MUSIC RECOMMENDATION SYSTEM USING MACHINE LEARNING ALGORITHMS

Lalithakumari Datti, Scholar Andhra University College of Engineering for Women, Visakhapatnam, Andhra Pradesh :: datti.kumari@gmail.com

Dunaboyana Srujana, Sushmita Gandi, Velala Priya, Y Harshitha Dhanapriya, Yoga Esarapu
Andhra University College of Engineering for Women, Visakhapatnam, Andhra Pradesh

ABSTRACT

In recent times,we can see that the Recommendationbasedsystems have became increasingly popular. They are utilized as a key in lots ofbusinesses for their benefit like Amazon and flip kart for selling products (e-commerce), wynk music and ganna.com for music streaming, for selling books, for movies, YouTube for videos recommendations.There are two ways to produce recommendation collaborative filtering and content based models.These algorithms aim to find similarity between various users various songs and artists. Other than these algorithms we have also used random forest algorithm and decision tree algorithm. Both of these algorithms aims to predict a decision based on different attributes. For improving the efficiency and accuracy of the model we have also used cross validation technique in which we have recursively changed the training and testing data set to get optimum results. For a specific user we had to consider thier song history frequency list liked songs. From all this information we had to predict what songs user might like.In this paper we have proposed a music recommendation system based on collaborative filtering method and simple method of ranking movies according to their popularity rating. The system has been developed using Python and machine learning algorithms.

Keywords : *Machine Learning, Collaborative filtering,Content Based Filtering, K-Nearest Neighbour, Cosine Similarity.*

1. INTRODUCTION

Nowadays lots of music industries like amazon music, wink music,gaana.com are using recommender systems and the old fashioned way of selling music has changed to a totally different cloud based .Now all the music resources are present in their cloud and users can listen to the songs directly from the cloud. But the issue is there are lot of songs present in the cloud system. so we need to classify all the songs based on different genres ,artists locations , age groups, languages and the main goal is to classify these set of songs in accordance to the taste of the user. Because user expects valuable return after the investment of time as well as money thereby we can attract a lot of customers by providing various valuable services of their interests For this project we are using various machine learning algorithms as well as data mining techniques. We have implemented various algorithms and compared the results with one another to find the effective algorithm that suits our model. The most common approaches people have used implementing various recommender systems are collaborative filtering and content based models. These algorithms aim to find similarity between various users various songs and artists. Other than these algorithms we have also used random forest algorithm and decision tree algorithm. Both of these algorithms aims to predict a decision based on different attributes. For improving the efficiency

and accuracy of the model we have also used cross validation technique in which we have recursively changed the training and testing data set to get optimum results. We have faced a lot of problems which making this project like the data we have chosen was too big (1.2 GB) so we had to create a subset of that data set for optimum usage. There were lot of outliers present in that data set which required complex data pre-processing.

Figure 1.(Recommendation Systems)

2.Algorithms:

2.1 Collaborative filtering:

Collaborative-filtering is commonly known as a social filtering process in which it generally filters the content and information by the use of recommendations of other users in similar domain. The basic idea behind this technique is that users who agree and are interested in the evaluation of a specific item in previous history or purchase in the form of use, access and purchase are more likely to agree or match again in future. A user willing to listen particular music for example can ask or search for recommendation from friends and people having similar tastes and interests.

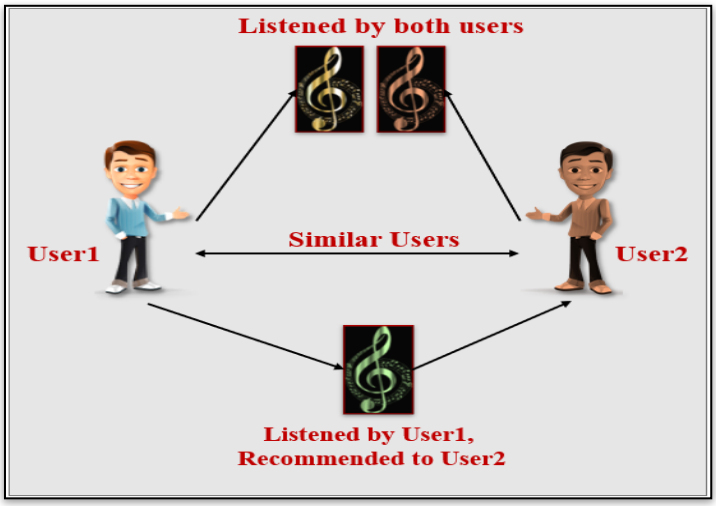

Figure 2.1(Collaborative filtering)

