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 2, February 2016

User Based Personalized Search for Service Recommender System with Bigdata

Shanmuga Priya.V, Jayakumar.P

M.E. Student, Dept. of C.S.E., Arunai Engineering College, Thiruvannamalai, India

Asst. Professor, Dept. of C.S.E., Arunai Engineering College, Thiruvannamalai, India

ABSTRACT:Big data is an emerging technology, as the name suggests is all about handling large amount of data., method and procedure the data within a tolerable elapsed time. A huge data is collected and processed to make some decision and also used to describe any type of data which may be structured, semi-structured, unstructured and if the data grows.Big Data became a challenge for IT companies.Service recommender system such as hotel reservation system and restaurant guides the ratings of services. They also provide the same service recommendation list to all the users. User's different preferences were not considered and also fail to meet the personalized requirements. Service utility is represented as a whole for all users which is based on single numerical rating in existing service recommender system.Scalability problem is solved ,but not provide favourable scalability and efficiency when the amount of data grows.In this paper, a new method is proposed called as KASR(Keyword Aware Service Recommendation)method is based on an algorithm is User based collaborative filtering algorithm Also to calculate the personalized rating of each candidate service for a user and also provide a personalized service recommendation list to appropriate service to users MapReduce is used in Hadoop for computing structure which is implement on a distributed computing.In KASR, keywords are used to denote both user's preferences and the quality of services.. This method is used to improve the scalability and efficiency ,if data grows.

KEYWORDS: Service recommendation, KASR ,personalized requirements, Map reduce, hadoop, collaborative filtering.

I. INTRODUCTION

Now a days, Bigdata is an new technology as the name suggest is all about handling large amount of data.Bigdata new technologies make it possible to realize value from Big Data. Using conventional software tools its impractical and complex to handle large datasets in bigadata.. There is foremost dispute in corporate company is bigdata. It must support application data to instrument and sensor data to public, picture and geospatial data.

The dimension of bigdata is volume, velocity or variety of data .Specifically, Big Data relate to data formation, storage, retrieval and analysis that is remarkable in terms of volume, velocity, and variety. In hotel reservation systems the same services ratings and the list of service recommendations are given to all user. In the traditional service recommender systems, they just provide the overall ratings about the hotel. To represent a service utility it fully based on a single numerical ratings.

The personalized requirements of each user are not considered in traditional service recommender system. They have not considered users' different preferences, without meeting users' personalized requirements. In current recommender system will leads to inefficiency,scalability issues.There are three main categories in the current service recommendation methods,they are content-based, collaborative, and hybrid recommendation. Content-based approach recommend services like to those the user prefer in the past.Collaborative filtering (CF) approach recommend services to the user that users with like tastes chosen in the past. Hybrid approaches coalesce content-based and CF methods in some different ways.

In Colloborative filtering systems, it provide similar relish and preference of the user, that can be classified as item-based CF and user-based CF.. By using this algorithm it will provide appropriate recommendation as per the need of each user.In proposed system ,a new method called KASR is based on user based collaborative algorithm.

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 2, February 2016

MapReduce is used in Hadoop for computing framework which is implemented on a distributed computing.

II. RELATED WORK

In many recommender systems residential in both the academy and industry. The author suggest a Bayesian-inference-based recommendation system for online social networks.The proposed Bayesian-inference-based recommendation is enhanced than the existing trust-based recommendations and is analogous to Collaborative Filtering recommendation. They also explain a variety of restrictions of present service recommendation methods, and discuss possible extension that can get better recommendation capability and make recommender systems valid to an even broader variety of applications. The majority existing service recommender systems are merely based on a single numerical rating to characterize a service's utility as a whole. Actually, evaluate a service through several criteria and taking into description of user response can help to formulate additional efficient recommendations for the users. With the advance growth of cloud computing software tackle such as Apache Hadoop, Map-Reduce and Mahout, it become probable to propose and implement the scalable recommendation systems in Big Data environment. The implementation of a CF algorithm on Hadoop .They resolve the scalability difficulty as a result of dividing data set. But their technique doesn't have good scalability and efficiency if the quantity of data grows. It present a equivalent user profile approach based on folksonomy information and implements a scalable recommender system by using Map-Reduce and cascade techniques .Item-based CF algorithmis used to propose a large-scale video recommendation system. They implement their proposed approach in Qizmt, which is a .Net Map-Reduce framework.KASR utilizes reviews of preceding users to get both of user preferences and the quality of various criterion of candidate services which makes recommendations more exact. Moreover, KASR on MapReduce has good scalability and efficiency.

