Review Paper on optimization of CF algorithms, for Book Recommendation

[1]Reet Garg, [2]Swanjali Singh, [3]Priyanshu Bhardwaj, [4]Aman Kumar, [5]Rahul Chaturvedi

[1]Student, [2]Student, [3]Student, [4]Student, [5]Assistant Professor
[1]Computer Science and Engineering,
[1]Krishna Engineering College, Ghaziabad, India

Abstract: Nowadays, the internet has become a very crucial part of our life. People are depending more and more on internet for the things they can buy online ass it saves their time and energy. In our current generation time is the only big thing which is lacking and it is the most precious proved thing in today's scenario. As the countries are being developed people are becoming more and more digital and getting uplifted by today's technology the traditional framework is being proven inconvenient as the people want to save their time and also expect technology to recommend and introduce the best things to them, the Recommender system is progressively made massive strides for public benefit. The Retrospection system is a substitute of customary recommendation system and thus has made an upheaval in the industries. It has capability to collect data, process and compute information which provide better choices to clients. Analysis of the experimental results of the piled data sets suggests that the many algorithms are stationed on the principle of retrospection system which grant satisfactory services to the readers. This paper trace about recommender system with advancements of multiple machine-learning based algorithms.

Index Terms - Retrospection system, Collaborative filtering technique, KNN approach, Memory-based system, Model-based system, Item-based collaborative approach, User-based collaborative approach, Gradient Descent, Rating-Matrix method.

I. INTRODUCTION

Collaborative filtering algorithm is most widely used customized retrospection technology in business recommendation systems. Retrospection system has become a part of people's routine life where people entrust on technologies for better deciding their choices and wants. The Collaborative Filtering algorithm intake data as input from reader's foregoing activities and traits (for instance, previous purchase history or order/star-ratings of the books) with the help of algorithm, readers construct artificial vision which reassures. Prior to, various methodologies has being developed and it keeps on revolutionizing. However, research still keeps on going due to the often use in many applications, which personalizes retrospection. To overcome, mass analyst passed-down numerous algorithms as- K-Nearest Neighbour (KNN) Algorithm, Alternating Least Squares method (ALSM), Singular Value decomposition (SVD) and many more. New approaches are being explored like the matrix factorization and further types of collaborative filtering (CF). Collaborative filtering technique is basically segregated into memory-based method of CF and model-based methods of CF.

Memory-based system influences, similar users based on similarity of ratings thus customer-book rating matrix is only one that matters as a result, rated output is quite revised on the basis of ratings. Whereas **Model-based** system influences, foregoing activities and traits act as input data which helps to develop new models on the basis of ratings thus retrospection system is adapted, skilled, proficient, competent and dynamic method.

II. RELATED WORK

Book Recommendation System using indigenous approaches. Alternating least square methods (ALS), K-Nearest Neighbour (KNN), as weighting approach, memory-based CF and many more. These approaches depend upon pre-exist data and facts based on customer's activities and history about the ratings for the books which are generated by the customer's feedback and response. These models and algorithms profusely use customer's review and feedback for assessment. Despite being uncertain and still always need a bit of improvement on which the work is still going to make these better and make the living and judgements of the people a bit less to bother.

2.1 ALTERNATING LEAST SQUARE METHOD (ALS) HANDLING

Till up-to date, multiple frameworks and plans have been recognized, developed, designed, configured for book retrospection system by the ALS algorithm. In this algorithm two squares are defined; one of them is known and the other one is not known. The comparison shows better results with the methods.

2.2 COLABORATIVE FILTERING(CF) USING K- NEAREST NEIGHBOUR (KNN) HANDLING

In this design and approach, the enactment of retrospection scheme based on order of the books given by the readers and similar relation is establish between the two readers or users which is then calculated and reviewed after that, this scheme will recommend books as per the functioning.

Later, resultant information (data) set is classified in two division, first division is Graded sample set and second division is Ungraded sample set by using K-Nearest Neighbours algorithm. In such a way, the genre which is asked in the searches and according to that by processing the information the item (book) is recommended to occult readers.

