Music Recommendation System

Submitted in the partial fulfillment for the award of

the degree of

BACHELOR OF ENGINEERING

IN

Computer Science and Engineering

Submitted by:

Puranjay Kwatra (19bcs6061)

Raghvendar Changotra(19bcs6083)

Anshul(19bcs6062)

Vishal (19bcs6073)

Rajeswar (19bcs6051)

Under the Supervision of:

Ajay Pal

Outline

- Introduction to Project
- Problem Formulation
- Objectives of the work
- Methodology used
- Results and Outputs
- Conclusion
- Future Scope
- References

Introduction to Project

Rapid development of mobile devices and internet has made possible for us to access different music resources freely. The number of songs available exceeds the listening capacity of single individual. People sometimes feel difficult to choose from millions of songs. There is a strong need of a good recommendation system. Currently, there are many music streaming services, like Pandora, Spotify, etc. which are working on building high-precision commercial music recommendation systems. These companies generate revenue by helping their customers discover relevant music and charging them for the quality of their recommendation service.

Thus, there is a strong thriving market for good music recommendation systems. Music recommender system is a system which learns from the users past listening history and recommends them songs which they would probably like to hear in future. We have implemented various algorithms to try to build an effective recommender system. We firstly implemented popularity based model which was quite simple and intuitive. Collaborative filtering algorithms which predict (filtering) taste of a user by collecting preferences and tastes from many other users (collaborating) is also implemented. We have also done experiments on content based models, based on latent factors and metadata.

Problem Formulation

In this project, we have designed, implemented and analyzed a song recommendation system. We used Million Song Dataset[1] provided by Kaggle to find correlations between users and songs and to learn from the previous listening history of users to provide recommendations for songs which users would prefer to listen most.

In this project, we will discuss the problems we faced, methods we have implemented, results and their analysis. We have got best results for memory based collaborative filtering algorithm. We believe that content-based model would have worked better if we would have enough memory and computational power to use the whole available metadata and training dataset.

Objectives of the Work

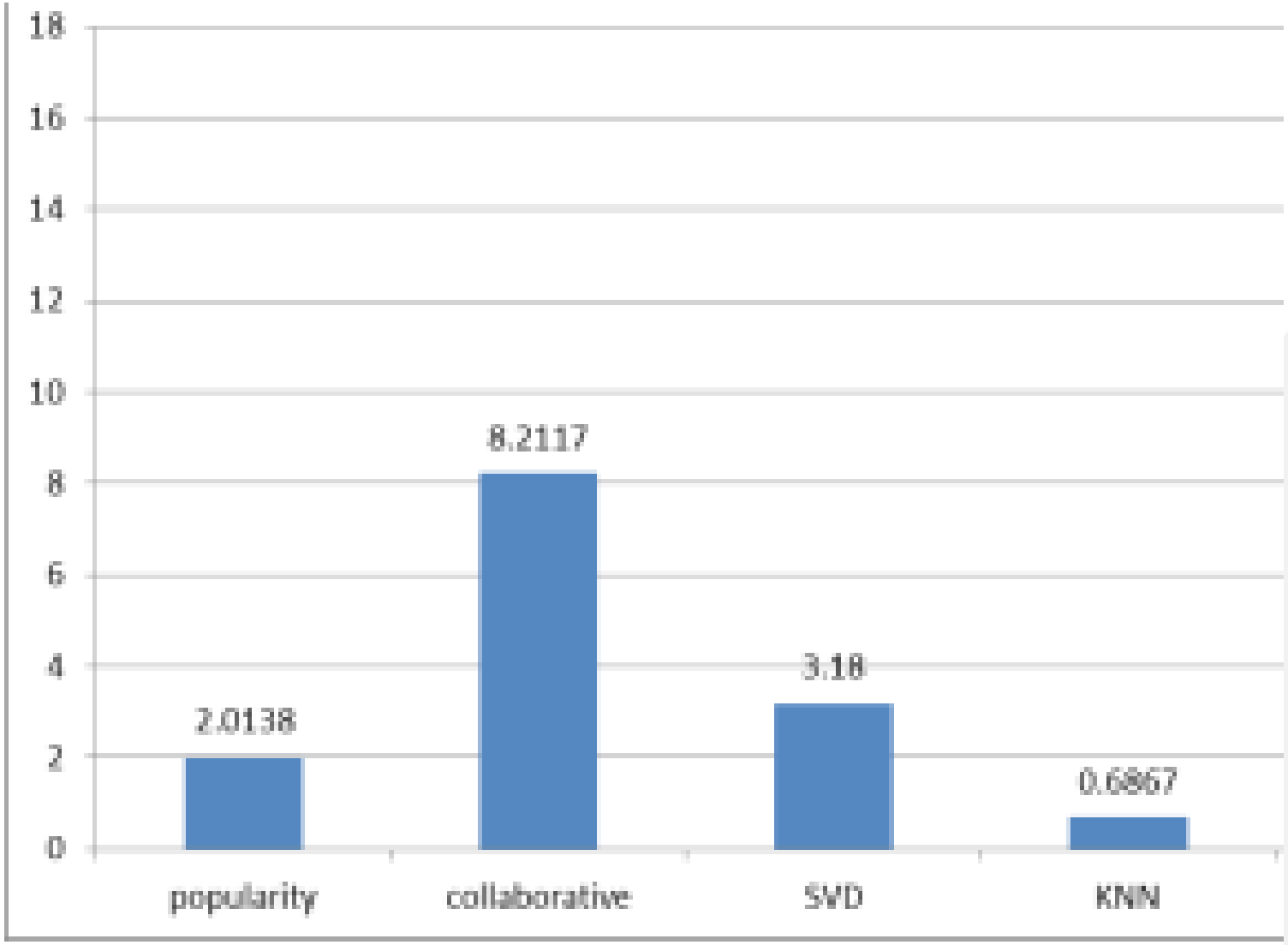
Rapid development of mobile devices and internet has made possible for us to access different music resources freely. The number of songs available exceeds the listening capacity of single individual. People sometimes feel difficult to choose from millions of songs. Moreover, music service providers need an efficient way to manage songs and help their costumers to discover music by giving quality recommendation. Thus, there is a strong need of a good recommendation system. In this project, we have designed, implemented and analyzed a song recommendation system.

