SONG RECOMEDATION SYSTEM USING COLLABORATIVE FILTERING

[1] S.INIYAN, [2] MOHIT JAIN, [3]ABHILASH SINGH

ABSTRACT--*A recommender system could be a scientific categorization of information separating a system that anticipate the "rating" or "inclination" a client would provide for a thing. Recommender structures (RS) use man-made thinking (AI) methodologies to outfit customers with things recommendations. For example, an online bookshop may use an AI (ML) figuring to describe books by type and after that endorse various books to a customer acquiring a specific book. With the ascent of computerized content conveyance, individuals presently approach music assortments on an exceptional scale. Business music libraries effectively surpass 15 million tunes, which incomprehensibly surpasses the listening ability of any single individual. With a large number of tunes to look over, individuals some of the time feels overwhelmed. Most normal RS are planned to utilize the idea of sifting methods and manage the tally and similitudes between the resemblances of the clients. Our methodology, right now, to upgrade the RS by consolidating the separating system with Collaborative Filtering.*

Keywords—*recommendations, system, using, collaborative, filtering*

I. INTRODUCTION

One of the prime features of web 2.0 is that it allows users to share with other users their opinions and viewpoints about almost everything on internet. Getting someone else's validated point of view can be of reasonable advantage with regards to choosing whether or not to contribute time, cash or exertion into something. This is one of the main thrusts behind the expanding accomplishment of network sites which permit enrolled clients to compose and peruse surveys about business items, for example, music, books, motion pictures, or buyer gadgets, for example, for example cameras or mobile phones. Client appraisals frequently comprise of a free – content survey and a general rating.

Recommender frameworks rose to manage the data over-burden issue by creating customized content recommendations to their clients. Given the present situation of the Web, where clients can give content by delivering explanations, remarks and surveys about any subject, there is a lot of rich and point by point data accessible that is made cooperatively by the network. Despite its unstructured and uncontrolled nature, client made

[1] *Assistant Professor, Department of Computer Science & Engineering SRM Institute of Science & Technology SRM Nagar, Kattankulathur, Kancheepuram, Tamil Nadu, India,603203*

[2] *UG Student, Department of Computer Science &Engineering SRM Institute of Science &Technology, SRM Nagar, Kattankulathur, Kancheepuram, Tamil Nadu, India,603203*

[3] *UG Student, Department of Computer Science & Engineering SRM Institute of Science & Technology SRM Nagar, Kattankulathur, Kancheepuram, Tamil Nadu, India,603203*

portrayals can be abused by data recovery and recommender frameworks errands, the need of space specialists to make organized metadata about

the things (ordering). In addition, one can generally acquire refreshed depictions about recently included things, which can shift over the time contingent upon the setting they are embedded (for example news about decided occasion will undoubtedly fluctuate quick, while depictions about motion pictures and books may have little variety through time).

II. PROPOSED METHODOLOGY

Generally, we might want to purchase an item that companions or partners have recommended. So how about we consider another case of a book shop caused unique to notice New Collection books, well-known books and so on. So the purchaser can rapidly pick a book. In a computerized world utilizing these sorts of techniques as suggestion frameworks, the item proprietor can prescribe things that clients may likewise enjoy and required. A proposal framework is a broad class of web applications that includes anticipating the client's reactions to the alternatives.

Recommender Systems are fundamentally calculations that expect to give the most pertinent and exact things to the client by sifting valuable stuff from a tremendous pool of data base. Proposal motors find information designs in the informational collection by learning shoppers' decisions and produces the results that co-identifies with their needs and premiums.

In Real-time models resemble Amazon, they have been utilizing a proposal motor for recommending the merchandise or items that clients may likewise like.