In this scheme, MYSQL database was taken in practice. All the information entered or provided by user's are free from danger. This information stored in user's database. The allotment set in motion using login unit with unique id (username or email) and password of user.

The **fig 1 illustrate** the functioning and effectiveness of collaborative filtering technique with the help of k-nearest neighbour.
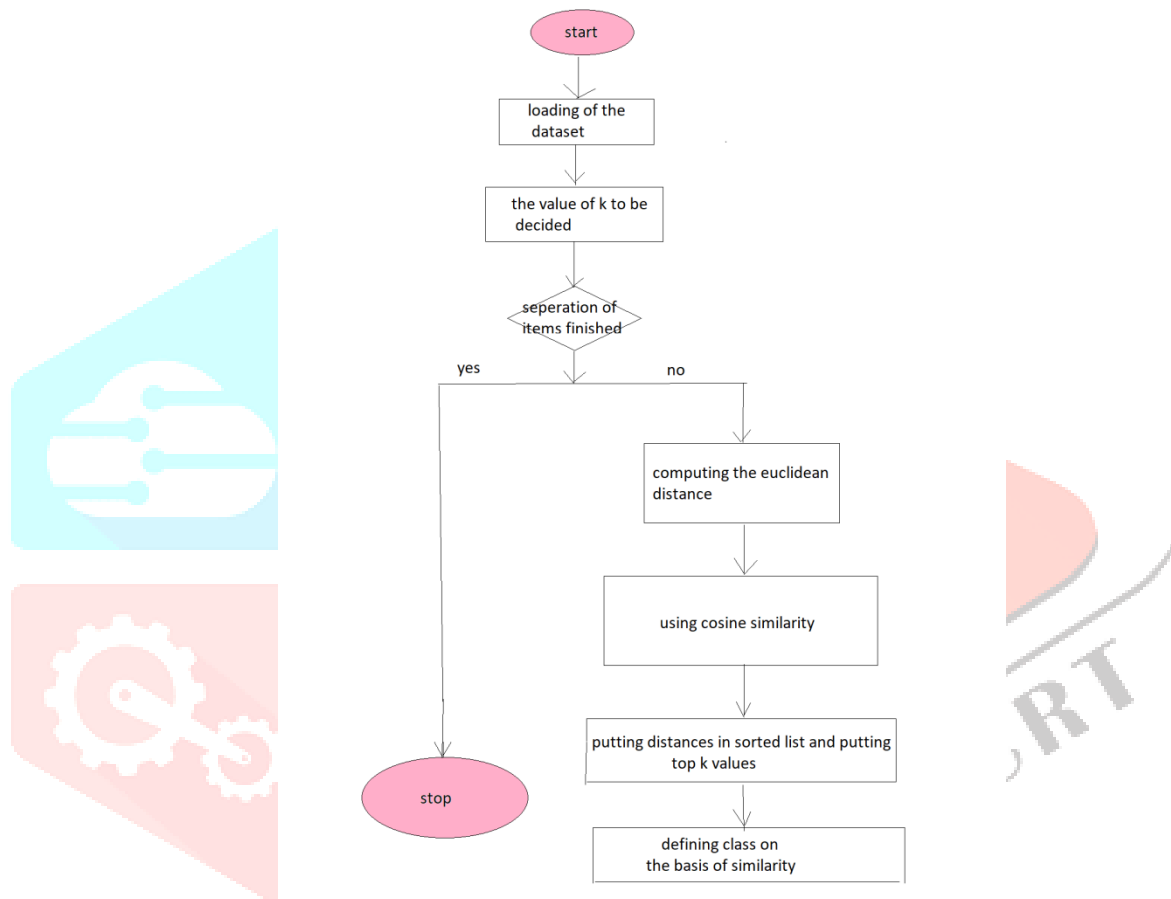

Fig 1 Flowchart of CF with KNN

2.3 MATRIX FACTORIZATION (MF) HANDLING

By selecting various parameters from the algorithm Denise Chen [1] proposed a useful way to generate latent features, this method is known Matrix Factorization. By multiplying two distinct quantities, latent feature is generated.

