

# Recipe Recommender Using Hadoop and Collaborative Filtering

Rohit Mundada<sup>1</sup>, V. S. Malemath<sup>2</sup>

Department of CSE, KLE Dr. M. S. Sheshgiri College of Engg., & Tech., Belagavi<sup>1</sup>

Professor, Department of CSE, KLE Dr. M. S. Sheshgiri College of Engg., & Tech., Belagavi<sup>2</sup>

**Abstract:** There is an extensive class of web applications that involve predicting user responses to options. Such a facility is called a recommendation system. In today's world recommender's system plays very vital role in recommending user required stuffs from the systems like shopping carts, hotel websites, tours and travels website, social media's and other entertainment media websites. We aim to build such a distributed recommendation system on top of the Hadoop framework. Since the data to be processed is usually massive and the computational workload will become heavier after adding and then returning rating parameters, optimizations specific to this modified algorithm and also the Hadoop framework will be the major concern of our project. Collaborative filtering algorithm is one of the most popularly used algorithm for recommendations systems, since Collaborative filtering computational complexity is very high thus it is too much time consuming to use this for large scale data. In this paper we implement Collaborative filtering for the map-reduced recipe data in cloud computing, we use Hadoop to process the large recipe data and then use Map-Reduced data to publish the recipe to users and based on the user interest and rating we recommending the recipe items to the users.

**Keywords:** Collaborative Filtering, Recommender System, Hadoop, Cloud Computing, Map-Reduce.

## I. INTRODUCTION

The measure of data on the planet is expanding much more rapidly than our capacity to process it. Every one of us has known the sentiment being overpowered by the quantity of new books, diary articles, and meeting procedures turning out every year. As of late, the measure of information in our reality has been becoming quickly and examining substantial information sets is purported "Huge information" turns into a key test. At that point what is "Enormous Data"?, Big Data alludes to information sets whose size is past the capacity of current innovation, strategy and hypothesis to catch, oversee and prepare the information inside fair slipped by time [1]. Today, enormous information administration emerges as a test for organizations. The answer for such a test is filtering progressively from giving equipment to provisioning more sensible programming arrangements. Enormous Data additionally conveys new open doors and basic difficulties to industry and scholastics. Presently the time has come to make the advancements that can help us filter through all the accessible data to find what is most profitable to us.

Like most huge information applications, the huge information propensity likewise postures overwhelming effects on server recommender frameworks. With the developing number of options administrations, adequately prescribing administrations those clientsfavoured has turned into a critical exploration issue. Administration recommender frameworks have been appeared as significant devices to help clients manage administrations over-burden and give fitting proposal to them. Case of such viable applications incorporates CDs, Books,

WebPages [3], [4], [5] and different delivers how utilize recommender framework.

Proposal is assuming an inexorably vital part in our life. Exact suggestion empower clients rapidly find attractive things without being overpowered by immaterial data. It is of awesome enthusiasm for sellers to prescribe to their potential client's items coordinating their interests, and ideally transform them into conferred purchasers. No big surprise, in the Netflix challenge, a change of 10% suggestion precision is honoured with 1 Million dollars. In actuality, individuals frequently depend on companions in their informal community for guidance before acquiring an item or devouring an administration. Because of steady and durable social ties, individuals are all the more ready to impart their own insights with their companions, and ordinarily trust suggestion from their companions more than those from outsiders and merchants. The Phenomenally well-known online interpersonal organizations, for example, face book, twitter and YouTube give novel approaches to individuals to impart and fabricate virtual groups.

## II. RELATED WORK

In this area we quickly introduce a portion of the exploration writing identified with community oriented sifting, recommender framework, information mining and personalization. Woven artwork [10] is one of the most punctual usages of community oriented sifting based recommender frameworks. This framework depended on the unequivocal suppositions of individuals from an

affectionate group, for example, an office bunch. Later a few rating-based mechanized recommender frameworks were produced. The GroupLens research system [13], [12] provides a pseudonymous collaborative filtering solution for Usenet news and movies. Ringo [14] and video Recommender [11] are email and online frameworks that produce proposals on music and films. There are different advances been connected to recommender framework which incorporates Bayesian systems, bunching and horting. Bayesian system make a model in light of a preparation set with a choice tree at every hub and edges speaking to client data. The subsequent model is little, quick and basically as precise as closest neighbour techniques [9]. Bunching strategies work by recognizing gatherings of clients who seem to have comparable inclinations. Once the bunches are made forecasts for individual can be made by averaging the assessments of alternate clients in that group. Horting is chart based methods in which hubs are clients, and edged between hubs show level of closeness between two clients [8]. Schafer et al [15] display a nitty gritty scientific categorization and case of recommender frameworks utilized as a part of E-trade and how they can give coordinated personalization and in the meantime can catch client dependability.

