IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Vol.7, Special Issue.1, Jan 2019 E-ISSN: 2347-2693

# Designing Distributed Recommender Systems using Map Reduce Paradigm -A Study

**P.S. Chakraborty**

Dept. of Information Technology, University Institute of Technology, The University of Burdwan, Burdwan, India

*Corresponding Author: psc755@gmail.com, Tel.: +91-9432132772

**Available online at: www.ijcseonline.org**

***Abstract*—** Nowadays Recommender Systems play an important role in E-Commerce domain. It helps customers to buy the right product or service by generating recommendations. In this paper, a detailed survey has been made regarding the works proposed by different researchers for designing recommender systems in distributed environment. A general framework for designing user-based and item-based recommender systems using map-reduce paradigm has been provide thereafter.

***Keywords*—**Recommender system, Collaborative filtering, Bigdata, Map-reduce,P2P network.

## I. INTRODUCTION

Introduction E-commerce Recommender systems help customers in finding appropriate product/service(s) from alternative ones available to them. In other words, it provides personalized recommendations which, in turn, make customers convenient in choosing the right product/service(s) from the online shop. Collaborative filtering based recommender systems [1,2] uses the previous purchase history of the customers and finds the most similar customers ( with the customer for which recommendation is to be generated ) then combining the opinion of *N* most similar customers (*N* is called the neighborhood ) for the target item. The paper is organized as follows. Section I provides a brief introduction on Recommender System. Section II discusses concepts of distributed Recommender Systems and then summarizes different distributed Recommender Systems proposed by researchers. Section III discusses frameworks for designing user-based and item-based recommender systems using map-reduce paradigm respectively and finally, the paper is concluded in section IV.

## II. RELATED WORK

In Distributed Recommender Systems, request is generated from any peer for asking recommendation. As the rating of different users for different products are distributed among the nodes of the system, the <target_userid, target_itemid> information is sent to all the nodes for calculating similarity between the target user (the user for whom the recommendation will be generated) and other users. No mention to need, it is the first phase of calculating prediction in collaborative filtering based recommender system. Once the user-user similarity calculation is completed, a second level of interactions among the nodes is made for calculating the predicted rating for the target item (the item for which the recommendation will be generated). Table 1. summarizes the works done by the researchers in this domain.

Table 1. Summary of major research works

| Refer-ence | Type of Network | Key contribution and Application Domain |
|---|---|---|
| [5] | Unstructur-ed P2P Network | Agent technology has been used in a P2P environment. |
| PocketLens project [6] | Structured P2P Network | Five alternative distributed CF framework have been designed and analyzed. |
| [7] | P2P File Sharing System | Similarity ranks between items are calculated by log-based user profiles. Similarity values are stored locally. |
| Vineyard system [8] | Centralized/ P2P Network | Provided a generalized service framework for distributed CF. Load balancing has been addressed by using CHORD protocol [3]. |
| PipeCF [9] | DHT-based P2P Network | Distributed Hash Table is used for storing and searching rating data from different nodes of the network. |
| [11] | P2P Networks | Content-based RS which uses image as well as text information for generating recommendations. |
| [10] | Grid technology | A RS framework using Service Oriented Architecture has been build for travel recommendation. |

Map Reduce programming paradigm is at the core of Apache Hadoop architecture [12]. It is a scalable computing model which consists of two major phases- map and reduce. In the map phase, a subset of data assigned to the node is

 

processed and output is generated as a <key, value> pairs which are then fed into the reduce phase. The reduce phase, then, combines the data coming from different nodes. In paper[13] a hybrid book recommender system has been proposed which uses map-reduce to find features of books. Authors of paper [14] propose system for recommendations using MySQL Cluster and Cassandra. They also consider different caching techniques and study the use of data replication and data distribution technologies.

## III. BUILDING RS USING BIG DATA TECHNOLOGIES

### *A. User-Based Collaborative Filtering using map-reduce*

In the first phase the similarity among users is calculated using Pearson correlation coefficient as given in equation 1.

