RESEARCHERID
THOMSON REUTERS

ICRTCET-2018]

ISSN 2348 – 8034
Impact Factor- 5.070

GLOBAL JOURNAL OF ENGINEERING SCIENCE AND RESEARCHES

COLLABORATIVE FILTERING RECOMMENDER SYSTEMS FOR MUSIC RECOMMENDATION

Chaitanya Varma
Department of Computer Science and Engineering, Manipal Institute of Technology, Manipal

ABSTRACT

With a vast growth of electronic data in this day and age, there is an increased demand for effective information retrieval and filtering tools. Information filtering systems such as recommendation systems are employed to foretell the bias or ratings for an item given by a user. Though there are several techniques for recommendation systems, collaborative filtering is highly effective to produce recommendations and is therefore very popular. Collaborative filtering is a method which emphasises the relationships amongst users and items to make predictions. Two approaches are commonly employed in collaborative fil-tering - user-based collaborative filtering along with item-based collaborative filtering. For user based collaborative filtering, the users recommendations are computed by determining other users with similar interests, while in item-based collaborative filtering, a comparison between the set of items and their ratings are computed to analyze the similarity to the target item undergoing investigation. A recommendation system computes a scoring function which consolidates the result of computing similarities between users and between items. In this paper, we have analysed the collaborative filtering technique for music recommendation systems and proposed a method to provide recommendations to listeners which are accurate and relevant to the user.

Keywords: *Collaborative Filtering, Recommender Systems, Music Recommendation*.

I. INTRODUCTION

A. Recommendation Systems and Their Goals

A recommendation system studies the past actions per-taining to a user to suggest user-specific objects that the user might be interested in. There are several algorithms which aim at building an effective recommendation system. Recommendation systems incorporate several steps to produce valid recommendations right from extracting information re-garding the user in question until the final recommendations are delivered to the user. Recommendation systems aim to ease human interactions with applications by increasing the degree of activity and enhancing the user experience. The ultimate goal of employing recommendation systems is to boost sales of products and revenue.

Thus the goals of recommendation systems are as follows.

- The recommender system must provide recommenda-tions pertinent to the user's interests: The suggestions made by the recommender system should have relevance to the user's interests.
- The recommender system must offer innovative rec-ommendations: Recommendation system should suggest items which are new to the user and not the content that he has already consumed.
- The recommender system must offer unexpected, interesting suggestions: The recommendation system should offer suggestions to the user which should help him discover items which may be of interest to him and at the same time, helping the user expand his/her interests and hence lead to diversification of sales for the company.
- The recommender system must incorporate variety in the recommendations made: The recommendation system should try to offer variety in the suggestion, oth-erwise, the user may get bored with the recommendations being limited to a specific group or genre.

B. Need for Recommendation Systems

The proliferation of mobile devices and penetration of the internet amongst customers has led to the development of several music applications which offer listening services with vast content. The content is of such a great capacity that it is not a viable option for anyone to listen to every song listed on the application to decide whether he likes it or not. Hence, it is difficult for users to choose songs to listen to from the vast music libraries. Furthermore, these music applications require an organized method to suggest songs to their cus-tomers and hence improve the overall listening experience by offering quality recommendations. Hence, there is a definite requirement of good recommendation systems in the market. Companies such as Spotify, Saavn, last.fm are conducting research on accurate music recommendation systems, and they need to help their customers discover relevant music and thus maintain their customer base [2].

C. Literature Review and Related Work

In [3]Elena Shakirova investigated the use of collaborative filtering methods for music recommendation systems and com-pared various evaluation metrics which estimate usefulness of the recommendation system.

In [4]Shefali Garg and Fangyan Sun put forth a design and implementation for a music recommender system and analyzed its performance. In their paper, they were under the belief that a content-based model could have superior performance as compared to collaborative filtering if there weren't constraints on memory and computing ability to employ the entire meta-data and training dataset.

In [8]Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl analyzed various item based recommender algo-rithms and looked into methods to compute item similarities such as item correlation and cosine similarity measure along with various methods to obtain recommendations such as weighted sum and regression methods.

II. THE APPLICATION OF COLLABORATIVE FILTERING FOR RECOMMENDER SYSTEMS

Collaborative filtering predicts the preferences of users using instances of their past preferences coupled with the preferences of other comparable users. Collaborative filtering achieves this by emphasizing the relationship amongst objects. For conducting collaborative filtering, we need to analyze the correlations amongst various items in question through a variety of methods such as calculating the tanimoto coefficient or estimating the Spearman rank correlation. After calculating the similarity values, we can then make predictions for user-item pairs absent from the current dataset. In Collaborative filtering, we need to analyze vast quantities of data regarding the behaviour of the user such as the past actions or inclina-tions and then use them to predict the user's inclinations using the data [2].

