ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

Implementation of Collaborative Filtering Techniques Based On Items

Jaisimha R[1], S Sneha[2], Tanushree P[3], Manu M N[4]

B.E Students, Department of Information Science and Engineering, SJB Institute of Technology, Bengaluru, Karnataka, India[1,2,3]

Professor, Department of Information Science and Engineering, SJB Institute of Technology, Bengaluru, Karnataka, India[4]

ABSTRACT: Collaborative filtering (CF) identifies common interest between users based on their rating and recommend items. This is based on the assumption that the users who liked a product in the past will also like a similar product in future. This helps in solving overload problem. They work. In traditional Collaborative Filtering, user-item matrix is based on single criteria for recommending items. We compare the accuracy of the items that is provided through single-criteria and multi-criteria. Results show that comparison of the techniques used, Multi-Criteria gives accurate recommendation over single-criteria.

KEYWORDS: Collaborative Filtering, Single criteria, Multi criteria, Recommendation system.

I. INTRODUCTION

Big data [1] is an amazingly expansive informational collection that might be investigated computationally to uncover examples, patterns, and affiliations, particularly identifying with human conduct and co-operations.

Big Data is characterized by four measurements spoken to by five V's (Value, Volume, Variety, Velocity, and Veracity) [10]. Volume is spoken to by the measure of content information that we are utilizing for rundown to create proposal. Assortment speaks to various sort of information separated from various sources like online journals, Facebook, twitter and also unique audit, assessment locales.

Analyst can compose their audits, comments, criticism in any configuration like structure, semi-organized, and unstructured that ought to be taken care of by the framework. Speed speaks to the seed of information era on web. Presently nowadays everyone is associated through web also expanding the notoriety of web based business locales has turned into the primary purpose behind expanding the speed of content information era on the web. Veracity speaks to the dependability of the information. Ordinarily survey, feeling, inputs are controlled or supported by various partners of online business.

Under big data, the most common and important part is the Recommender system. Recommender System that helps recommending personalized items to users based on their interest. These systems usually use data mining as the basic process where it can be defined as a method of finding new information from large volume of unprocessed data. These in turn use Clustering techniques. Clustering can be defined as a method of organising objects that are similar to each other into groups and that these objects are not similar to other groups Recommender system is divided into the following:

Collaborative Filtering system: It is the most commonly used technology which is popular and efficient. They identify common interest between users based on their rating and recommend items. This is based on the assumption

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

that the users who liked a product in the past will also like a similar product in future.
Content based system: As the name says, the content of an item is known completely and then it is recommended to the user.
Demographic based system: It categorizes users based on demographic (a particular sector of population) classes. These systems also use people to people correlation but uses different data.
Utility based system: It is based on the usefulness of an object to a user. The problem here is how to provide utility for the user. It helps in factoring non-product attributes.
Knowledge based system: It works on functional knowledge i.e. they know how a particular item is suitable and appropriate for a user.
Hybrid system: This combines a number of recommender system techniques for users. The main advantage is it combines different systems and eliminates weakness.

Collaborative Filtering [2] performs on user's preference on several items to build user-item matrix. It can be classified as memory based and model based. Under memory based, we have user based and item based. Memory based provides better accurate reference but they have scalability problem.

In single criteria, a user rates an item based on a single category. But users may want to rate on multiple categories. Due to this drawback in single criteria, Multi-criteria item based collaborative
filtering system is used. Adomavicius and Kwon [1] introduced in multi criteria to get the correlation between different users. The main contribution of the paper gives us accurate recommendation of items to users by using multi criteria system.
Some of the challenges faced by Collaborative Filtering are:
Data Sparcity: Large datasets are used for recommender system in which user item matrix used by collaborative filtering will be vey large and sparse and faces challenges in recommendation.
Scalability: As the no of items increase, the numbers of users also increase thereby resulting in inaccurate information and it gets expensive.
Cold-Start problem: When new users and new items arise, recommendation cannot be given properly because very little information is available about the user and new items will have no rating.
Low Confidence value: There is no guarantee as to whether precise recommendation is given to the user or not.
Serendipity loss: Extracting information about new items and recommending them to new users is difficult.

