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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

Toward personalize User-based Recommendation System for big data Application

Nasimkothiwale, Prof. Shriharikhatawkar

Student, Dept. of CSE., Shivaji University, Maharashtra, India,

Assistant Professor, Dept. of CSE., Shivaji University, Maharashtra, India

ABSTRACT: As the usage of internet upsurges, people comes with multiple choices. Recommendation system has given help to them by suggesting best services. It is software tools and techniques providing suggestions to users. The suggestion provided are aimed to support users in various decision making processes such as what item to buy, which music to listen, where to travel and so on. At that time huge amount of data is exposed. Data comes from different resources. The bulky data is called as "Big Data". Big Data is a collection of huge amount of datasets. They are having different type of data which is large in size and complex in structure that's why it is difficult to capture, manage and process the data within a short period. Recommendation system is used to data management term. The proposed system is implemented to solve challenges related to the big data application. Movies recommendation system is developed in proposed system. MovieLense Dataset is used to work with proposed system. UBSR implemented on Hadoop has favorable for scalability & efficiency in big data environment. After over all analysis of result proposed system shows accurate recommendation as per user satisfaction.

KEYWORDS: Recommender system, UBSR, preferences, keyword, Big Data, MapReduce, Hadoop.

I. INTRODUCTION

Big Data applications where data collection has grown tremendously and is beyond the ability of commonly used software tools to capture, manage, and process within a tolerable elapsed time [1]. The most fundamental challenge for Big Data applications is to explore the large volumes of data and extract useful information or knowledge for future actions. The solution to such a challenge is shifting increasingly from providing hardware to provisioning more manageable software solutions.

Every day, people are inundated with choices and options. What to buy? Which book to buy? Where to travel? Which blog post to read? Which movie to watch? And so on [3]. Each of these questions has many alternative solutions. With the growing number of alternative services, effectively recommending services that users preferred have become an important research issue.

Recommender systems have been shown as valuable tools to help users deal with services overload and provide appropriate recommendations to them [4]. Generally speaking, comparing with existing methods, personalize user-based recommendation system utilizes reviews of previous users to get both of user preferences and the quality of multiple criteria of candidate services, which makes recommendations more accurate[1]. Moreover, personalize user-based recommendation system implemented on MapReduce has promising scalability and efficiency.

II. RELATED WORK

The authors ShunmeiMeng, Wanchun[1] presents personalize recommendation list and as per user interest recommend the most appropriate items to the users. In keyword based recommendation system collaborative filtering algorithm is implemented on hadoop to raise appropriate recommendation. Hadoop is used for to improve scalability and efficiency.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

The authors, X. Yang, Y. Guo, and Y. Liu [2], implemented Bayesian-inference-based recommendation system which is used for social networking sites. This is excellent than existing trust-based recommendation system. They show that active user gives reviews to every item and these reviews can see all users which are connected to active user.

In [3], Adomavicius and Tuzhilin give an overall structure of recommender systems. It describes the current generation recommendation approaches. In next generation recommendation system gives the solution on limitation of current generation recommendation system. This recommendation system gives idea about how to improve and how to make robust system. It is flexible in even broader range of application.

The [4] existing recommendation system only single rating criterion is applicable to generate recommendation list. But if without considering single one criteria, considering multi-criteria is most effective than other. Here rating, features and attribute of items are used to produce most appropriate recommendation. The feedbacks of users are also tack into account at the time of recommendation.

The authors Z D Zhao and M. S. Shang of [5] developed CF algorithm on Hadoop parallel processing paradigm. This is favourable to scale the large application by dividing the datasets to solve the scalability problem. The MapReduce and cascading technologies are used to implement scalable recommendation also in falksonomy information if presents a parallel user profiling approaches.

M. Hu, H. Singh, D. Rule, M. Berlyant, and Z. Xie Y. Jin [6] presenting a large scale video recommendation system implemented by using item-based CF algorithm. They implement their proposed approach in Qizmt, which is a .Net MapReduce framework, thus their system can work for large scale video sites.