A. DATA COLLECTION

Data collection is an organized way to deal with gathering and putting away data from an assortment of sources to get a total and exact image of the ideal zone. Data collection permits a substance or an endeavor to respond to important inquiries, survey results and make expectations about future results and patterns. Precise Data collection is essential to keeping up the integrity of research works, settling on taught business choices and guaranteeing quality. Reviews, meetings and open posts are essential devices for social event data. Today, with assistance from Web associations are likewise ready to gather information from cell phones, site logs, and so forth. The information can be gathered from different open-source sites, web-based social networking pages utilizing web scratching or from unequivocal and understood databases.

The dataset used in the current work was collected from an opensource website http://millionsongdataset.com/. The dataset contains 2 files: One contains the song_id, user_id and listen count and the other contains the song details. The dataset contains features like song_id, user_id, listen_count, song name, album name, artist, year, etc.

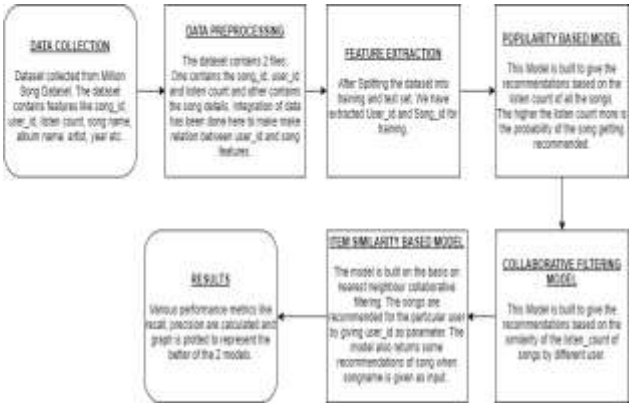


Figure 1: System Architecture.

B. DATA PREPROCESSING

Data preprocessing is an information mining process that includes adjusting crude information into a justifiable configuration through a progression of predefined techniques and steps. Information preprocessing is required on the grounds that Real-World Data are commonly loud, deficient, conflicting, and so forth. Errands in information preprocessing by and large incorporate data cleaning, transformation, and reduction.Integration of data has been done here to make make relation between user_id and song features.

C. FEATURE EXTRACTION

Feature extraction is a strategy of dimensionality decrease utilizing which an underlying arrangement of crude info information is diminished to progressively sensible dimensionally diminished gatherings for preparing purposes. Feature extraction is the assortment of strategies that select or potentially join factors into highlights, adequately lessening the measure of information that must be handled but at the same time unequivocally and exhaustively portraying the first informational collection. The procedure of feature extraction is valuable when one plans to cut down the number of utilities required for handling without missing out on significant data. We have extracted User_id and Song_id for training.

D. Popularity based Model

Popularity based model is basically used to generate recommendation on the basis of popularity which means the item which is most popular are recommended to the user. This approach is helpful in cases when large audience likes the same product in our case, songs. But if the likes and disliked of songs differ heavily then it is not very efficient. In this project, the popularity based model ranks the songs according to the listen_count of the songs and then recommends the songs to the user.

E. Collaborative Filtering

Collaborative Filtering uses clients' authentic inclination on a lot of things. Since it depends on verifiable information, the center supposition here is that the clients who have concurred in the past will in general additionally concur later on. Regarding client inclination, it generally communicated in two classes. Express Rating are kind of points given by a client to a thing on a numerical scale, similar to 5 stars for a particular song or a movie. This is the immediate input from clients to show that they like a thing. By understanding Ratings, recommends clients' inclination in a roundabout way, for example, site visits, clicks, buy records, regardless of whether tune in to a music track, etc.

III. NEAREST NEIGHBORHOOD

The most typical method of Collaborative Filtering is called the Nearest Neighborhood algorithm. The nearest neighborhood algorithm makes user of item-similarity based collaborative filter.