III. PROPOSED SYSTEM

In the paper ,five modules are used to describe the system architecture.

1.Big Data and Environment
2.Batching and Preprocess
3.Digging in Big Data &Service Recommender Application
4.MapReduce and Hadoop
5.KASR and Analysis

1. *Big Data and Environment:*

A massive amount of data is retrieved from open source datasets that are freely obtainable from the Travel Recommendation Applications. Big Data Schemas were analyze and a function ruling of the Schema is determined. Java API is used to read and manipulate the CSV(Comma separated values)that itself developed by us which is developer gracious ,light weighted and easily changeable.

2. *Batching and Preprocess:*

The conventional vision of Service Recommender Systems that show Top-K Results are displayed with Paginations with which a user can steer Back and Forth of the Result sets. Every Services Ratings and Reviews of all hotels are listed. POS(Parts of Speech) Tagger and Chuncker Process done on each and every review of all hotels for all countries in a analogous and disseminated Manner as Batch jobs. The master Job is divide into n no of little Batch jobs based on the slave machines associated with the master. POS Tagger tags every words of a review with its tags and the Chunker Process will receive POS tagged output as input for Grouping the Words based on meaning of the Review.

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 2, February 2016

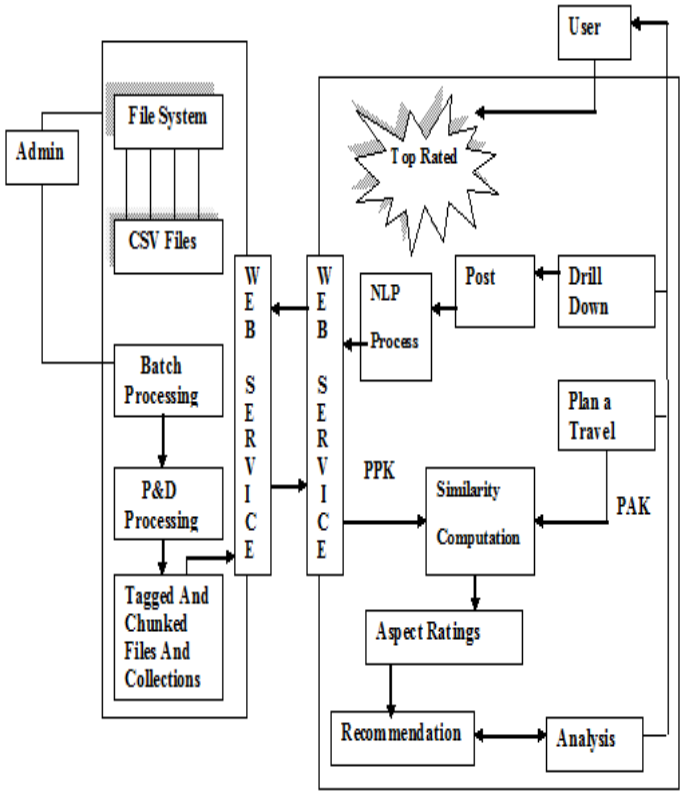

Fig.1 System Architecture

3. Service Recommender Application:

The CSV Files in distributed Systems are invoke through Web Service operation in the Server Machine of the host Process during a Web Service Client Process in the Recommendation System. The data that retrieve toward the Recommendation Systems are provided with a clean GUI and can be query on demand.Recommendation Application invoke Web Service which use light weighted traversal of data with XML. The Users can give review each hotel and can post commentary also. The Reviews gets updated to the CSV Files as it get retrieved. A User can map or list a Travel highlighting his necessities in a complete way that shows the preference keywords set of the Active User. Keyword Candidate List and Candidate Services List based on Domain Thesaurus. The Domain Thesaurus can be updated frequently to get exact results of the Recommendation System.