The authors [7] proposed a trust-aware system for generating personalized user recommendations in social networks. Its foundations lie on a reputation mechanism that is mathematically formulated, comprising both local and collaborative rating formation. The proposed system provides users with personalized positive and/or negative recommendations that can be used to establish new trust/distrust connections in the social network.

The author [8] proposed location-aware recommender system they introduces some special types of ratings those are spatial ratings and non- spatial ratings. Those techniques are efficient, scalable and accurate one for recommendation system as compared with traditional recommendation system. They deals with scalability problem solve by applying those technology on Hadoop.

In this paper author [16] presents personalized recommendations are used to support the activities of learners in personal learning environments and this technology can deliver suitable learning resources to learners. This paper models the dynamic multi-preferences of learners using the multidimensional attributes of resource and learner ratings by using data mining technology to alleviate sparsity and cold-start problems and increase the diversity of the recommendation list. The proposed method outperforms current algorithms on accuracy measures and can alleviate cold-start and sparsity problems and also generate a more diverse recommendation list.

III. SYSTEM ARCHITECTURE

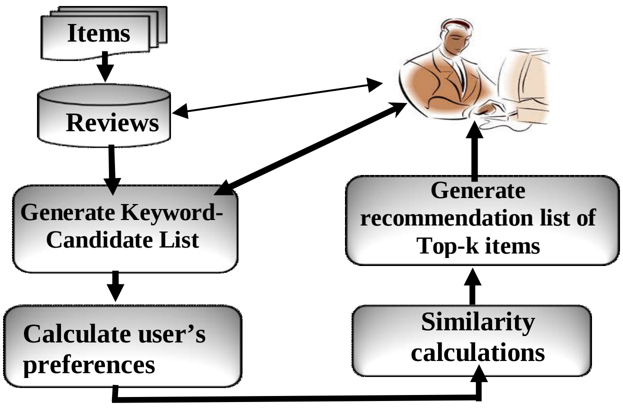

Fig.1.System Architecture

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

A. *Models:*

- Generate Keyword-Candidate List and Domain Treasures.
- Calculate user's preferences/choices.
- Similarity calculations.
- Then finally, Generate recommendation list of Top-k items.
- Implementation on map-reduce.

B. *Generate Keyword-Candidate List and Domain Treasures*

The keyword-candidate list is a set of keywords about users preferences, Keywords in the keyword-candidate list can be a word or multiple words related with the quality criteria of candidate services. A domain thesaurus is a reference word of the keyword-candidate list that lists of words grouped together according to the similarity of keyword meaning, including related and contrasting words and antonyms [2].

Here purposed system presents the simple recommendation list as per the rating selected by the user. The collaborative filtering algorithm is adopted to measure the similarity in between active user.

$$P_{a,u} = \frac{\sum_{i=1}^{m}(r_{a,i} - \bar{r}_a) \times (r_{u,i} - \bar{r}_u)}{\sqrt{\sum_{i=1}^{m}(r_{a,i} - \bar{r}_a)^3 \times \sum_{i=1}^{m}(r_{u,i} - \bar{r}_u)^3}}$$

C. *Calculate user's preferences/choices*

In this step, the preferences of active users (current user) are formalized into their corresponding preference keyword sets respectively.

The keyword extraction process [6] is done this step. And the Porter Stemmer algorithm (keyword stripping) is used to calculate user preferences/ choices.

D. *Similarity calculations*

The two similarity computational methods are introduced in our recommendation system [7].

- ✓ Approximate Similarity Computational

  A Jaccard coefficient algorithm is adopted to generate appropriate similarity computation.

$$sim(APK, PPK) = Jaccard(APK, PPK) = \frac{|APK \cap PPK|}{|APK \cup PPK|}$$

- ✓ Exact Similarity Computational

  A cosine-based approach is applied in the exact similarity computation.

$$sim(APK, PPK) = Cos(W_{AP}, W_{PP}) = \frac{W_{AP} \bullet W_{PP}}{||W_{AP}||_2 \times ||W_{PP}||_2}$$

E. *Then finally, Generate recommendation list of Top-k items*

Based on the similarity of the active user and previous users, further filtering will be conducted.