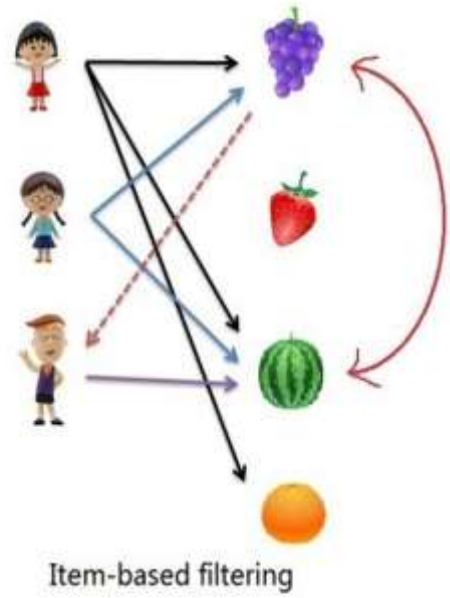


Figure 2: Item-Similarity Based Collabrative Filtering

In Item-based CF, two things are comparative when they got comparative evaluations from an equivalent client. At that point, we will make forecast for an objective client on a thing by computing weighted normal of appraisals on most X comparative things from this client. One key favorable position of Item-based CF is the solidness which is that the evaluations on a given thing won't change fundamentally extra time, in contrast to the interest of people. Likeness or resemblance between items are calculated using cosine similarity metric. The rating for target item i for an active user a can be calcilated by using a simple weighted average as

$$p_{a,i} = \frac{\sum_{j \in K} r_{a,j} w_{i,j}}{\sum_{j \in K} |w_{i,j}|}$$

where K is the neighborhood of most similar items rated by active user a, and w(i,j) is the similarity between items i and j.

IV. MATRIX FACTORIZATION

Since adaptability and sparsity are the 2 greatest difficulties for typical collaborative filtering strategy, it is a further developed technique which decay the first scanty framework to low-dimensional grids with inactive elements or highlights (latent features or factors) and decreased sparsity. This is called as Matrix Factorization. Apart from figuring out how to solve the sparsity and scalability problem, a very good reason exists why we should represent users' preference using the low-dimensional matrices. Suppose a user gives a good ratings to songs Love Me by Justin Bieber, Somebody to Love by Justin Bieber, U Smile by Justin Bieber. These are necessarily 3 different views or opinions but shows that the particular user likes song from Justin Bieber and may ne more songs Justin Bieber that the user might like so we should recommend those songs to the user. Unlike particular songs, the latent factors are showed using help of higher level attributes and one of the latent feature of factor in this case is the Artist's name. What matrix factorization in the long run gives us is how a lot of a client is lined up with a lot of idle highlights, and how much a motion picture fits into this arrangement of idle highlights. This provides a edge to the matrix factorization method over nearest neighbour method that even same songs are not rated by two users, it's still very plausible to find similarity or connection between two users because of the reason that they sharing common underlying or similar underlying interest.

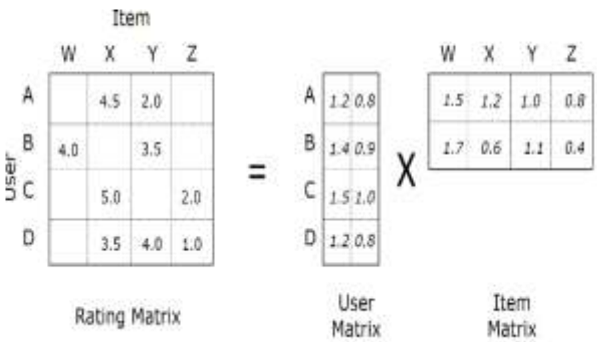


Figure 3: Matrix Factorization

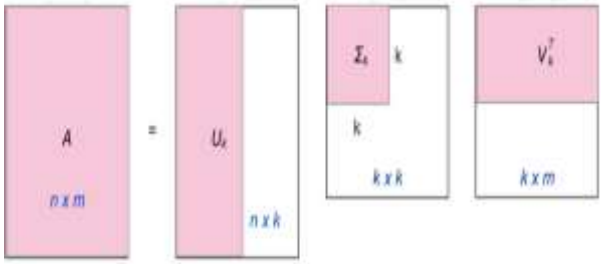
To understand how a matrix factorization works, there is a need to learn or understand what Singular Value Decomposition (SVD) is. According to Linear Algebra concepts, a matrix R(real matrix) may be broken down or decomposed into 3 matrices namely V, U, and Σ. Taking the example of songs, V is the value of m x r songs latent feature matrix, the user latent-feature matrix U is value r x n, Σ is r x r diagonal matrix which contains the singular values of original matrix R to represent how a specific feature is of very great value in predicting the user preference.