The general sketch of this phase using map-reduce framework is shown in figure 1. Where map and reduce functions are used twice in a chained manner.

$$s(a,b)=\frac{\sum_{i\in R(a,b)}(r_{a,i}-\bar{r}_a)\times(r_{b,i}-\bar{r}_b)}{\sqrt{\sum_{i\in R(a,b)}(r_{a,i}-\bar{r}_a)^2}\times\sqrt{\sum_{i\in R(a,b)}(r_{b,i}-\bar{r}_b)^2}} \quad (1)$$

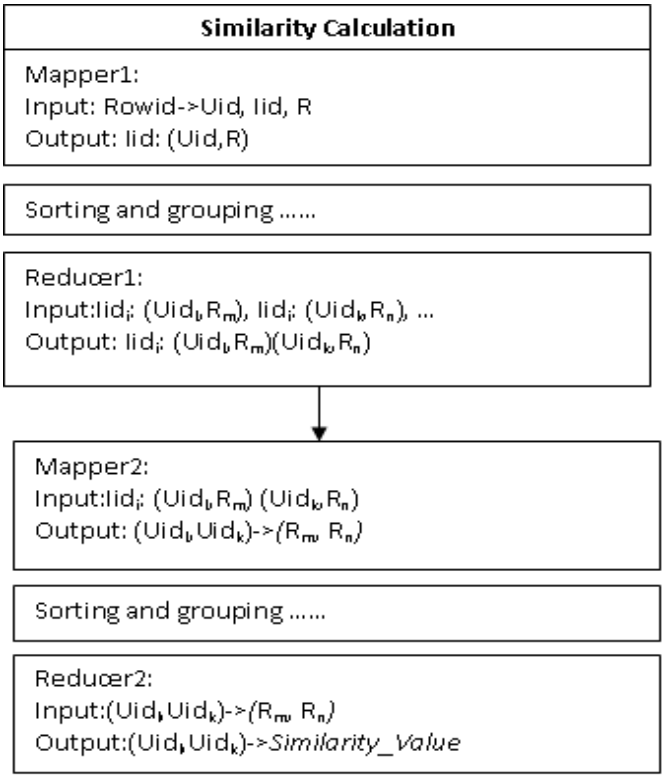


Figure 1.

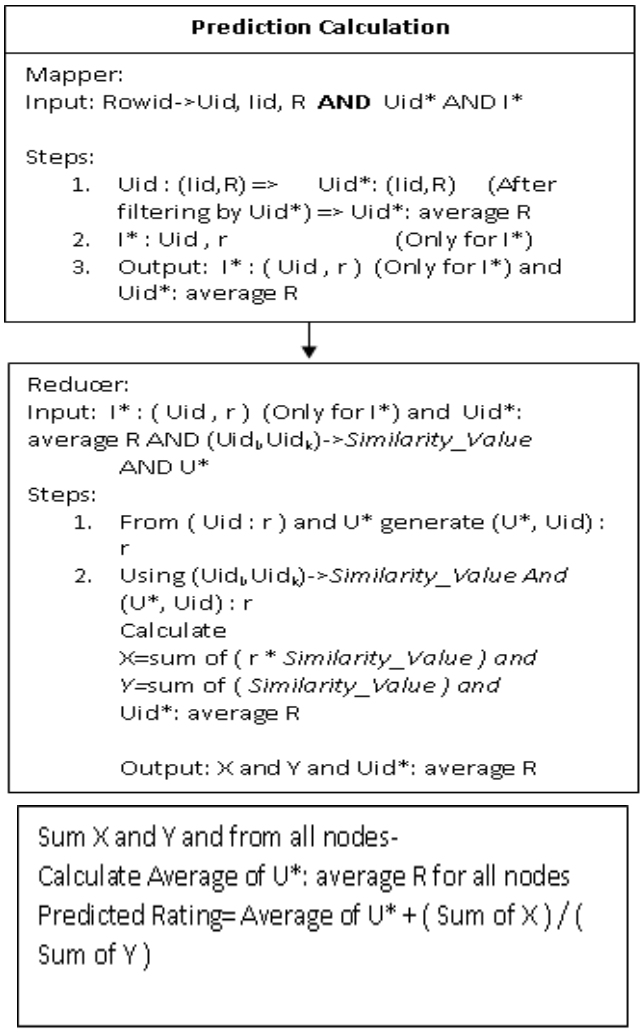


Figure 2.