Collaborative Filtering with the help of matrix factorization, provide a mean which enumerate and establish the relationship between the user's entities with their items

One of the best procedures to calculate minimum value cost function RMSE is gradient descent. This approach is retrospect approach by which minimum value cost function can be minimised.

Gradient Descent is a function to find the local maxima in a problem it is one of the approaches used in artificial intelligence. It is used to reduce the cost function of a problem so the solution to the problem can be found easily.

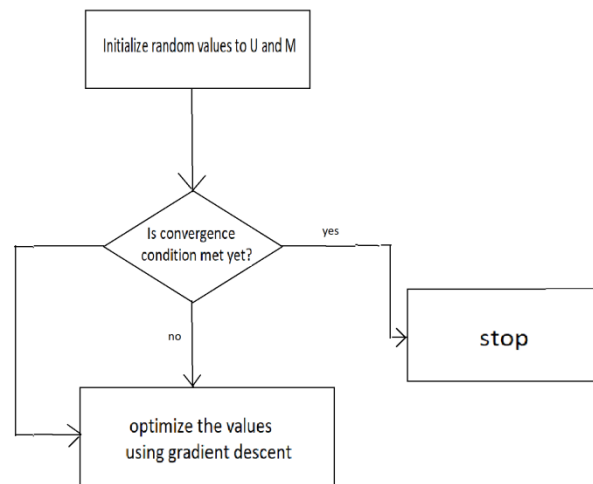The **fig 2,** illustrate the functioning and effectiveness of Gradient Descent.

Fig 2 Gradient Descent Flowchart

This system is well-organised, systematic and productive to recommend books than the memory based and content-based recommendation systems as there are some limitations in this system. In content-based CF sometimes we do not have detailed enough dataset of items so that this method can work well. There is a lack of dataset which is inappropriate at times and do not produce the effective results.

2.4 WEIGHTING SCHEME FOR COLLABORATIVE FILTERING.

Jie Zhou, Quanquan Gu, Chris Ding [3] proffered a matrix factorization, which has non-negative weight, to standardize collaborative filtering approach. In this scheme, they composed two graphs one for the items and another for the users. From the experiment they concluded that there are not many uses of some of the collaborative filtering algorithm and only few are used in the world of developing good recommendation systems.

Analysis on distinct sets composed those populous methods outrun the planned collaborative filtering approach.

2.5 USER-BASED COLLABORATIVE FILTERING HANDLING

C. Kadie, D. Heckerman and J. S. Breese [4] had done factual analysis on the planned collaborative filtering approach. Memory-based scheme is well known for their work of predicting new ratings by giving the average ratings amidst identical users. As a result of this, extensive outlay(ratings) from identical users is unhandy in this approach, which results in insufficiency coming in the data. Therefore, recommendation technique become below par. This framework thereby treats singleton on the basis of ratings on items that come from the users and the customers. The final rating is brought to the estimation by merging foresight from three specific derivations: firstly, foresight on distinct item ratings but made by the Xerox user, secondly, foresight on the ratings of the dupe item but this time users are peculiar and, thirdly, ratings foresight on data from distinct users but homogenous items are taken into consideration. This scheme on completion and formation act as powerhouse due to inefficiency in data, distant ratings used in numerous ways, supplementary ratings acquired from homogenous users purchasing same items are occupied as a framework to get the foresight done in an easier fashion. Experiments demonstrate that the proposed and the put forward approaches are forcibly powerful which determine data inefficiency and are thereby proved to provide accurate recommendation.