International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 06, 2020
ISSN: 1475-7192

$$R = U\Sigma V^T$$

$$U \in IR^{n \times r}, \quad \Sigma \in IR^{r \times r}, \quad V \in IR^{r \times m}$$

To grade the Σ matrix values by reducing absolute value and shorten matrix Σ to first k singular values, there is a need to rebuild the matrix as matrix A. The k singular values should be selected such that most of dissimilarity or variation within the original matrix R should be captured in A, so that A ≈ R i.e., A is the approximation of R. The variation should be minimal between A and R. This is actually the idea of Principle Component Analysis.



V. RESULTS

The experiments are performed on Intel i5 processor with 4GB memory. A huge database of 2000000 entries is used. Out of which around 200000 entries are taken into consideration for the experimental purposes. The experiment is conducted on 200000 entries contains close to 7000 unique users and 10000 unique songs data.

The recommendation are provided to the user according to the input provided by the user. The input in this case are mostly user_id and song_name. The top 10 recommendation are provided to the user in both cases and it's accuracy can be determined by the satisfaction of the user. The user if likes the song then the recommendations are efficient.

While working with popularity based model the top 10 recommendations generated were as shown:

	user_id	song	score	Rank
6413	[illegible]	Sehr kosmisch - Harmonia	163	1.0
8173	[illegible]	Undo - Björk	142	2.0
1849	[illegible]	Dog Days Are Over (Radio Edit) - Florence + Th...	139	3.0
8881	[illegible]	You're The One - Dwight Yoakam	134	4.0
6095	[illegible]	Revelry - Kings Of Leon	128	5.0
6483	[illegible]	Secrets - OneRepublic	123	6.0
2424	[illegible]	Fireflies - Charttraxx Karaoke	107	7.0
3230	[illegible]	Horn Concerto No. 4 in E flat K495: II. Romance...	104	8.0
7969	[illegible]	Tive Sim - Cartola	97	9.0
3196	[illegible]	Hey_ Soul Sister - Train	88	10.0

Figure 4: Popularity based recommendation.

Similarly, while working with item similarity based model the top 10 recommendations generated for particular user when given user_id as input and when song_name is givn are shown:

Received: 22 Feb 2020 | Revised: 13 Mar 2020 | Accepted: 05 Apr 2020

(a) user_name

```
song = 'The Real Slim Shady - Eminem'
is_model.get_similar_items([song])
```

```
no. of unique songs in the training set: 8958
```

	user_id	song	score	rank
0		My Name Is - Eminem	0.203704	1
1		Mockingbird - Eminem	0.203125	2
2		Without Me - Eminem	0.154930	3
3		Forgot About Dre - Dr. Dre / Eminem	0.152174	4
4		Terre Promise - O'Rosko Raricim	0.132075	5
5		Superman - Eminem / Dina Rae	0.130435	6
6		I'm Back - Eminem	0.128205	7
7		Eenie Meenie - Sean Kingston and Justin Bieber	0.109375	8
8		Te Amo - Rihanna	0.107143	9

(b) song_name

Figure 5: Item-simailarity based recommendations on (a)user_name (b)song_name

To compare the popularity based model and collaborative filtering model on the basis of efficiency we have used performance metrics like precision and recall.

VI. CONCLUSION

The music recommender system will be large-scale and personalized. Main idea is to learn from user's listening history and features of songs and predict songs that a user would like to listen to. On comparing, the item-based collaborative filtering approach outperforms the popularity based model. Matrix factorization is also applied on the small subsets. Future work for the project will include making the matrix factorization method using SVD efficient while working on large number of entries and combining the image processing and collaborative filtering to give much more accurate recommendations by performing CNN algorithms on the album art of the songs.