4. MapReduce and Hadoop:

(1) User preferences are captured by a keyword-aware method:

In this step, the preferences of active users and previous users are solemn into their equivalent preference keyword sets correspondingly. Active user refers to a present user needs recommendation.

a)Active user's preference:

An active user can provide his/her preference in relative to candidate services by select keyword from a keyword-candidate list, which replicate the quality criteria of the services he/she is concerned about. In addition the active user should also select the importance degree of the keywords. The importance degree of the keywords as '1' denotes the 'general', 3 denote the 'essential' and '5' denotes the 'very essential '.

b) Previous users Preferences :

According to the keyword-candidate list and domain thesaurus the reviews are extracted from the preference of a earlier user for a candidate service. Moreover a review of the previous user will be dignified into the preference keyword set of user.

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 2, February 2016

(2) Keyword extraction process:

a) Preprocess:

Firstly, tags and stop words in the reviews snippet collection should be detached to avoid affect the quality of the keyword extraction in the next stage. And the Porter Stemmer algorithm is used to eliminate the commoner morphological and inflexional endings from words in English.

b) Keyword extraction:

In this stage, all review will be changed into a consequent keyword set according to the keyword-candidate list and domain thesaurus. If the review contain a word in the domain thesaurus, then the equivalent keyword should be extract into the preference keyword set of the user well-known as PPK.

(3) Compute the similarity:

The Third step is to identify the review of previous users(PPK) who have similar tastes to an active user by discovery neighbourhood of the active user(PAK) based on the similarity of their preferences. Before similarity evaluation, the review not related to the active user's preferences will be filtered out by using set theory. If the set is empty when the intersection of the preference keyword sets of the active user and a previous user then the preference keyword set of the previous user will be filtered out.

(4)HADOOP platform:

Hadoop is a open source,free,Java based programming framework that support the process of big data sets in a disseminated computing environment. It is aspect of the Apache project sponsored by means of the Apache software foundation .The active users preferences and previous user preferences are transformed into equivalent keyword sets PAK and PPK respectively.PAK denotes the keyword set of the active user preferences and PPK denote the the keyword set of the previous user preferences.

Active user:

The current user need recommendation is referred as an active user. By select the keyword from a keyword-candidate list is based on the preference of the active user and also the significance degree of the keywords as '1' denotes the 'general', 3 denote the 'essential' and '5' denotes the 'very essential '.

Previous user:

The user who has visited the particular hotel and give his/her review that will be updated. Extract the keywords from the reviews.In the review of snippet collection the meaningless tags and stop words are removed.

Porter stemmer algorithm is used to remove morphological from the words.According to domain thesaurus and candidate lists all review will be transformed. The review contains a keyword same as in the domain thesaurus then the equivalent keyword should be extracted into PPK.

Similarity computation:

Active user is to be identified based on the reviews of previous user,who have similar tastes.Based on the similarity preferences able to find the neighbourhood of the active user .If the set is empty when the intersection of PPK and PAK it leads to there is no similarity keyword in between the active and previous user.The active and previous user will be removed if the reviews are not similar.

(5)Analysis in KASR:

The Chunked Reviews of the similar user list is retrieved and the Keywords equivalent to the User is analyzed for its Valence and Arousal. Valence Means Weather the Keywords Means a positive or Negative thing and Arousal answers, how much it is? Ratings are given for each Domain based on the Valence and Arousal for each User of each hotel. The Overall Hotel Rating is now manipulated by taking average values of each rating of several users of a particular hotel. Now ranking is done for all hotels based on Ratings and will be sorted based on Bubble Sort Algorithm to have the Most appropriate personalized Recommendation for the User. The Results will be analyzed with Graphical Views so as to understand easier.

