets recommendation on the basis of their personal choices and it is done as mentioned in our architecture.

Comparing the Models

When compared the models using precision recall curves, the precision-recall plot is a modelwide evaluation measure that is based on two basic evaluation measures – recall and precision. Recall is a performance measure of the whole positive part of a dataset, whereas precision is a performance measure of positive predictions.





For Small data sets **For Large data sets**

SUMMARISING

From the above plot comparison, we see that item personalisation recommendation we get to know that the personalised recommendation becomes more precise in case of large data sets. So, the model is far more precise and user friendly in terms of precise recommendation as compared the popularity-based mode.

REFERENCES

1. Collaborative Filtering–Based Recommender System: Approaches and Research Challenges Ritu Sharma*, Dinesh Gopalani**, Yogesh Meena*** Department of Computer Science & Engineering, Malaviya National Institute of Technology, Jaipur, India Email: {*2015rcp9027, ** dgopalani.cse, *** ymeena.cse} @mnit.ac.in
2. Collaborative Filtering for Recommender Systems Rui-sheng Zhang, Qi-dong Liu, Chun-Gui, JiaXuan Wei, Huiyi-Ma School of Information Science & Engineering Lanzhou University Lanzhou, China wliuqidong@126.com
3. A Research of Job Recommendation System Based on Collaborative Filtering Yingya Zhang, Cheng Yang, Zhixiang Niu Information Engineering School Communication University of China Beijing, China zyy90@aliyun.com, cafeeyang@163.com, niuzhixiang4321@126.com
4. A Recommender System Architecture for an Inter-application Environment Eduardo Jose Marcelino Vicente dos Santos, Byron Leite Dantas Bezerra Jefferson Silva de Amorim and Arthur Inacio do Nascimento Polytechnic School of Pernambuco University of Pernambuco 50.750-470, Recife - PE, Brazil Email: ejmvs@ecomp.poli.br,byronleite@ecomp.poli.brjeffersonsilvadeamorim@gmail.com, arthur.inas@gmail.com
5. Recommendation systems: Principles, methods and evaluation Author links open overlay panelF.O.IsinkayeaY.O.FolajimibB.A.Ojokohc
6. A Comparative Study of Collaborative Filtering Algorithms Joonseok Lee, Mingxuan Sun, Guy Lebanon May 14, 2012
7. Amazon.com recommendations: item-to-item collaborative filtering G. Linden ; B. Smith ; J. York
8. User-based collaborative-filtering recommendation algorithms on Hadoop ZD Zhao, MS Shang - … Discovery and Data Mining, 2010. WKDD …, 2010 - ieeexplore.ieee.org
9. Explaining collaborative filtering recommendations JL Herlocker, JA Konstan, J Riedl - … of the 2000 ACM conference on …, 2000 - dl.acm.org

10. A survey of collaborative filtering techniques X Su, TM Khoshgoftaar - Advances in artificial intelligence, 2009 - dl.acm.org