The rest of the paper is as follows. In section II, we discuss related work and summarize on collaborative filtering. In section III, we mention about the design. In section IV, we describe our proposed method. Experimental results are in section IV followed by conclusion is section V.

II. RELATED WORK

This section provides various contributions of different authors based on collaborative filtering, recommender system and big data. The authors [4] used various techniques to implement collaborative filtering. They used agglomerative hierarchical clustering which merges clusters going up the hierarchy.Stemming algorithm and mash up services were also used. The authors [5] use hybrid recommender system where they mainly obtain the results based on individual and group trust. They compare individual and group trust and therefore the amount of prediction accuracy depends on the amount of group trust. If the impact reduces the error rate increases. M. OmairShafiq [5] proposed event segmentation in an effective and efficient way. The method was evaluated and tested on large scale log data based on real life case study. Performance evaluation was also performed and effectiveness was checked. Min Seon Lee, Earl Kim et.al [6] improved machine learning by using Hadoop ecosystem. A prediction model was created from the collected data that could recommend courses taken by students going to the next semester. It also provides personalized recommendation to users. In [7], CF types and their main challenges are presented. Using Apache Mahout, Single criteria CF is used to recommend items to users. Logliklihood similarity was used to process huge data. Prathana Rao H

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

M and Satyanarayna S [1] specified on how useful recommendation system are for providing items based on users interest. They also told about how useful it will be if done correctly, it will maximize the value and protect important corporate reputation.

III. DESIGN

A. *Apache Mahout:*

It is a machice learning library containing algorithms for open source Apache Hadoop [11]. In machine learning, it helps in providing scalable performance and to build scalable algorithms. In ApacheHadoop to produce distributed scalable algorithms, Mahouts implemented MapReduce paradigms is used.

B. *Hive:*

In Hadoop Hive [11] is used for SQL queries over petabytes of data in Hadoop. Using Hive tools data, summarization, querying and analysis of big data can be done.

C. *Apache Spark:*

Apache software foundation introduced spark for increasing their speed of Hadoop computational registering programming process [11]. Spark is not a changed variant of Hadoop. But Spark is dependent on Hadoop in such a way that it has its own specific clustering management. Spark in Hadoop can be used in two ways- one is capacity and other is handling. Spark uses Hadoop for capacity reasons. Spark has its own calculation for cluster management.

IV. PROPOSED METHOD

Single criteria needs overall ratings of the user to collect their preferences but in MC a user can rate an item on multiple categories

$$R: Users\ x\ Items \rightarrow R_0 X R_1 \dots X R_c \quad (1)$$

Where R_0 is the overall ratings and R_i is for each criterion, i and c shows the number of criteria.

Adomavicius and Kwon [3]says about similarity and aggregation function based approach. The methodology is used to find the similarities between items and to calculate multi-dimensional distance matrix.

In [1] the performance is evaluated by Euclidean distance.

$$dis(i,j) = \sqrt{\sum_{s=0}^{c} abs(\mathrm{r_i} - \mathrm{r_j})^2} \quad (2)$$

According to [1] multi-criteria ratings gives user preference for different categories in item and overall rating depends on the other user ratings on each category. By this we can find the aggregation of different categories to give overall ratings.

Aggregation is mainly done for giving recommendation to user and to do this recommendation we use linear regression technique where overall rating is estimated using equation [3].

$$r_0 = w_1 r_1 + w_2 r_2 + \cdots + w_c r_c \quad (3)$$

A prediction value is produced for each criteria and linear regression based function is used to estimate overall final rating.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