ENHANCEMENT:

The Natural Language Processing is implemented to analyze the reviews of the previous user. The NLP Process comprise Tokenizing a Sentence or a word, POS&Tagging,extraction of Nouns and Verbs, Synonym recovery and Spell Check of extract Keywords using WordNet Dictionary .Valence and Arousal will be implemented for calculating Ratings of Aspects of a Hotel. The BigData manipulation from CSV through Our Own JAVA API enforces developer friendly access.

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 2, February 2016

IV. EXPERIMENTAL RESULTS

The experiments are designed and analysed to evaluate the accuracy and scalability of KASR. To evaluate the performance of KASR in accuracy, we compareKASR with other two well-known recommendation methods:user-based algorithm using Pearson Correlation Coefficient and item-based algorithm using PCC,which are called as UPCC and IPCC respectively.Three metrics are used to evaluate the accuracy mean absolute error (MAE) , mean average precision (MAP) and discounted cumulative gain (DCG).Two groups of experiments are conducted to evaluate compare KASR with UPCC and IPCC in MAE, MAP and DCG to evaluate the accurateness of KASR. The other is to explore the scalability of KASR.

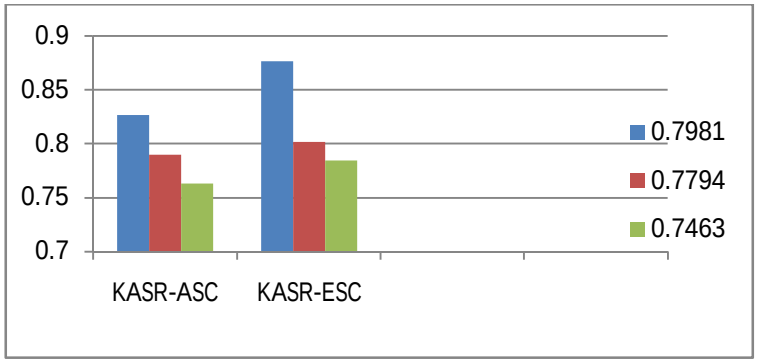

Fig.2.a. Comparison of KASR-ESC and KASR-ASC

Accuracy Evaluation:

MAE is a statistical accuracy metric frequently used in CF methods to compute the prediction quality. And the normalized mean absolute error (NMAE) metric is used to calculate the accuracy.It point up the MAE and NMAE values of UPCC,IPCC, KASR-ASC and KASR-ESC. It could be create that the MAE and NMAE vaules of KASR-ASC and KASR-ESC are much lower than UPCC and IPCC.

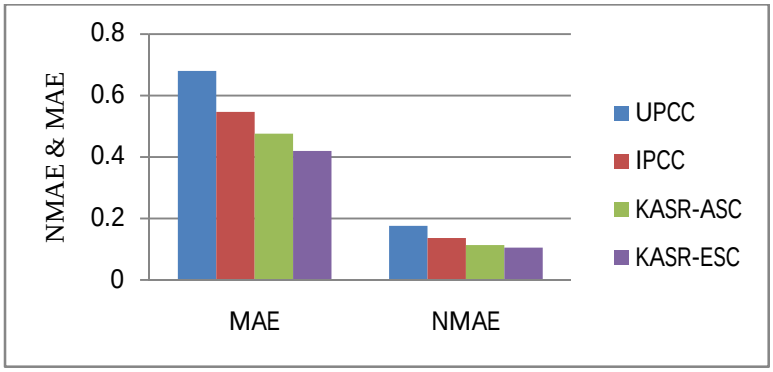

Fig.2.b.Comparison of UPCC, IPCC, KASR-ASC and KASR-ESC in MAE

Scalability Evaluation:

To verify the scalability of KASR, experiment is conducted respectively in a cluster of nodes ranging from 1 to 8. There are four synthetic data sets used in the experiments (128M, 256M, 512M and 1 Gdatasize). show the speedup of KASR .From we can see that the speedup of KASR increased relative linearly with the growth of the number of nodes.