Shang, Ming Sheng and Zhi-Dan Zhao [5] personate collaborative filtering approach on cloud computing platform, which designated as Hadoop which was used to rectify the scalability issue in collaborative filtering algorithm. In the first partition, it is necessary that users will acquire output more accurately. The segregation takes place in two halves. In first half, a smart frame-up is done by surveyor which act as fitter image in the starting and in the second half, tasks are taken in consideration evenly, a s a result entire system(processors) will finish their tasks in stipulated pace and remit linear speed-up.

2.6 ITEM-BASED COLLABORATIVE FILTERING

KaivanShah[6] stated in his paper that this approach of item-based collaborative filtering that calculates similarity among the items. In this approach, the main focus is on what items the target user gets the most out of all the available items. Item-based CF is other variety of retrospection method which checks on the similar items based on the item's, users have already liked or gave a better review or feedback about it. The underlying and essential act to measure the relationship among user's items. In this scheme, the Private Pearson approach and the Private Cosine approach are expected which can expertly measures Pearson approach and Cosine approach among user's item in considering the privacy of all the users.

M.Deshpande and G.Karypis [7] experimentally concluded that utmost recommendation approaches which take itemset to compute retrospection foundation than those provided by user-based collaborative filtering algorithms.

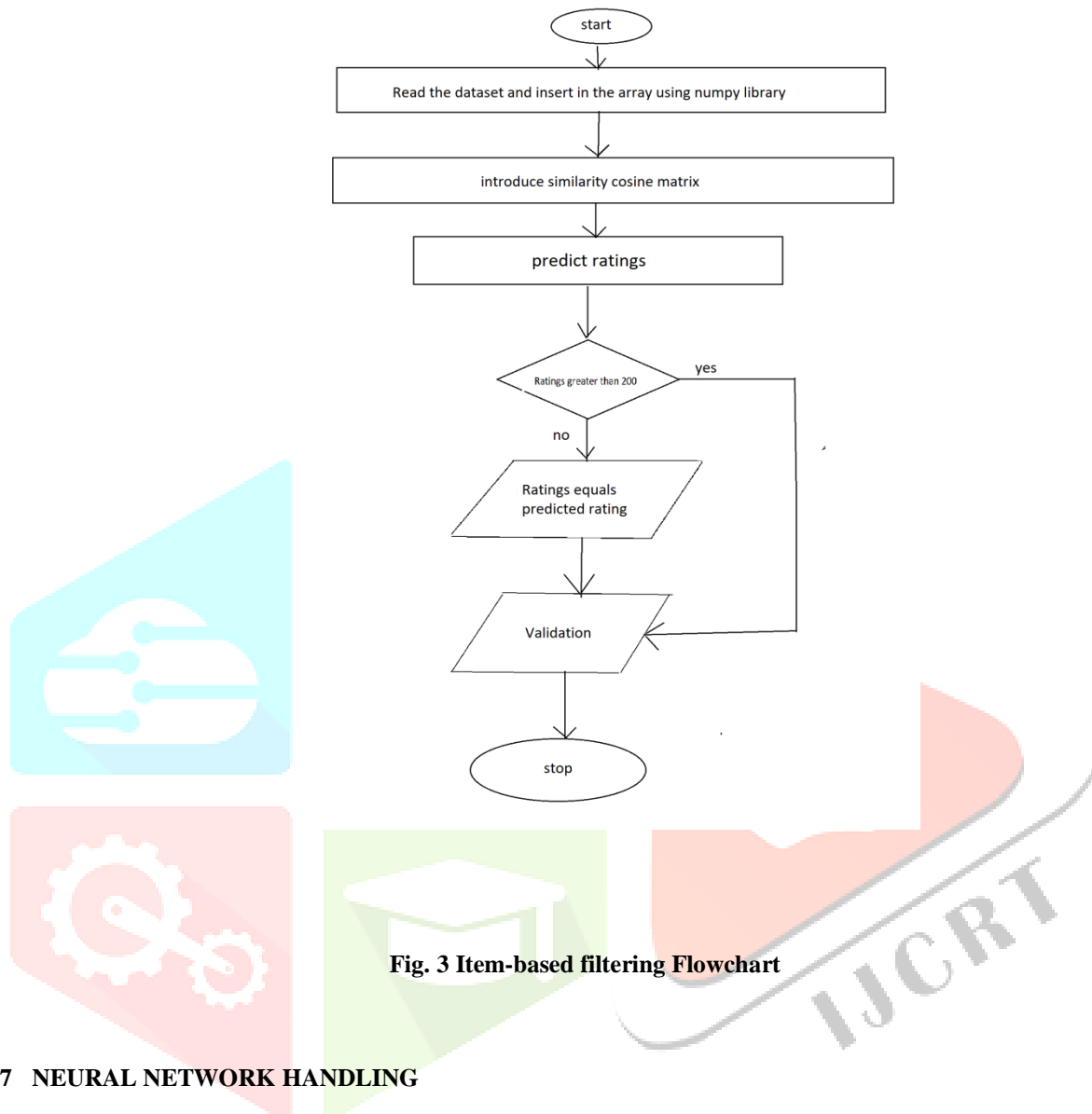The fig **3**, illustrate the functioning and effectiveness of item-based collaborative filtering.