### III. CF AND MAP REDUCE

#### A. Collaborative Filtering:

Collaborative Filtering calculation is a great customized proposal calculation; it's broadly utilized as a part of numerous business recommender frameworks. Community filtering calculation is a calculation in view of the accompanying three suppositions thought: People have comparable inclinations and interests. Their inclinations and interests are steady. We can anticipate their decision as indicated by their past inclinations. On account of the above suppositions, the cooperative sifting calculation depends on the correlation of one client's conduct with other client's conduct, to discover his closest neighbour's, and as per his neighbour's advantages or inclinations to foresee his interests or inclinations. The initial step of community oriented separating calculation is to acquire the client's history profile, which can be spoken to as an appraisals framework with every passage the rate of a client given to a thing.

An evaluations network comprises of a table where every line speaks to a client, every section speaks to a particular motion picture, and the number at the convergence of a line and a segment speaks to the client's appraising quality. The nonappearance of a rating score at this convergence shows that client has not yet appraised the thing. Inferable from the presence issue of scanty scoring, we utilize the rundown to supplant the framework. Second step is to ascertain the likeness amongst clients and discover their closest neighbour's. There are numerous closeness measure strategies. The Pearson relationship coefficient is the most broadly utilized and served as a benchmark for CF. Generally, we use the Cosine similarity measure method, its calculating equation as follows:

$$\text{Sim}(x,y) = \frac{\sum_{s \in S_{xy}} r_{x,s} r_{y,s}}{\sqrt{\sum_{s \in S_{xy}} r_{x,s}^2} \sqrt{\sum_{s \in S_{xy}} r_{y,s}^2}} \quad (1)$$

Where  $r_x$  is rating of user  $x$  and  $r_y$  is rating of user  $y$  on item  $s$ ,  $S_{xy}$  indicates the items that user  $x$  and  $y$  co-evaluated.

The Last step is to calculate the item rating. The rating is computed by weighted average of the ratings by the neighbours.

$$r_{x,s} = \bar{r}_x + \frac{\sum_{y \in S_{xy}} (r_{y,s} - \bar{r}_x) \text{sim}(x,y)}{\sum_{y \in S_{xy}} \text{sim}(x,y)} \quad (2)$$

$\bar{r}_x$  is the average rating of user  $x$ .

From the above strides we can find that the count procedure of Collaborative Filtering calculation would devour figuring time and PC assets. At the point when the information set is substantial, the figuring procedure would proceed for a few hours or much more. In this manner, we propose new technique that is actualizing the Collaborative Filtering calculation on Hadoop stage to take care of that issue.

#### B. Map-Reduce Overview:

The Map-Reduce model is a disseminated execution model which is proposed by Google com. Here, we present the Map-Reduce show and depict its working instrument on Hadoop stage. The Map-Reduce models are enlivened by the Lisp programming dialect outline decrease operations. The Map-Reduce model dynamic the count procedure into two stage: map stage and lessen stage. In the Map, composed by the client, takes an arrangement of information key/esteem combines, and creates an arrangement of yield key/esteem sets.

The Map-Reduce library amasses together all halfway values connected with the same transitional key and passes them to the Reduce stage. In the Reduce, the capacity additionally composed by the client, acknowledges a middle of the road key  $I$  and an arrangement of qualities for that key. It consolidates together these qualities to shape a conceivably littler arrangement of qualities. Commonly, none yield esteem or one and only is delivered per Reduce summon. In this passage we quickly break down the booking system of the Hadoop. In the Hadoop stage, at default the info information set size of one mapper is under 64MB (or a record, however the document size is under 64MB), when the document is bigger than 64MB, the stage would part it into various little records which measure under 64MB consequently. For one info record, the Hadoop stage instates a mapper to manage it, the document's line number as the key and the substance of that line as the worth. In the guide arrange, the client characterized process manage the information

key/esteem and pass the middle key/quality to the decrease stage, so the lessen stage would actualize them. At the point when the records square are figured totally, the hadoop stage would slaughter the relating mapper, if the archives are not complete, the stage would picks one document and introduces another mapper to manage it. The Hadoop stage ought to be course the above procedure until the guide errand is finished.