Content based filtering is the most popular algorithm when it comes to recommender systems. It is common sense that if a user listens to a song and another song is similar to that one then user may also like that also. Basic idea behind this is to predict on the basis of similarity .

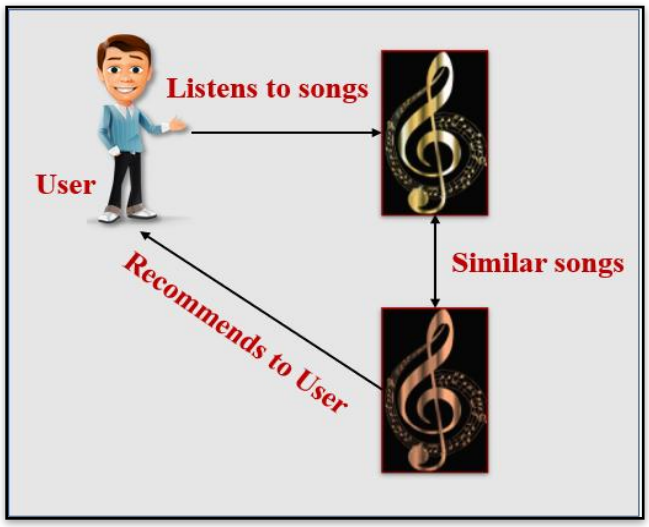

Figure 2.2(Content Based Filtering)

i. Item Profile Creation: Here initially, item profile is created in order with the help of it's feature. In case of movie, music meta data available can be used for item profile creation. 17
ii. User Profile Creation: User profile is created , based on their interaction with the item for example with the help of their rating on the items. Hence user profile is created with the help of the item profile either by taking the average of item-profile or weighted average of item profile.

2.3 K NEAREST NEIGHBOR

The k means is a generic and easy to implement clustering-algorithm with minimal complications and complexities. In clustering algorithms, the data is accepted and taken as input data as input and then it will use the mathematical precision of calculations and methods to search particular groups which are of similar items as predefined as input or users as using that particular data as predefined in input. For example supposing we are provided a set of 6 people having age group as follows: 4,7,17,27,46,48. This information defines number of persons and the ages of those particular persons .If we are asked to split that group into sets, generally one will split the given data into minors (5,7,17) and non-minors(44 and 68).This type of separation or classification is based on a factor which is age .But we cannot call this clustering but this what exactly is being done: clustering like with similar based on the data set given. Clustering-algorithms such as k-means algorithm generally perform the same thing as

defined earlier, but with huge amount of data at a large amount of scale. It is a good and easy method to arrange andaugment.

This algorithm aims to group together various songs or various users based on various attributes. This first finds the similarity between objects and then group them together. These groups are named as clusters. One thing about this algorithm is that previously unknown clusters or groups can be discovered using this. Nearest cluster for each object is calculated and then they are assigned the nearest cluster. In k means clustering we are provided with a set of data items relevant or irrelevant to the domain, with specific attributes. Our primary goal is to arrange and categorize these set of items into groups based on given constraints and parameters. For us to attain this, we will have to use the k-Means algorithm to obtain required and optimum results.

The algorithm will make category of the items into k groups of similarity. This algorithm will work as in the given order:

- Firstly, initializing of k points will be done, which is done uncertainly or randomly. Further we will the categorize all items to the closest mean and then update the mean'scoordinates,whicharetheaveragesofitemscategorizedinthatmeansofar.
- We the repeat the process for any given number of iterations and then at end, we have ourclusters.