The second phase of the User-Based Collaborative Filtering is generating the predicted rating for the target user using the similarity matrix derived in the first phase. The predicted rating is calculated using the formula given in equation 2.

$$\tilde{pr}_{a,j}=\bar{P}_a+\left(\sum_{i\_rates\_j,i\in NU}(r_{i,j}-\bar{P}_{N_i})*\omega_{a,i}\right)*\left(\sum_i|\omega_{a,i}|\right)^{-1} \quad (2)$$

where $\bar{P}_a$ , $r_{i,j}$ , $\bar{P}_{N_i}$ ,$N_i$ and $\omega_{a,i}$ denotes the average ratings of active user $U_A$, actual rating of neighbor $U_i$ on product $I_j$, average ratings of neighbor $N_i$ and correlation between the active user, $U_A$ and its $i^{th}$ neighbor, $N_i$ respectively. The general map-reduce framework for prediction calculation is given in figure 2.

 

*B. Item-Based Collaborative Filtering using map-reduce*

In the first phase of Item-Based Collaborative Filtering algorithm the similarity among the items is calculated using adjusted cosine similarity as given in equation 3.

$$s(a,b)=\frac{\sum_{i\in R(a,b)}\left(r_{a,i}-\bar{P}_i\right)\times\left(r_{b,i}-\bar{P}_i\right)}{\sqrt{\sum_{i\in R(a,b)}\left(r_{a,i}-\bar{P}_i\right)^2}\times\sqrt{\sum_{i\in R(a,b)}\left(r_{b,i}-\bar{P}_i\right)^2}} \quad (3)$$

The general sketch of this phase using map-reduce framework is shown in figure 3.where map and reduce functions are used twice in a chained manner.

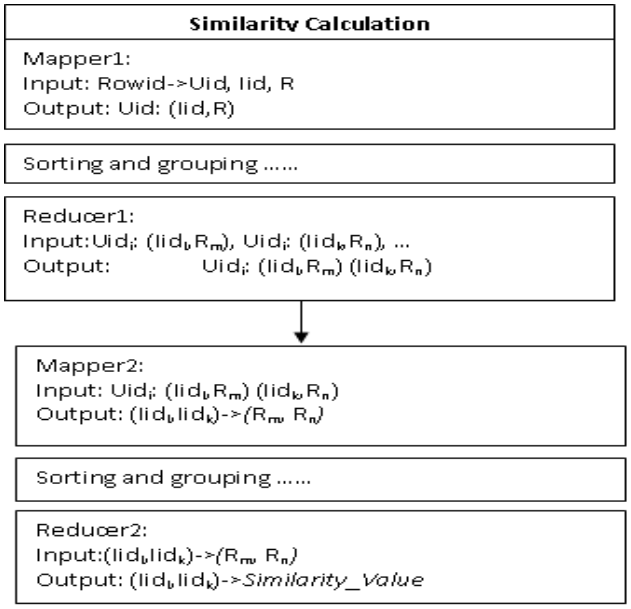


Figure 3.

In prediction calculation (PC) phase, prediction is calculated by equation 4.

$$\hat{r}_{a,j}=\frac{\sum_{i,i\in NI} r_{a,i} * \omega_{i,j}}{\sum_{i,i\in NI}\omega_{i,j}} \quad (4)$$

*NI* is the set of similar items to item $I_j$ and $\omega_{i,j}$ is the degree of similarity between $I_i$ and $I_j$. Figure 4. Shows the map reduce computation for this phase.

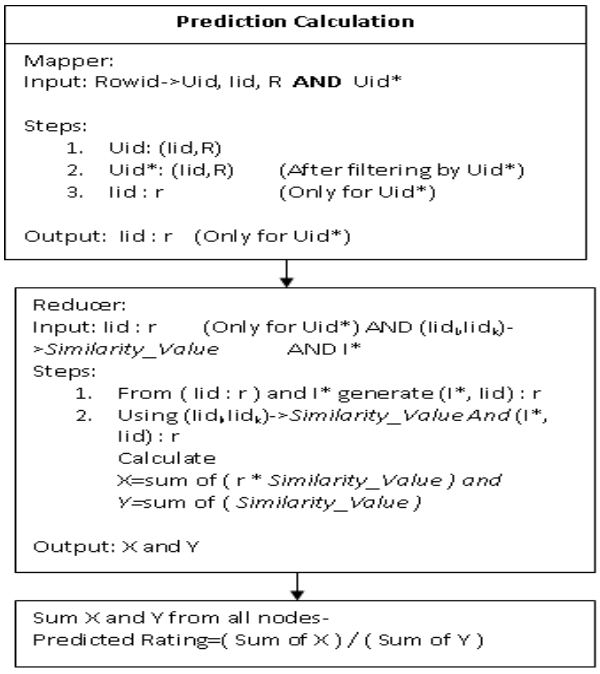


Figure 4.