Given a threshold δ, of sim(APK, PPKj) < δ, , the preference keyword set of a previous user PPKj will be filtered out, otherwise PPKj will be retained.

Once the set of most similar users are found, the personalize ratings of each candidate service [5] for the active user can be calculated.

F. *Implementation on map-reduce*

To improve the scalability and efficiency of our recommendation method for Big Data environment it is Implementation on hadoop map-reduce framework [3].

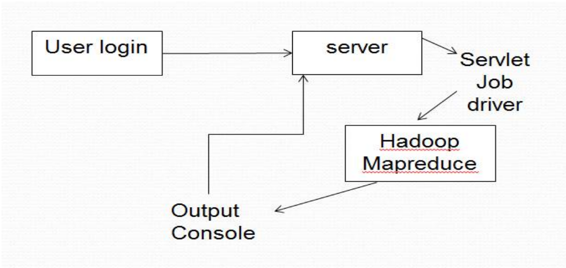

Fig.2.Hadoop Processing on servlet

IV. Proposed Algorithm

ALGORITHM OF UBSR:

Input: The preference rating set of the active user APK
The rating set RS = {rs1; rs2; :::rsn }
The Threshold δ in the filtering phase
The Top-k number
Output: The service with the Top-k highest rating {tws1; tws2; ……..twsk}

for each rating RSi Є RS
Rj =Φ, sum = 0, r =0
for each rating Rj of services RSi
process the rating into a previous preference set PPKj
if PPKj ∩ APK ≠ Φ Then
insert PPKj into Rj
end if
end for
for each rating set PPKj Є Rj
sim (APK, PPKj) =SIM (APK, PPKj)
if sim (APK, PPKj) < δ then
remove PPKj from Rj
else sum = sum + 1, r = r + 1
end if
end for
prj = r/sum
get prj
end for
sort the personalized ratings set prj
return the services with the Top-K highest ratings {tws1; tws2; …….twsk}

V. Result Analysis

In this section, experiments are conducted and studied to estimate the accuracy and scalability of UBRS. To observe the performance of UBRS in accuracy, we compare UBRS with other three well-known recommendation methods:

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

1. Keyword aware service recommendation system (KASR).
2. User-based algorithm using Pearson Correlation Coefficient (UPCC)
3. Item-based algorithm using Pearson Correlation Coefficient (IPCC).

Three metrics are used to observe the accuracy:

1. Mean absolute error (MAE),
2. Mean average precision (MAP)
3. Discounted cumulative gain (DCG).

As to the scalability, a well-accepted scalability metric, Speedup, is accepted to determine the overall behavioral in the scalability of UBRS

The experiments are divided into two parts that observe the accuracy and scalability of UBSR.

1. Accuracy Evaluation
2. Scalability Evaluation

The experiments are going too conducted on MovieLens Dataset.

Accuracy Evaluation:

1. *Comparison of UBSR with UPCC, IPCC and KASR in MAE.*

MAE is a statistical accuracy metric often used in CF methods to measure the prediction quality and accuracy. The lower the MAE presents the more accurate predictions.

Fig. 3 shows the MAE values of UBSR, UPCC, IPCC and KASR. It could be found that the MAE value of UBSR is much lower than KASR, UPCC and IPCC. Thus our methods UBSR can provide more accurate predictions than Existing methods KASR, UPCC and IPCC.

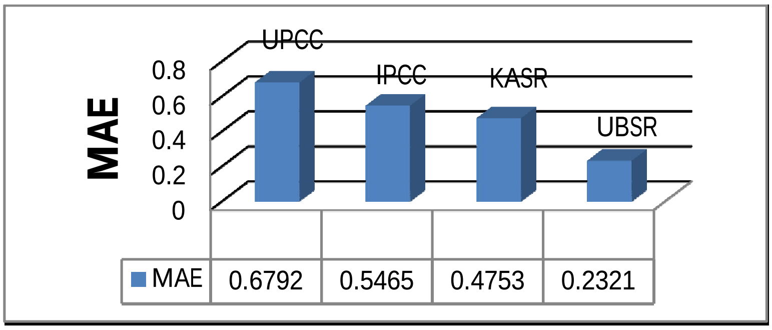

Fig.3.the MAE values of UBSR

2. *Comparison of UBSR, KASR, UPCC and IPCC in MAP and DCG.*

To observe the quality of Top-K service recommendation list, MAP and DCG are used as performance measurement metrics. And the higher MAP or DCG presents the higher quality of the predicted item recommendation list.