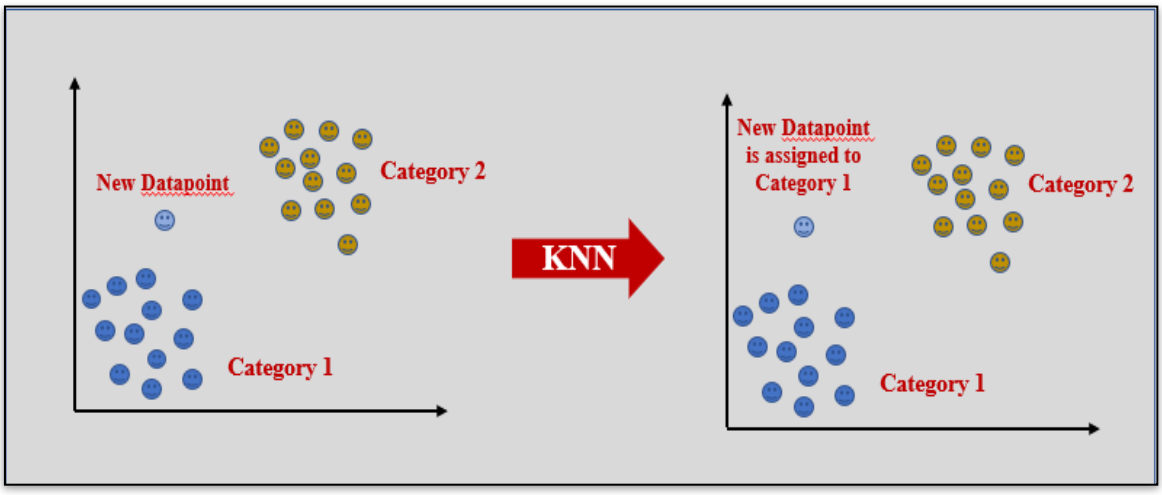

Fig 7: K Nearest Neighbor

Algorithm

We are describing the process to show how it works. W takes care of all with only a single line of code. In order to keep the procedure straight and simple, we will us now consider 30 users in a given Training set or given set of Class, each user will have 2 or more than 2 scores. We now will create separate non-identical clusters based on the scores of the first tests performed.

- Now this is how the k-means algorithm will work with this particularsetup:
- It will firstly randomly it will take 2 or more than 2 different points that will serve as initialcluster.
- Then it will calculate the distance between each of the given 30 data points from the given eachcenter.
- Now it will allot each of the data points to the cluster with the closestcenter.
- Now recalculate each of given centers of clusters using just those given data points that were assigned tocluster.
- Recalculate distance of each of 20 data points from each of new cluster center.
- Itwillthenreassigneachofthedatapointstotheclusterandthenwithclosestcenter.

Further it will continue repeating process till no such data points need to be assigned again even after the calculation of the new cluster centers has beendone.

Does k- means produces optimumresult?

The primary stage is randomly choosing starting cluster center locations. This means that the received output can sometimes be affected and changed to some degree level by that particular starting point. In some rare cases, the outcome can be substandard to other clustering outputs. This means all clustering outputs will not necessarily be same or identical. In practice, we should firstly set up a random input and then reproduce the result. Moreover, in formation of this post, very commonly we receive undesirable results. The method to obtain desired results was to set up random seeds orinput.

Advantages Of K-MeansClustering

- Quick, robust and easy to analyze for effectiveimplementation

- Relatively it is more effective andoptimum
- Gives more desirable results when the data set is different or well separated from one another

2.4 COSINE SIMILARITY

Cosine similarity is a measure of similarity between two non-zero vectors of an inner product space that measures the cosine of the angle between them. Similarity = (A.B) / (||A||.||B||) where A and B are vectors. The cosine similarity is beneficial because even if the two similar data objects are far apart by the Euclidean distance because of the size, they could still have a smaller angle between them. Smaller the angle, higher the similarity.