We used Million Song Dataset provided by Kaggle to find correlations between users and songs and to learn from the previous listening history of users to provide recommendations for songs which users would prefer to listen most in future. Due to memory and processing power limitations, we could only experiment with a fraction of whole available dataset. We have implemented various algorithms such as popularity based model, memory based collaborative filtering, SVD (Singular Value decomposition) based on latent factors and content based model using k-NN. Memory based collaborative filtering algorithm gave maximum mean average precision. We believe that content-based model would have worked better if we would have enough memory and computational power to use the whole available metadata and dataset.

Methodology used

- We have implemented four different algorithms to build an efficient recommendation system
- It is the most basic and simple algorithm. We find the popularity of each song by looking into the training set and calculating the number of users who had listened to this song. Songs are then sorted in the descending order of their popularity
- Collaborative based Model
  - Collaborative filtering involves collecting information from many users and then making predictions based on some similarity measures between users and between items

- SVD Model
  - Listening histories are influenced by a set of factors specific to the domain (e. g. genre, artist). These factors are in general not at all obvious and we need to infer those so called latent factors[4] from the data. Users and songs are characterized by latent factors.
- KNN Model
  - In this method, we utilize the available metadata. We create a space of songs according to their features from metadata and find out neighborhood of each song. We choose some of the available features (e.g., loudness, genre, mode, etc.) which we found most relevant to distinguish a song from others.
- Evaluation Metrics
  - We used mean Average Precision (mAP) as our evaluation metric. The reason behind using this is that this metric was used in the Kaggle challenge which helps us to compare our results with others. Moreover, precision is much more important than recall because false positives can lead to a poor user experience.


18
16
14
12
10
8
6
4
2
0
8.2117
3.18
2.0138
0.6867
popularity
collaborative
SVD
KNN

Results and Outputs

- Upto 10 slides depending upto the project results.

Conclusion

- This is a project of our Artificial Intelligence course. We find it is very good as we got a chance to practice theories that we have learnt in the course, to do some implementation and to try to get a better understanding of a real artificial intelligence problem: Music Recommender Sys4 tem. There are many different approaches to this problem and we get to know some algorithms in detail and especially the four models that weve explained in the paper. By manipulating the dataset, changing the learning set and testing set, changing some parameters of the problem and analyzing the result, we earn a lot practicing skills. Weve faced a lot of problem in dealing with this huge dataset, how to explore it in a better way and we also had difficulties in some programming details. However, with lot of efforts, we have overcame all of these

Future Scope

- Mention what advancements are possible with respect to your work in future.
- Use 1 slide for this

References

- Font Size for List of references to be used is *16* with Times New Roman.
- 2-3 slides to be used.