## IV. CONCLUSION

In this paper the working principle of a distributed recommender system has been mentioned first. Then some important works of researchers in this domain have been summarized. After the introduction of big data technologies, map reduce paradigm has been used in designing distributed recommender systems to increase scalability. In section III, some of the works based on map-reduce paradigm have been highlighted. Finally, a general framework based on map-reduce computing has been shown for designing user-based and item-based collaborative filtering algorithm. In future, designing of a model based collaborative filtering framework using map-reduce paradigm has been proposed.

## REFERENCES

[1] Shardanand, U. and Maes, P. Social information filtering: algorithms for automating "word of mouth". In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'95). ACM Press/Addison-Wesley Publishing Co., New York, NY, 210–217,1995.

[2] J. Herlocker, J. A. Konstan, J. Riedl, "Explaining Collaborative Filtering Recommendations", in Proceedings of ACM Conference on Computer Supported Cooperative Work, Philadelphia, PA, 2000.

[3] Resinck., P., Neophytos, I., Mitesh, S., Peter, B., John, R., 1994. GroupLens: An Open Architecture for Collaborative Filtering of Netnews. Proceedings of the 1994 ACM conference on Computer Supported Cooperative Work, Chapel Hill, North Carolina, United States, p.175-186, 1994.

[4] Stocal, I., et al. (2001). Chord: a scalable peer-to-peer lookup service for Internet applications. In: ACM SIGCOMM, San Diego, CA, USA, pp. 149–160.

 

[5] A. Tveit, "Peer-to-Peer Based Recommendations for Mobile Commerce", in Proceedings of the International Workshop on Mobile Commerce, Rome, Italy, 2001.

[6] B. N. Miller, J. A. Konstan, J. Riedl, "PocketLens: Toward a Personal RecommenderSystem", in ACM Transactions on Information Systems, Vol. 22 (3), 2004.

[7] J. Wang, J. Pouwelse, R. Lagendijk, and M. R. J. Reinders, "Distributed collaborative filtering for peer-to-peer file sharing systems," in Proceedings of the 21st Annual ACM Symposium on Applied Computing (SAC06), 2006.

[8] T. Oka, H. Morikawa, and T. Aoyama, "Vineyard: A collaborative filtering service platform in distributed environment," in SAINT-W '04: Proceedings of the 2004 Symposium on Applications and the Internet- Workshops (SAINT 2004 Workshops). Washington, DC, USA: IEEEComputer Society, 2004, p. 575.

[9] P. Han, B. Xie, F. Yang, and R. Shen, "A scalable P2P recommender system based on distributed collaborative f iltering," Expert Systems With Applications, vol. 27, no. 2, pp. 203–210, 2004.

[10] P. Liu et al., The Knowledge Grid Based Intelligent Electronic Commerce Recommender Systems, IEEE International Conference on Service-Oriented Computing and Applications (SOCA'07).

[11] F. Yuan et al., A Novel Collaborative Filtering Mechanism for Product Recommendation in P2P Network, Third International IEEE Conference on Signal-Image Technologies and Internet-Based System.

[12] Apache hadoop, http://hadoop.apache.org/.

[13] S. Vinodhini, Building Personalised Recommendation System With Big Data and Hadoop Mapreduce, International Journal of Engineering Research & Technology (IJERT), Vol -3, Issue 4, April - 2014.

[14] D. Valcarce et al., A Distributed Recommendation Platform for Big Data, Journal of Universal Computer Science, vol. 21, no. 13, 2015.

[15] J. Soni, A Hybride Product Recommendation Model Using Hadoop Server for Amazon Dataset, Advances in Computational Sciences and Technology, pp. 1691-1705, Volume 10, Number 6, 2017.

[16] Amazon Dataset: http://jmcauley.ucsd.edu/data/amazon/.

## Authors Profile

Partha Sarathi Chakraborty has done his M..E.(in I.T.) degree from West Bengal University of Technology and M.C.A from I.G.N.O.U.. Before that he obtained Master degree in Economics from Kalyani University. For the last twelve years he is serving as Assistant professor of Dept. of I.T. of University Institute of Technology, The University of Burdwan. He obtained his Ph.D degree from the Dept. of Computer Science, The University of Burdwan.

 