From Fig 4 and 5, shows that the MAP values and DCG values of UBSR are comparatively higher than KASR, UPCC and IPCC. It also could be found that the DCG values increase when K increases or the MAP values decrease when K increases.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

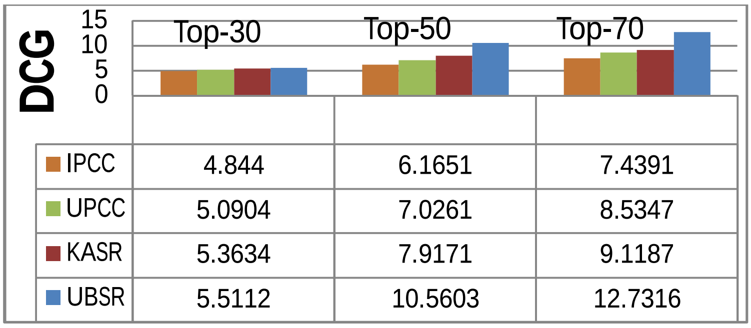

Fig.4.the DCG value of UBSR

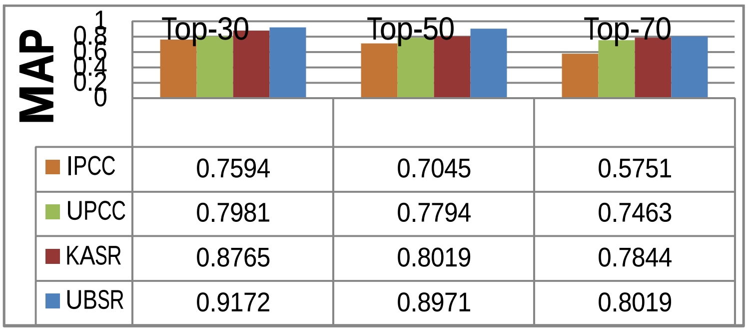

Fig.5.the MAP value of UBSR

For a more examination, the percentage values that UBSR verses KASR, UPCC and IPCC in MAP and DCG of Top-K (k=30, 50, 70) recommendation list are calculated and listed in Table 1 and Table 2.

Table 1 presents the percentage values of UBSR verses KASR in MAP of Top-K recommendation list (e.g., When K=30, UBSR outperforms KASR 4.65% ((0.9172-0.8765)/0.8765 = 4.65%) in MAP of Top-30 recommendation list).

Table 1 the percentage values of UBSR in MAP

MAP	Top-30	Top-50	Top-70
UBSR/KASR	4.56 %	11.9 %	2.23 %
UBSR/UPCC	14.9 %	15.1 %	7.45 %
UBSR/IPCC	20.8 %	27.3 %	39.4%

Table 2 present the percentage values of UBSR verses KASR, UPCC and IPCC in DCG of Top-K (K =30, 50, 70) recommendation list. It could be found that UBSR can provide more accurate service recommendation list than KASR, UPCC and IPCC.

Table 2 the percentage values of UBSR in DCG

DCG	Top-30	Top-50	Top-70
UBSR/KASR	2.7	33.3	72.5
UBSR/UPCC	8.2	50.3	84.3
UBSR/IPCC	13.7	71.3	111.5

Generally speaking, UBSR perform better than traditional methods KASR, UPCC and IPCC, in MAE, MAP and DCG. Thus the user-based personalized service recommendation lists provided by our method would satisfy users better.

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

International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 7, July 2016

VI. CONCLUSION AND FUTURE WORK

During this work, problem of accuracy, inefficiency and speedup is identified by the literature survey. Then proposed system which gives solution to problem identified. Proposed system also used for to improve the scalability and inefficiency in "Big Data" environment, proposed system implemented it on a MapReduce framework in Hadoop platform.Comparing with existing system, UBSR generate accurate recommendation. UBSR implemented on Hadoop has favourable for scalability & efficiency.