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 2, February 2016

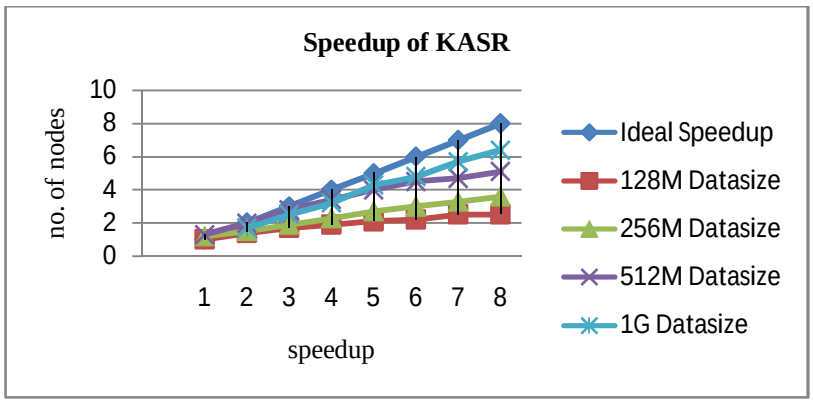

Fig.2.c. Speedup of KASR

Meanwhile, larger data set obtained a improved speedup.When the datasize is 1G and the number of nodes is 8,the speedup value reaches 6.412, which is 80.15 percent(6.412/8 ¼ 80.15%) of the ideal speedup. The experimental result shows that KASR on Map-Reduce in Hadoop platform has excellent scalability over Big Data and performs better with larger data set.Overall, these experimental results show that KASR performs well in accuracy, and KASR on Mapreduce framework has good scalability in "Big Data" environment.

V. CONCLUSION AND FUTURE WORK

In existing service recommender systems, there is an issue called scalability and inefficiency.To overcome the issue by using a method KASR-keyword aware service recommendation method. By select the keyword from a keyword-candidate list is based on the preference of the active user and also the significance degree of the keywords as '1' denotes the 'general', 3 denote the essential and '5' denotes the very essential .User-based collaborative filtering algorithm is use to make appropriate recommendation.To obtain meaningful keyword by doing the POS tagging and chunking process. From the reviews of previous user. we can separate the positive and negative preference's using Natural Language Processing. Valence and Arousal will be implemented for calculating Ratings of Aspects of a Hotel. The implementation of KASR on mapreduce framework, parallel computing in hadoop platform is to deal the bigdata issues in existing system so that in proposed system used to improve the scalability and efficiency.

REFERENCES

1. W. Dou, X. Zhang, J. Liu, and J. Chen, "HireSome-II: TowardsPrivacy-Aware Cross-Cloud Service Composition for Big Dat Applications," IEEE Trans. Parallel and Distributed Systems,2013.
2. 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. 2013
3. Z. Zheng, X. Wu, Y. Zhang, M. Lyu, and J. Wang, "QoS Ranking Prediction for Cloud Services," IEEE Trans. Parallel and Distributed Systems, vol. 24, no. 6, pp. 1213-1222, June 2013
4. Z. Luo, Y. Li, and J. Yin, "Location: A Feature for Service Selection in the Era of Big Data," Proc. IEEE 20th Int'l Conf. Web Service,pp. 515-522, 2013
5. 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.
6. L. Zhang, "Editorial: Big Services Era: Global Trends of Cloud Computing and Big Data," IEEE Trans. Services Computing, vol. 5,no. 4, pp. 467-468, Fourth Quarter, 2012.
7. G. Kang, J. Liu, M. Tang, X. Liu, and B. cao, "AWSR: Active Web Service Recommendation Based on Usage History," Proc. IEEE 19th Int'l Conf. Web Services (ICWS), pp. 186-193, 2012.
8. J. Manyika et al., "Big Data: The Next Frontier for Innovation,Competition, and Productivity," 2011.
9. K. Lakiotaki, N.F. Matsatsinis, and A. Tsoukis, "Multi-CriteriaUser Modeling in Recommender Systems," IEEE Intelligent Systems,vol. 26, no. 2, pp. 64-76, Mar./Apr. 2011.
10. M.Bjelica,"Towards TV Recommender System Experiments with User Modeling," IEEE Trans. Consumer Electronics, vol.56, no. 3,pp. 1763-1769, Aug. 2010.