V. EXPERIMENTAL RESULTS

The data that is given by different users is being pre-processed in big data and it will recommendations that is needed by the users and it will also check the time efficiency of the collaborative filtering system. Here we are using the book data set that is collected from Amazon web services. On the book data each user can rate with respect to 3 criteria i.e. author, type of book, publication. The user can also give a overall rating for the book. In the data set, 5 level rating scale (from 1 to 5) is used. In the experiment we use a two set data rating that is the rating given for most items i.e. top 20 ratings and then another subset is taken i.e. top 10 ratings. The main motivation is to check the accurate

recommendation that is got when items are rated in single criteria and in multi criteria. The following are the data that is being given.

```
vagrant@vagrant:~/data/book$wc -l book.csv
271379 book.csv
vagrant@vagrant:~/data/book$ wc  -l
book-rating.csv
1149772 book-rating.csv
vagrant@vagrant:~/data/book$ wc -l user.csv
278860 user.csv
```

Rating that is given by different users on multi-criteria. The below table gives the user X item rating matrix of a random user.

User(u)	Item 1	Item 2	Item 3	Item i
U1	2	3	2	4
U2	?	5	4	5
U3	1	5	4	?

Table 1: User X item rating matrix (m*x*n)

```
vagrant@vagrant:~/data/book$ head book-
rating.csv

276725; 034545104; 0
276726; 0155061224; 5
276727; 0446520802; 0
276729; 052165615; 3
276729; 0521795028; 6
276733; 2080674722; 0
276736; 3257224281; 8
276737; 0600570967; 6
276744; 038550120; 7
276745; 342310538; 10
```

By doing the aggregation method from the rating that are the top most recommendation is being provided to the users as following:

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Website: www.ijircce.com

Vol. 5, Issue 4, April 2017

```
vagrant@vagrant:~/data/book$ head book.csv
0195153448;ClassicalMythology;Mark P. O.
Morford;2002;Oxford University Press;
0002005018;ClaraCallan;Richard Bruce
Wright;2001;HarperFlamingo Canada;
0060973129;Decision in Normandy;Carlo
D'Este;1991;HarperPerennial;
```

VI. CONCLUSION

Collaborative filtering method is basically used by users to rate items so that recommendation in social network. In this work we propose collaborative filtering using multi-criteria for different items according to real data-based experiments for items on multi-criteria accurate recommendation is got, where as compared to single-criteria. Scalability [9] of multi-criteria rating is been improved and that can be used with large scale data information system. Better recommendation is provided when ratings are given on multiple categories than on single category.

REFERENCES

[1] Prarthana Rao H M and Sathyanarayana S, "A Review on Big Data : Recommendation System," 2016.
[2] J. L. Herlocker, J. A. Konstan, et. al, "An algorithmic framework for performing collaborative filtering," in Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, Berkeley, CA, USA, 1999.
[3] G. Adomavicius and Y. Kwon, "New recommendation techniques for multi-criteria rating systems," IEEE Intelligent Systems, vol. 22, no. 3, pp. 48-55, 2007.
[4] S. Ancy and Vysali Unnikrishnan, "A Study on Clustering and Collaborative Filtering", 2015.
[5] M. Omair Shafiq, "Event Segmentation using MapReduce based big data clustering," 2016.
[6] Min – Seon Lee, Earl Kim, Choong – Sung Nam and Dong Ryeol Shin, "Design of Educational big data application using spark," 2017.Alper Bilge and Cihan Kaleli, "A Multi-Criteria Item-Based Collaborative Filtering Framework," 11th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2014.
[7] Manu M N and Ramesh B, "Single-Criteria Collaborative Filtering Implementation using Apache Mahout in Big Data," JSCE International Journal of Computer Science and Engineering, vol. 5, 2017.
[8] Afsaneh Khosravani and Mehrdad Jalali, "Recommending a method based on hybrid individual and group trust in collaborative filtering", 2015.
[9] Alper Bilge and Cihan Kaleli, "A Multi-Criteria Item-Based Collaborative Filtering Framework," 11th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2014.
[10] M. N. Manu and K. R. Anandakumar, "A current trends in big data landscape," 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Madurai, 2015.
[11] Manu M N and Ramesh B, "A Preliminary Study on Big Data Landscape," International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET), Vol. 6, Issue 1, January 2017.