#### IV. MAP-REDUCE FOR RECOMMENDER

In this section we present how to implement Map-Reduce for JSON Recipe data where we find out the recipe items list and count based on Preparation time, Ingredients, Date of Publish and Description.

Since we used Recipe Data Instead of Netflix Movies Data to work on Collaborative filters, In the proposed model we first implement map reduce algorithms of Hadoop for Recipe Data where we will get count and list of recipe items based on Preparation Time needed, Ingredients Contains, And Description given by user. And our approach consists of Four Phases.

##### A. Map Phase:

This is the first phase in this phase the Hadoop environment estimate the algorithms memory and other resource consumption specified in the Datanode the number of mapper it can be initialized. The Hadoop mapper setup function to build Recipe List based on Preparation time and ingredients used.

Here mapper will read the recipe JSON dataset file line by line and in each line it checks for Preparation time of recipe given a and for each occurrence of preparation time the key and values will be created if the preparation time repeats the increment in the value will takes place.

And same procedure will be repeated for Ingredients and description. The Process will continue until all data will be read and completes the task.

##### B. Reduce Phase:

In this stage the hadoop stage would produces a few reducers naturally. The reducer gathers the Recipe subtle elements and sorts them as indicated by the given arrangement.

##### C. Web application:

In this stage we created web application using J2EE architecture where the application takes hadoop map reduced output files as input to database and present the recipes on the web front end to the user whenever user gives rating the ratings will be stored in the database.

##### D. Collaboration Filtering & Recommendations:

This stage will filter the data from database according to the user ratings and interests.

And final recommendations will be made to the user when user searches for any recipe item in the web application.

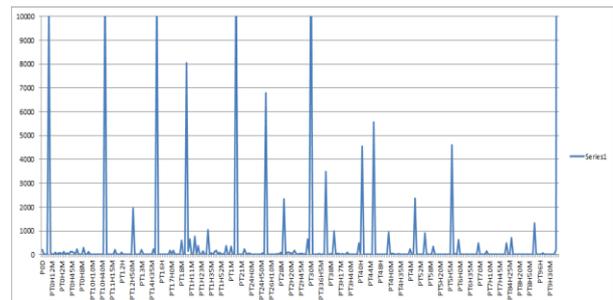


Fig 1. Preparation Time Count.

In the Fig 1 we can see the graph of Map-Reduced Recipe Data based on Preparation Time these data will be given to MySQL database to retrieve them throw web application whenever user search for dish which need to be prepared in given limited time. Using these Map-Reduced data we can present the user about making of dish based on limited time.

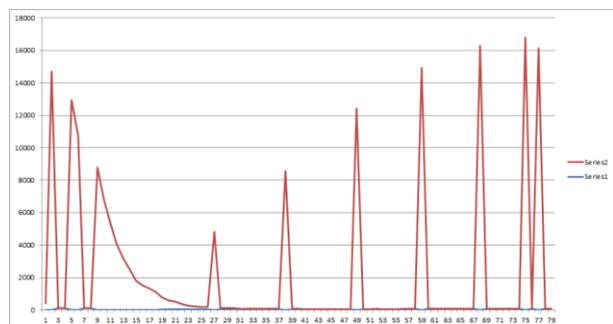


Fig 2. Ingredients Count Dishes.

In the Fig 2 we can graph plotted based on number of ingredients used in making of dish for x-axis specify the number of ingredients in making dish y-axis specifies total number of dishes with respect x-axis values.

#### V. CONCLUSION

In this paper we have proposed a method for user based recommendation system in big data environment, the keywords used are users 'preferences' and user-based collaborative filtering algorithms is adopted to generate appropriate recommendations. Our method aims at presenting a personalized service recommendation list and recommending the most appropriate service(s) to the users.

Moreover, to improve the scalability and efficiency of Recommender system in Big Data environment, we have implemented it on a Map-Reduce framework, in hadoop platform. Finally, the experimental results demonstrate the scalability and accuracy of service recommender system over existing approaches.

In future instead of storing the hadoop sequence file in the HDFS, use Hbase or Sqoop for retrieving the data real time. Also, by combining both the item-based and the user-based collaborative filtering, recommender system can predict accurate recommendations to user.