REFERENCES

1. Hyun-Tae Kim, Eungyeong Kim, Jong-Hyun Lee and Chang WookAhn. "A Recommender System Based on Genetic Algorithm for Music Data." 2010 2nd International Conference on Computer Engineering and Technology (Vol. 6, pp. V6-414), IEEE.
2. Kunal Shah, AkshaykumarSalunke, SaurabhDongare, KisandasAntala. "Recommender Systems: An overview of different approaches to recommendations." In Innovations in Information, Embedded and Communication Systems (ICIIECS), 2017 International Conference on (pp. 1-4). IEEE.
3. Joonseok Lee, Kisung Lee, Jennifer G. Kim. "Personalized Academic Research Paper Recommendation System". Proceeding at SRS'15, August 10, 2015, Sydney, NSW, Australia.
4. PawelMatuszyk., Myra Spiliopoulou. "Stream-based semi-supervised learning for recommender systems. Gastroenterology." Machine Learning June 2017, Volume 106, Issue 6, pp 771–79.
5. Nakamura, K., Fujisawa, T., &Kyoudou, T. (2017, October)."Musicrecommendation system using lyric network " In Consumer Electronics (GCCE), 2017 IEEE 6th Global Conference on (pp. 1-2), IEEE.
6. Nikhil Sonavane, Ms. SonaliSonavane. "Song Recommender System Using Machine Learning." National Conference on Computational Intelligence and Deep Learning (NCCIDL - 19), International Journal on Future Revolution in Computer Science & Communication Engineering Volume: 5 Issue: 2, pages 22 – 25.
7. Young Sung Cho, Song Chul Moon, Si Choon Noh, KeunHoRyu. "Implementation of personalized recommendation system using k-means clustering of item category based on RFM." 2012 IEEE International Conference on Management of Innovation & Technology (ICMIT).
8. Hao Wang, Naiyan Wang, Dit-Yan Yeung. "Collaborative Deep Learning for Recommender Systems." KDD '15 Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Pages 1235-1244.
9. Rohini Nair, KavitaKelkar. "Implementation of Item and Content based Collaborative Filtering Techniques based on Ratings Average for Recommender Systems." International Journal of Computer Applications (0975 – 8887) Volume 65– No.24, March 2013
10. Lee, J., Shin, S., Jang, D., Jang, S. J., & Yoon, K. (2015, January). "Music recommendation system based on usage history and automatic genre classification." In Consumer Electronics (ICCE), 2015 IEEE International Conference on (pp. 134-135). IEEE
11. YasharDeldjoo, Markus Schedl, BalázsHidasi, Peter Knees. "Multimedia Recommender Systems." ,RecSys '18 Proceedings of the 12th ACM Conference on Recommender Systems, Pages 537-538.

12. Hong-Jian Xue, Xin-Yu Dai, Jianbing Zhang, Shujian Huang, Jiajun Chen. “Deep Matrix Factorization Models for Recommender Systems”. IJCAI'17 Proceedings of the 26th International Joint Conference on Artificial Intelligence. Pages 3203-3209.
13. VishwaGosalia, BhaveshChatnani, Aman Gupta, PranalSoni “Movie Recommendation System”, International Research Journal of Engineering and Technology (IRJET) Volume: 05 Issue: 02 | Feb-2018.
14. JianxunLian, Fuzheng Zhang, Xing Xie, Guangzhong Sun. “CCCFNet: A Content-Boosted Collaborative Filtering Neural Network for Cross Domain Recommender Systems”. Proceeding at WWW '17 Companion Proceedings of the 26th International Conference on World Wide Web Companion Pages 817-818.
15. Ruchika, Ajay Vikram Singh, Mayank Sharma "Buildingan Effective Recommender System Using Machine Learning Based Framework." In Consumer Electronics (ICCE), 2015 IEEE International Conference on (pp. 134-135). IEEE.