



Collaborative Recommendation System Based On Hadoop

Harsh Varudkar

Student, Shah & Anchor Kutchhi Engg. College, Mumbai, Maharashtra, INDIA.

Ms. Shivani M Deosthale

Professor, Shah & Anchor Kutchhi Engg. College, Mumbai, Maharashtra, INDIA.

Ms. Jalpa Mehta

Professor, Shah & Anchor Kutchhi Engg. College, Mumbai, Maharashtra, INDIA.

ABSTRACT

In day-to-day life people rely on recommendation made by other people using surveys, guides, spoken words, news and media. Recommendation system implements algorithms for making recommendation. It provides assistance to choose best item from many products as per interest. Collaborative recommendation system is one of the popular method in recommendation system. It uses users' rating to evaluate recommendations. The complexity of collaborative recommendation system for computation is much higher as it has to compute large scale data and trail large scale systems. Hence for solving scalability of recommendation system, we implemented collaborative recommendation system over distributed computation platform, Hadoop.

KEYWORDS : Collaborative filtering, Recommender systems, Hadoop, Map-Reduce, Mahout

INTRODUCTION

21st century is internet age. Websites, online services, digitalization of business tends rapid growth of the information. Users face difficulty to find the relevant information they want. Recommendation system uses algorithms to predict recommendation on basis of users' interest and patterns. Recommendation systems are used in almost all field of IT, e.g. Amazon, LinkedIn, and Facebook. Recommendation system provides users their interest based data which will make easy for users to choose their product or service which make more purchase hence recommendation system not only useful for customer but also at providers end. Recommendation system has mainly 3 types, 1. Collaborative Recommendation system 2. Content based recommendation system 3. Hybrid recommendation system [1]. Each type has its own properties and requirement which define which recommendation suits what kind of scenarios. Collaborative recommendation system is one widely used recommendation system. It is a very popular technique in commercial recommendation system [2]. In Collaborative Recommendation system users past transaction has been analyzed and according to such behavior recommendation has been provided with highest rating score. Quality of Collaborative Recommendation system is depends upon users rating quality and quantity of data, more amount of data provides best predictions. Collaborative recommendation only depends upon ratings of users make it more clear and affordable to implement (Alspector, et al. 1997). Here, we implemented Memory based Collaborative recommendation system. It does not need feature of product or content to give recommendation. Collaborative recommendation system is highly popular many practical application like recommending music, apps, music, social tags, book etc. [3]. Content based recommendation system needs features and ratings to generate recommendation, Content based recommendation bit costly as we need to create features for each product and it much be updated which make it costly [4].

Hybrid Recommendation system used multiple feature from multiple recommendation system. It combines collaborative and content based recommendation to make a recommendation system. Hybrid recommendation system is used in scenarios where requirement are very discrete and doesn't fit into a single recommendation system.

REVIEW OF LITERATURE

Collaborative recommendation system can be implemented by Memory based and Modal based. Memory based collaborative

recommendation system make predication stored in memory. Modal based collaborative recommendation system use collection of ratings to learn the model which is then used to make rating predictions. Collaborative recommendation system uses approach to find similar users of similar taste. For finding similarity we use users' past transaction on basis of those ratings we find similar users by comparing it with community data [5].

In Collaborative recommendation system, we use algorithms for finding similarity between users. The basic approach is that user A and Users B's taste are correlated if both users rate n items similarly. For this we need to gather ratings from users then correlate between pair of users to determine a user's neighborhood and on basis of that recommendation has been generated. There are many similarity algorithms for finding similarity e.g. Pearson correlation, Euclidean correlations, Loglikelihood correlation etc. [6].

Collaborative recommendation system computation is very complex and in case of large scale dataset there will be scalability issue. Hadoop is a distributive environment. There are 2 modules of Hadoop, 1st is MapReduce and 2nd is Hadoop Distributed File System (HDFS). HDFS is a distributed file system that is very reliable as it stored data on different nodes and keep replica to recover in critical scenarios. Mapreduce is computation component of Hadoop. Mapreduce uses parallel computing approach. MapReduce read data line by line by using mapper function and divide it in multiple chunks, on other hand reducer use his chunks for computation over nodes and generate results. Hadoop is an open source project by apache. Hadoop provides scalable computational environment as well as very reliable storage HDFS [7].

Mahout is an open source project of apache. It is a very popular tool in machine learning field. It is useful to create machine learning modal. Mahout can also use to implement recommendation system. Mahout supports many algorithms like, Manhattan distance, Pearson correlation, Euclidean correlations, Loglikelihood correlation, etc.

PROPOSED WORK

We implemented memory based collaborative recommendation system with the help of mahout and Hadoop for better scalability. Implementation steps are as follows.

1. Dataset

Here for user ratings we use Movielens[8] dataset. This dataset is available on Grouplens. This datasets are available in multiple size and it has format userid:itemid:ratings.

2. Finding Similarity

We use datasets to find similarity between users by using similarity algorithm. We are implementing Pearson Correlation Loglikelihood, Euclidean, Tanimoto. We implemented such algorithm by using mahout.

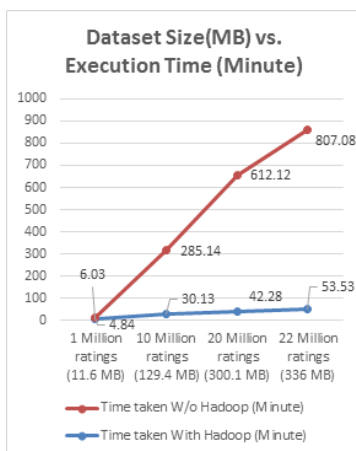
3. Storage

After computation the generated recommendation file is stored in HDFS. For further visualization we can use this recommendation.