**REFERENCES**

- [1]. ShunmeiMeng, Wanchun Dou, Xuyun Zhang, and Jinjun Chen. KASR:A keyword-Aware Service Recommendation Method on Map-Reduce for Big Data Applications. In Proceedings of IEEE Transaction on Parallel and Distributed Systems, Vol. 25, No. 12, December 2014, pp. 3221-3231
- [2]. Zhi-Dan Zhao, Ming-Sheng Shang. User-Based Collaborative-Filtering Recommendation Algorithms on Hadoop[C]. In Proceedings of the Third International Conference on Knowledge Discovery and Data Mining, (2010) 478 – 481.
- [3]. G. Linden, B. Smith, and J. York, “Amazon.com Recommendations: Item-to-Item Collaborative Filtering,” IEEE Internet Computing, vol. 7, no. 1, pp. 76-80, Jan. 2003.
- [4]. M. Bjelica, “Towards TV Recommender System Experiments with User Modeling,” IEEE Trans. Consumer Electronics, vol. 56, no. 3, pp. 1763-1769, Aug. 2010.
- [5]. M. Alduan, F. Alvarez, J. Menendez, and O. Baez, “Recommender System for Sport Videos Based on User Audiovisual Consumption,” IEEE Trans. Multimedia, vol. 14, no. 6, pp. 1546-1557, Dec. 2012.
- [6]. J. Manyika et al., “Big Data: The Next Frontier for Innovation, Competition, and Productivity,” 2011.
- [7]. W. Dou, X. Zhang, J. Liu, and J. Chen, “HireSome-II: Towards Privacy-Aware Cross-Cloud Service Composition for Big Data Applications,” IEEE Trans. Parallel and Distributed Systems, 2013.
- [8]. Aggarwal, C. C., Wolf, J. L., Wu, K., and Yu, P. S.(1999). Horting Hatches an Egg: A new Graph-Theoretic Approach to Collaborative Filtering. In Proceedings of the ACM KDD'99 Conference. San Diego, CA. pp. 201-202.
- [9]. Breese, J. S., Heckerman, D., and Kadie, C.(1998). Empirical Analysis of Predictive Algorithms for Collaborative Filtering. In Proceedings of the 14th Conference on Uncertainty in Artificial Intelligence, pp. 43-52.
- [10]. Goldberg, D., Nichols, D., Oki, B. M., and Terry, D.(1992). Using Collaborative Filtering to Weave an Information Tapestry. Communication of the ACM. December
- [11]. Hill, W., Stead, L., Rosenstein, M., and Furnas, G.(1995). Recommending and Evaluating Choices in a Virtual Community of Use. In Proceedings of CHI '95.
- [12]. Konstan, J., Miller, B., Maltz, D., Herlocker, J., Gordon, L., and Riedl, J. (1997). Group Lens: Applying Collaborative Filtering to Usenet News. Communications of the ACM, 40(3), pp. 77-87.
- [13]. Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994). GroupLens: An Open Architecture for Collaborative Filtering of Netnews. In Proceedings of CSCW '94, Chapel Hill, NC.
- [14]. Shardanand , U., and Maes, P. (1995). Social Information Filtering: Algorithms for Automating ‘Word of Mouth’. In Proceedings of CHI '95. Denever, Co.
- [15]. Schafer, J. B., Konstan, J., and Riedl, J. (1999). Recommender Systems in E-Commerce. In Proceedings of ACM E-Commerce 1999 Conference.

Universiyt Dharwad in the year 1993,did his MS in Software Systems from BITS Pilani Rajasthan in 1998 and received his Ph.D in Computer science from Gulbarga University,Gulbarga in 2009. His research interests are Document Image Processing , medical and pattern recognition.He has published more than 75 articles reviewed international journals and conferences.

**BIOGRAPHIES**

**RohitMundada**received B.E. degree in Information Science Engineering from Visvesvaraya technological university of Belgaum in 2014. Currently pursuing M. Techdegree in Visvesvaraya Technological University of Belgaum.His research interest area are in BigData

Analytics and Cloud Computing.



**Dr.Virendra S Malemath** is currently working as Professor in Computer Science & Engg, KLE Dr. M S Sheshgiri college of Engg & Tech, Belagaum.He did his Bachelors of Engg in Electronics & Communication Engg from Karnataka