A. Types of Collaborative Filtering

Collaborative filtering methods are of basically two types -

- Memory Based - Memory-based methods examine user and related item data to provide recommendations and can be further subdivided into other methods includ-ing Content-Based methods along with hybrid methods. Content-based methods suggest objects based on the history of the user in question while ignoring the history of other users [3]. Hybrid methods merge collaborative filtering and content based method for yielding recom-mendations with improved accuracy. Memory-based Col-laborative Filtering has three basic phases which include correlation evaluation, production of nearest neighbor-hoods followed by score forecast [4].
- Model Based - Here, a technical model is proposed for ranking behaviour of users by estimating the parameters of the model from the available ranking data rather than employing the raw ranking data to make predictions.

B. Advantages and Disadvantages of Collaborative Filtering

Collaborative filtering does not depend on computer an-alyzable content and hence it can accurately recommend complicated items including music, movies, products without the actual need to "understand" the item in question [5]. Collaborative filtering algorithms can scale remarkably well and additional data can be added for analysis. The drawback of collaborative filtering is a performance decrease as the data becomes sparser which is quite common with large datasets and therefore leads to reduced scalability.

ISSN 2348 – 8034
Impact Factor- 5.070

III. METHODOLOGY FOR MUSIC RECOMMENDER SYSTEM

To analyze the techniques of implementation of recom-mender systems using collaborative filtering, we understood the working of a song recommendation system [6]. For this, we employed the million song dataset available on Kaggle. The Kaggle Million Song Dataset is open; with content analysis and meta-data available. It has about 48 million (userid, songid, play count) triplets accumulated using past listening usage of over a million users including the metadata (280 GB) of millions of songs. It was published by The Laboratory for the Recognition and Organization of Speech and Audio at Columbia University. This data was employed to determine correlations amongst users and songs while learning from past listening records of the users to suggest the users with new and relevant songs [7].

Due to the large amount of data present in the dataset, we limited ourselves to work with the validation set for our analysis and employed the metadata of about 5,000 songs, which amount to about 1.5 GB. Our emphasis was on classi-fiers that could matter the most to characterise songs such as the year; genre, popularity, artist. To aid computations, we changed usernames along with song names which were originally strings to integers [8].

A. Similarity Measure

To check the similarity amongst songs and users, we re-quired similarity criteria. Commonly used similarity measures include Cosine similarity, but it has the drawback that it considers every user with equal weight and this cannot be applied to a real-world scenario. The users must be assigned a lower weight if he/she has a wide range of interests and does not stick to a particular genre, while on the other hand, a user is assigned a higher weight if has very specific taste in music or specific genres.

Let I(r) denote a set of entities rated by a user 'r' and U(a) denote a set of users who listened to a song 'a'.

The similarity between users 'r' and 's' is determined as follows-

$$W_{rs} = \frac{I(r) \cap I(s)}{I(r)^{\beta} I(s)^{1-\beta}}, \ \beta\,(0, 1)$$

Furthermore, the conditional probability of similarity be-tween items 'a' and 'b' shall be

$$X_{ab} = \frac{U(a) \cap U(b)}{U(a)^{\beta} U(b)^{1-\beta}}, \ \beta\,(0, 1)$$

B. Scoring Function

For creating the scoring strategy, a weighted score model was proposed for consolidating the data for the various users and items [9].

For a user-centric recommender system, the score for any particular item 'a' is proportional to the similarity measure amongst the user 'r' (for whom the recommendation is being prepared) and other users 's' who have listened to item 'a'. It is computed as -

$$k^{U}_{ra} = \qquad f(w_{rs})$$

For an item-centric recommender system, the score depends on the similarity between item 'a' and items listened by user 'r'. It is computed as –

$$k^{I}_{ra} = \qquad f(w_{ab})$$

ISSN 2348 – 8034
Impact Factor- 5.070

The locality of scoring function indicates those entities which have high similarities. To calculate locality, an exponential function is employed [10]. The function

$$f(x) = x^q ,\ q\ N$$

is used to increase the similarity for those items have little similarity while decreasing the similarity measure of non-similar items even further.

C. Evaluatation Metric

Stochastic consolidation was employed to consolidate the user-centric and item-centric lists. This was achieved through the random choice of one using the probability distribution followed by providing recommendations from high scoring objects from the probability distribution [11]. In certain cases in which the past listening experience for a user is not enough to engage user-centric recommender method, the item-centric algorithm can produce valuable results when the number of songs is much lesser as compared to the users [12].

Mean Average Precision - (mAP)was employed for the eval-uation metric because the need for increased precision is more desirable as compared to recall as false positives may lead to faulty recommendations. Mean Average Precision provides the mean for the proportion of accurate recommendations using high weights for top recommendations. The three phases for calculating Mean Average Precision include -

In the beginning, the precision for each 'i' is calculated providing us with a proportion for accurate recommendations of the top 'i' amongst predicted rankings [13].

$$Pi(r, \alpha) = \frac{1}{\alpha} \sum_{j=1}^{i} M(r, \alpha(i))$$

Next, average precision for every user at each k is calcu-lated.

$$AP(r, \alpha) = \frac{1}{nr} \sum_{i=1}^{t} P_i(r, \alpha).M(r, \alpha(i))$$

Lastly, the average for all users is computed.

$$mAP = \frac{1}{n} \sum_{r=1}^{n} AP(r, \alpha_r)$$

where n is the user count.