When plotted on a multi-dimensional space, the cosine similarity captures the orientation (the angle) of the data objects and not the magnitude.

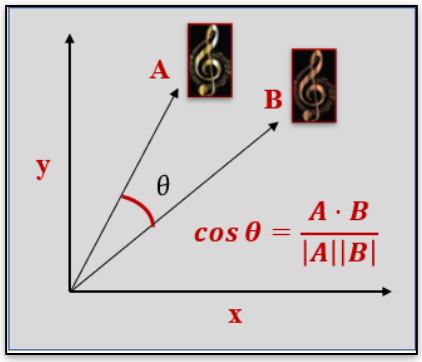

Fig 8: Cosine Similarity Measure

Output

```
In [20]: recommedations.recommend(recommendation)

The 4 recommended songs for Children Go Where I Send Thee are:
Number 1:
Born To Be Loved By You by Roy Orbison with 0.482 similarity score
--------------------
Number 2:
Born To Love by Nazareth with 0.359 similarity score
--------------------
Number 3:
Born To Run by Neil Young with 0.299 similarity score
--------------------
Number 4:
Born To Lose by Zebrahead with 0.271 similarity score
--------------------
```

Figure 5(Result)

5. CONCLUSION

There are of the recommender system is vast and covers various parameters. It is developing and emerging in modern day generation of e services and commerce. But, simultaneously there is a need to develop and optimize the working and output of recommender system. Several service providers facilitate the users with a list of items. But this is not sufficient because customers have different preferences and choices which may mainly depend upon various factors and constraints. Also, in many cases it may not be possible to recommend specific items to particular users. Therefore, there is a scope for incorporating the concept of multiple dimensions in music recommender systemparticularly.

Most of the products and services provided by the various e-commerce sites are expensive and therefore less used by customers. This leads to the inability to rate an item or set of items correctly and specifically. Thus, traditional recommender system techniques are not optimum. This lays a way towards further work and development in building an optimized recommender system which also takes into considerations the feedback and all other constraints related to recommendation.

.

6.REFERENCES

1. Álvarez, P., F. J. Zarazaga-Soria, and S. Baldassarri. "Mobile music recommendations for runners based on location and emotions: The DJ-Running system." *Pervasive and Mobile Computing* (2020): 101242.

2. Patel, Ashish, and Rajesh Wadhvani. "A comparative study of music recommendation systems." *2018 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*. IEEE, 2018.

3. Volokhin, Sergey, and Eugene Agichtein. "Understanding music listening intents during daily activities with implications for contextual music recommendation." *Proceedings of the 2018 Conference on Human Information Interaction & Retrieval*. 2018.

4. Sarda, Pranav, et al. "Emousic: emotion and activity-based music player using machine learning." *Advances in Computer Communication and Computational Sciences*. Springer, Singapore, 2019. 179-188.

5. Nassar, Nour, Assef Jafar, and Yasser Rahhal. "A novel deep multi-criteria collaborative filtering model for recommendation system." *Knowledge-Based Systems* 187 (2020): 104811.

6. Klec, Mariusz, and Alicja Wieczorkowska. "Music Recommendation Systems: A Survey." *Recommender Systems for Medicine and Music* 946 (2021): 107.

7. Nassar, Nour, Assef Jafar, and Yasser Rahhal. "Multi-criteria collaborative filtering recommender by fusing deep neural network and matrix factorization." *Journal of Big Data* 7 (2020): 1-12.

8. Paul, Dip, and Subhradeep Kundu. "A survey of music recommendation systems with a proposed music recommendation system." *Emerging Technology in Modelling and Graphics*. Springer, Singapore, 2020. 279-285.

9. Åman, Pirkka, and Lassi A. Liikkanen. "A survey of music recommendation aids." *Proceedings of the 1st Workshop on Music Recommendation and Discovery (WOMRAD) colocated with ACM RecSys*. 2019.

10. Paul, Dip, and Subhradeep Kundu. "A survey of music recommendation systems with a proposed music recommendation system." *Emerging Technology in Modelling and Graphics*. Springer, Singapore, 2020. 279-285.