Fig. 3 Item-based filtering Flowchart

2.7   NEURAL NETWORK HANDLING

Meysam Shamshiri, Goh Ong Sing and Yogan Jaya Malay Kumar [8] proposed a better recommendation algorithm strategy by accumulating approaches to handle neural networks. As a result, their proffered system was stored by them in two different methodologies: Precision and error which in turn were compared with other algorithms like MLP and Bayesian networks They have used RMSE and RSE methods. The final result comes to be that their method outperforms other algorithms.

2.8   SINGLE VALUED DECOMPOSITION HANDLING

S.K. Raghuwanshi, R.K. Pateriya [9] stated a better recommendation algorithm in his paper that accumulate matrix-factorisation with gradient descent and considered an advance version of SVD. It plays an active role in convergence of skimpy datasets whereby other algorithms are hardly sub-merge and by a fluke slow in convergence. It break-down skimpy dataset of dimension(N) into a new dataset of dimension(K)such that K<N.

The K-Dimension dataset consists of only considerable feature, spare domains exclude from memory of k dataset. It has a well-defined structure in the form of 2D-Matrix where row represents users and column represents items. The data in matrix will be ratings that users direct to corresponding items.

The **fig 4**, illustrate the functioning and effectiveness of single valued decomposition.
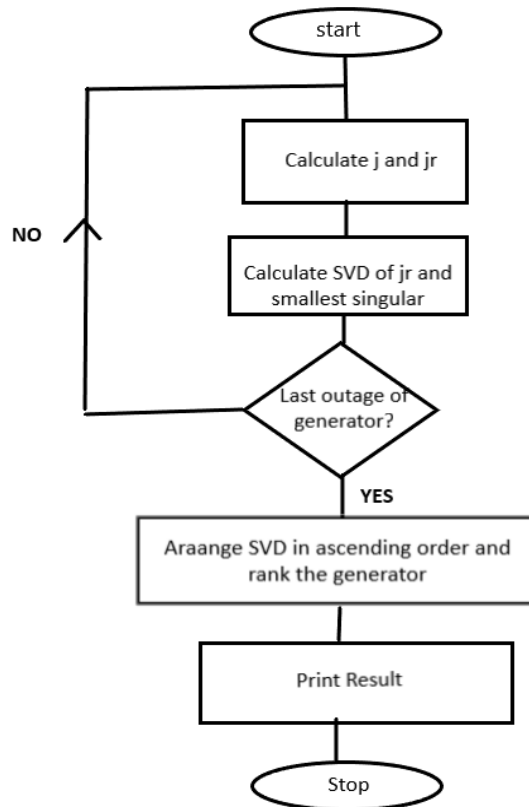

Fig. 4 single valued decomposition Flowchart

III. COMPARISON WITH OTHER ALGORITHMS

Victor Caneiro, Fidel Cacheda, Vreixo Fermoso and Diego Fernandes [10] explore information into halves of dataset, first half contains training dataset and the second half contains an evaluation dataset. In this scheme, first set is bend toward training algorithm and with such information collected from the algorithm it was computed and the recommendation done after-forth as a result original dataset with evaluated dataset present in the system. For retaining comparison between model-based algorithm, memory-based algorithm and normal-predictor conclude by ending statement that's Model-based is remarkably well-optimized algorithm for the recommendation system among all. The memory-based approach works on user's items.