In future work, proposed system will do further research in how to deal with the case where term appears in different categories of a context and how to distinguish them.

REFERENCES

1. ShunmeiMeng, Wanchun Dou, Xuyun Zhang, and Jinjun Chen, Senior Member, IEEE KASR: A Keyword-Aware Service Recommendation Method on MapReduce for Big Data Applications IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 25, NO. 12, DECEMBER 2014.
2. X. Yang, Y. Guo, and Y. Liu, Bayesian-Inference Based Recommendation in Online Social Networks, IEEE TRANS. PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 24, NO.4, PP. 642-651, APR. 2013.
3. A. Tuzhilin and G. Adomavicius,"Toward the Next Generation of Recommender Systems:A Survey of the State of the Art and Possible Extensions," IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 17, NO. 6, PP. 734-749, 2005.
4. G. Adomavicius and Y. Kwon, New Recommendation Techniques for Multicriteria Rating Systems, IEEE Intelligent Systems, vol. 22, no. 3, pp. 48-55, May/June 2007.
5. Z D Zhao and M. S. Shang, "User Based Collaborative Filtering Recommendation Algorithms on Hadoop," In the third Internation-alWorkshop on Knowledge Discovery and Data Mining, pp. 478-481, 2010.
6. M. Hu, H. Singh, D. Rule, M. Berlyant, and Z. Xie Y. Jin, "MySpace Video Recommendation with Map-Reduce on Qizmt," Proeedings of the 2010 IEEE Fourth International Conference on Semantic Computing, pp. 126-133, 2010.
7. MagdaliniEirinaki, Malamati D. Louta, Member, IEEE, and IraklisVarlamis, Member,IEEE "A Trust-Aware System for Personalized User Recommendations in Social Networks"IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS, VOL. 44, NO. 4, APRIL 2014 409.
8. Sarwat, Justin J. Levandoski, Ahmed Eldawy, and Mohamed F. Mokbel IEEE LARS*: An Efficient and Scalable Location-Aware Recommender System Mohamed TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 26, NO. 6, JUNE 2014.
9. Y. Chen, A. Cheng, and W. Hsu, "Travel Recommendation by Mining People Attributes and Travel Group Types from Community-Contributed Photos," IEEE Trans. Multimedia, vol. 25, no. 6,pp. 1283-1295, Oct. 2013.
10. C. Lynch, "Big Data: How do your data grow?" Nature, vol. 455, no. 7209, pp. 28-29, 2008.
11. Y. Pan and L. Lee, "Performance Analysis for Lattice-Based Speech Indexing Approaches Using Words and Subword Units," IEEE Trans. Audio, Speech, and Language Processing, vol. 18, no. 6,pp. 1562-1574, Aug. 2010.
12. G. Kang, J. Liu, M. Tang, X. Liu, and B. cao, "AWSR: Active Web Service Recommendation Based on Usage History," Proc. IEEE19th Int'l Conf. Web Services (ICWS), pp. 186-193, 2012.
13. G.M. Amdahl, "Validity of the Single-Processor Approach to Achieving Large Scale Computing Capabilities," Proc. Spring JointComputer Conf., pp. 483-485, 1967.
14. H. Liang, J. Hogan, and Y. Xu, "Parallel User Profiling Based on Folksonomy for Large Scaled Recommender Systems: An Implimentation of Cascading MapReduce," Proc. IEEE Int'l Conf. DataMining Workshops, pp. 156-161, 2010.
15. B. Issac and W.J. Jap, "Implementing Spam Detection Using Bayesian and Porter Stemmer Keyword Stripping Approaches," Proc. IEEE Region 10 Conf. (TENCON '09), pp. 1-5, 2009
16. An Effective Recommendation Framework for Personal Learning Environments Using a Learner Preference Tree and a GA MojtabaSalehi, Isa NakhaiKamalabadi, and Mohammad B. GhaznaviGhoushchi, Member, IEEE IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES,VOL. 6, NO. 4, OCTOBER-DECEMBER, 2013.