We use VMWare Virtual Machine of Ubuntu OS 14.04, Hadoop Single node 0.20. For mahout execution, Apache maven 3.0.5 has been used. We implemented different similarity algorithm to find similarity and here we have shown example of Pearson correlation similarity runtime time on different size of datasets is given in below Table 1.

**TABLE – 1
EXECUTION TIME OF DIFFERENT DATASET**

| Dataset | Time taken With Hadoop (Minute) | Time taken W/o Hadoop (Minute) |
|-------------------------------|---------------------------------|--------------------------------|
| 1 Million ratings (11.6 MB) | 6.03 | 4.84 |
| 10 Million ratings (129.4 MB) | 30.13 | 285.14 |
| 20 Million ratings (300.1 MB) | 42.28 | 612.12 |
| 22 Million ratings (336 MB) | 53.53 | 807.08 |



CONCLUSION & FUTUREWORK

For better recommendation in collaborative filtering we need good quality of community data which use for finding similarity, as much as huge data we have as good recommendation will be generated. Handling huge amount of data is very complex. We implemented Collaborative recommendation system over Apache Hadoop Yarn. We can easily see from experiment that as data size increased Hadoop will provide more scalability.

By using Hadoop and mahout, we can process large dataset and provide recommendation efficiently. But in case of real time scenario we need to use flume or sparks to combine recommendation system with real time data, which will be useful for new users who can directly get recommendation as soon as it gives few ratings.

REFERENCES:

1. Robin Burke, 1. Department of Information Systems and Decision Sciences, California State University, Fullerton, CA, Hybrid Recommender Systems: Survey and Experiments, User Modeling and User-Adapted Interaction, November 2002, Volume 12, Issue 4, pp 331-370, <http://link.springer.com/article/10.1023/A:1021240730564>
2. Harsh Varudkar. 2016. Hadoop based collaborative recommendation system. In Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies (ICTCS '16). ACM, New York,

- NY, USA,, Article 137, 3 pages. DOI: <http://dx.doi.org/10.1145/2905055.2905353>
3. Paritosh Nagarnaik, Prof. A.Thomas, "Survey on Recommendation System Methods", IEEE SPONSORED 2ND INTERNATIONAL CONFERENCE ON ELECTRONICS AND COMMUNICATION SYSTEM (ICECS 2015)
4. Neethu Raj, Suja Rani M S, "An Overview of Content Recommendation Methods "International Journal of Innovative Research in Computer and Communication Engineering
5. B. N. Miller, J. A. Konstan, and J. Riedl, "PocketLens: towards a personal recommender system," ACM Transactions on Information Systems, vol. 22, no. 3, pp. 437-476, 2004. Dr. Senthil Kumar Thangavel, Neetha Susan Thampi, Johnpaul C I, "Performance Analysis of Various Recommendation Algorithms Using Apache Hadoop and Mahout", International Journal of Scientific & Engineering Research, Volume 4, Issue 12, December-2013.
6. Xiaoyuan Su and Taghi M. Khoshgoftaar "A Survey of Collaborative Filtering Techniques", Hindawi Publishing Corporation, Advances in Artificial Intelligence Volume 2009
7. F. Maxwell Harper and Joseph A. Konstan. 2015. The MovieLens Datasets: History and Context. ACM Transactions on Interactive Intelligent Systems (TIIS) 5, 4, Article 19 (December 2015), 19 pages. DOI=<http://dx.doi.org/10.1145/2827872>
8. Poonam Ghuli, Atanu Ghosh, Dr.RajashreeShettar, "A Collaborative Filtering Recommendation Engine in a Distributed Environment", 2014 International Conference on Contemporary Computing and Informatics (IC3I).
9. Xiwei Wang, Erik von der Osten, Xuzi Zhou, Hui Lin, "A Case Study of Recommendation Algorithms", 2011 International Conference on Computational and Information Science
10. Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl "Item Based Collaborative Filtering Recommendation Algorithms" Proceeding WWW '01 Proceedings of the 10th international conference on World Wide Web Pages 285-295 ACM New York, NY, USA ©2001
11. Xingyuan Li, "Collaborative Filtering Recommendation Algorithm Based on Cluster", 2011. Proc of International Conference on Computer Science and Network Technology, 2011, pp.2682-2685.
12. M. O'Mahony, N. Hurley, N. Kushmerick, and G. Silvestre, "Collaborative recommendation: a robustness analysis," ACM Transactions on Internet Technology, vol. 4, no. 4, pp.344-377, 2004.
13. J. L. Herlocker, J. A. Konstan, A. Borchers, and J. Riedl, "An algorithmic framework for performing collaborative filtering," in Proceedings of the Conference on Research and Development in Information Retrieval (SIGIR '99), pp. 230-237, 1999.
14. Carlos E. Seminario David and C. Wilson, "Case Study Evaluation of Mahout as a Recommender Platform", Proc ACM RecSys, 2012, pp.45-50.
15. Ailin Deng, Yangyong Zhu and Bole Shi, "A Collaborative Filtering Recommendation Algorithm Based on Item Rating Prediction", Proc Journal of Software, vol.14, No.9, 2003.
16. Y. Y. Yao, "Measuring retrieval effectiveness based on user preference of documents," Journal of the American Society for Information Science, vol. 46, no. 2, pp. 133-145, 1995.
17. YU Chuan, XU Jieping and DU Xiaoyong, "Recommendation Algorithm combining the User-Based Classified Regression and the Item-Based Filtering", Proc ICEC, 2006, pp.574-578.
18. Murthy, Arun. Apache Hadoop YARN: Moving beyond MapReduce and Batch Processing with Apache Hadoop 2. Upper Saddle River, NJ: Pearson, 2014.