The Proficient significance and accurateness of matrix density offers merits in matrices and sparse data and not for limited data and training thus model-based algorithms are prime solution in recommendation system

The **fig 5**, illustrate the tabular Comparison among Algorithms.

ALGORITHM CATEGORY	USER AVERAGE, ITEM AVERAGE	USER BASED, ITEM BASED	REGRESSION SVD	OTHERS
Size	Very Low	Low	High	Fair
Density	Very Low	High	Very High	Fair
Sparse	Poor	Very Poor	Very Good	Fair
Accuracy	Poor	Fair	Good	Good
Testing	Very fast	Fair	Very Good	Fair
Training	Good	Very slow	Very Good	Fair
Memory consumption	Low	High	High	High
Adjustable parameters	No	Few	Many	Few
Overall merits	Computation takes little time	Do not need to train	Perform best with sparse data	Perform well with asymmetric data
Overall Demerits	Accuracy is very poor	Testing takes very long time	Many parameters should be adjusted	Computation takes too long time

Fig. 5 Tabular comparison among Algorithms

IV. FUTURESCOPE WITH CONCLUSION

The review paper contains summarised and compiled analysis of desirable approaches, methods and technologies that can be opt to design, create, and formulate retrospection system. Different approaches are KNN model with CF, User-based filtering, Item-based filtering, Alternating Least Square Methods, and for measuring the performance Matrix Factorization method (MF) is used. The overall performance ratio of MF is better whether there is large dataset (>2,000) or mass user accessing the system once at a time.

In future Scope, the book recommendation system can be improved by applying and introducing machine learning in it with different libraries of python like matplotlib, math show. Comparison of all the algorithms is done so as to conclude which algorithm is the best to use our book recommendation system to improve the methodology which was used in past.

V. REFERENCES

[1] Denise Chen Data Scientist at Statistics Canada (Ottawa, CA)

[2] Gu, Quanquan, Jie Zhou, and Chris Ding. "Collaborative Filtering: Weighted Nonnegative Matrix Factorization Incorporating User and Item Graphs", Proceedings of the 2010 SIAM International Conference on Data Mining, 2010.

[3] J. S. Breese, D. Heckerman, and C. Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In Proc. of UAI, 1998.

[4] Zhao, Zhi-Dan & Shang, Ming Sheng. (2010). User-Based Collaborative-Filtering Recommendation Algorithms on Hadoop. 3rd International Conference on Knowledge Discovery and Data Mining, WKDD 2010.

[5] Kaivan Shah International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056 Volume: 06 Issue: 05 | May 2019.

[6] M.Deshpande and G. Karypis. Item-based top-n recommendation algorithms. ACM Trans. Inf. Syst., 22(1):143-177, 2004.

[7] Meysam Shamshiri, Goh Ong Sing2 and Yogan Jaya Kumar. "A Recommender System Based on Group Method of Data Handling Neural Network", International Journal of Computer Information Systems and Industrial Management Applications. ISSN 2150-7988 Volume 11 (2019) pp. 028-038.

[8] Willmott, C.J., Matsuura, K.: Advantages of the mean [1]absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. Climate Res. **30**(1), 79–82 (2005)

[9] S.K. Raghuwanshi, R.K. Pateriya/Journal of King Saud University – Computer and Information Sciences 33 (2021) 447–452.

[10] Fidel Cacheda, Victor Canneiro, Diego Fernandes and Vreixo Fermossa. University of Coruna ACM Transactions on the Web, Vol. 5, No. 1, Article 2, Publication date: February 2011.