IV. TESTING AND RESULTS

Based on the similarity measure, scoring function and evaluation metric to be employed as outlined in the previous section, we ran simulation based experiments and recorded the results in real time. The most accurate results for stochastic aggregation of item-centric technique using values of q and α as 3.0 and 0.150 respectively, while user-centric technique using values 5.0 and 0.300 for q and α respectively [14].

ISSN 2348 – 8034
Impact Factor- 5.070

V. CONCLUSION AND FUTURE WORK

Through this paper, with an aim to enhance accuracy, quality along with the scalability of a recommender system, an approach using collaborative filtering was proposed. We demonstrated the advantages of using collaborative filtering on the Million Song Dataset for providing the user with valuable recommendations which meet the goals that we outlined for an efficient recommender system. We achieved our aforementioned objective while proving the strength of our proposed model.

Our proposal has a lot of scope in other application not only limited to music recommendation, but also for other applications such as movie recommendation [15], product recommendation [16], online dating [17]. Companies such as Pandora, last.fm and other online music services are using such recommendation systems and hence a lot of research is being conducted to improve the recommendation so that customers can have an improved listening experience [18].

However, there can be improvements which can serve as future work. Our technique lacks personalization, and this needs to be worked upon [19]. Furthermore, there are a large number of songs which have very few users listening to them and because of our algorithm, they do not get recommended. This leads to a skewing of the recommender system in which popular songs are recommended even more and less popular songs are not recommended. This can be adjusted by explicitly disregarding song recommendations which are very popular and thus helping users discover new songs [20]

REFERENCES

1. *A Collaborative Filtering Recommendation Engine in a Distributed Environment, Poonam Ghuli, Atanu Ghosh, Dr. Rajashree Shettar*
2. *A Review Paper on Machine Learning Based Recommendation System Bhumika Bhatt, Prof. Premal J Patel, Prof. Hetal Gaudani*
3. *Collaborative Filtering for Music Recommender System, Elena Shakirova*
4. *Recommender System, Shefali Gar,Fangyan SUN*
5. *Z. Shi, Y., Larson, M., Hanjalic, A. A (2014). Collaborative filtering beyond the user-item matrix: A survey y of the state of the art and future challenges. ACM Computing Surveys (CSUR), 47(1), 2014, pp 3-10.*
6. *Seth Sorensen, Accuracy of Similarity Measures in Recommender Systems, Ph.D. Thesis, Department of Computer Science, University of Minnesota, Morris, 2011.*
7. *Xiaoyuan Su and Taghi M. Khoshgofftaar," A Survey of Collaborative Filtering Techniques" Hindawi Publishing Corporation Advances in Artificial Intelligence Volume 2009 , Article ID 421425, 19 pages doi:10.1155/2009/421425*
8. *Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl" Item Based Collaborative Filtering Recommendation WWW10, May 15, 2001, Hong Kong.ACM.*
9. *McFee, B., Bertin Mahieux,T., Ellis, D. P., Lanckriet, G. R. (2012, April). The million song dataset challenge. In Proceedings of the 21st international conference companion on World Wide Web (pp. 909916).ACM.*
10. *Aiolli, F. (2012). A preliminary study on a recommender system for the million songs dataset chal- lenge. PREFERENCE LEARNING: PROBLEMS AND APPLICATIONS IN AI*
11. *Koren, Yehuda. Recommender system utilizing collaborative filtering combining explicit and implicit feedback with both neighborhood and latent factor models.*
12. *T. Bertin et al., The Million Song Dataset, Proc. of the 12th International Society for Music Information Retrieval Conference, 2011*
13. *P. Resnick and H. R. Varian: Recommender Systems, Communications of the ACM, vo1.40, pp.56-58, 1997 Therapeutics, vol. 1(4), pp. 232, (2005).*
14. *Deshpande, M., Karypis, G.: Item-based top-N recommendation algo-rithms. ACM Transaction on Information Systems 22(1), 143 177 (2004)*
15. *Goldberg, D., Nichols, D., Oki, B. M., and Terry, D. (1992). Using Col-laborative Filtering to Weave an Information Tapestry. Communications of the ACM. December.*

ISSN 2348 – 8034
Impact Factor- 5.070

16. Birtolo, C., Ronca, D., Armenise, R., Ascione, M. (2011), Personalized suggestions by means of collaborative filtering: A comparison of two different model-based techniques, In NaBIC, IEEE (pp. 444450).
17. Recommendation systems: Principles, methods and evaluation F.O.Isinkaye,Y.O.Folajimi,B.A.Ojokoh
18. J.A. Konstan, J. Riedl Recommender systems: from algorithms to user experience.
19. B. Schafer, D. Frankowski, J. Herlocker, S. Sen Collaborative filtering recommender systems.
20. D. Jannach, M. Zanker, A. Felfernig, G. Friedrich Recommender systems an introduction.