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 Personalised Search with Big Data

C.Pabitha, V.Juli Stephy

Assistant Professor, Dept. of C.S.E, Valliammai Engineering College, Chennai, India

PG Scholar, Dept. of C.S.E, Valliammai Engineering College, Chennai, India

ABSTRACT:Personalized service recommendation list and recommending the mostappropriate services to the users effectively. Specifically, keywords are used to indicate users' preferences, and a user-based Collaborative Filtering algorithm is assumed to generate appropriate recommendations. To increase its scalability and efficiency in big data environment,Keyword-Aware Service Recommendation method, named KASR, to address scalability and inefficiency problem in BigData with traditional service recommender systems, which fails to meet users' personalized requirements and various Preferences.NLP process comprises tokenizing a sentence or a word, part of speech tagging, extraction of nouns and verb, synonyms, retrieval and spell check of extraction keyword using wordnet dictionary. Valance and arousal will be implemented for calculating rating of aspect of the hotel.

KEYWORDS: Natural language processing, personalized service recommendation, collaboration filtering.

I. INTRODUCTION

In recent years, the amount of data in our world has been increasing explosively, and analysing large data sets the so-called "Big Data" becomes a key basis of competition underpinning new waves of productivity growth, innovation, and consumer surplus. Then, what is "Big Data"?, Big Data refers to datasets whose size is beyond the capability of current technology, method and theory to capture, man-age, and process the data within a tolerable elapsed time. With the growing number of alternative services, effectively recommending services that users wished have become an important research issue. Service recommender systems have been shown as valuable tools to help users deal with services overload and provide appropriate recommendations to them. Service recommender systems have been shown as valuable tools for providing appropriate recommendations to users. In the last decade, the amount of customers, services and online information has grown rapidly, yielding the big data analysis problem for service recommender systems. Consequently, traditional service recommender systems often suffer from scalability and inefficiency problems when processing or analysing such large-scale data. Moreover, most of existing service recommender systems present the same ratings and rankings of services to different users without considering diverse users' preferences, and therefore fails to meet users' personalized requirements. Current recommendation methods usually can be classified into three main categories: content-based, collaborative, and hybrid recommendation approaches. Content-based approaches recommend services similar to those the user preferred in the past. Collaborative filtering (CF) approaches recommend services to the user that users with similar tastes preferred in the past. Hybrid approaches combine content-based and CF methods in several different ways. In CF based systems, users receive recommendations based on people who have similar tastes and preferences, which can be further classified into item-based CF and user-based CF. In item-based systems; the predicted rating depends on the ratings of other similar items by the same user. While in user-based systems, the prediction of the rating of an item for a user depends upon the ratings of the same item rated by similar users. And in this work, we will take advantage of a user-based CF algorithm to deal with our problem.

II. RELATED WORK

The recommender is based on audio-visual consumption and does not depend on the number of users, running only on the client side. This avoids the concurrence, computation and privacy problems of central server approaches in scenarios with a large number of users, such as the Olympic Games [1]. The system has been designed to take advantage of the information available in the videos, which is used along with the implicit information of the user and the modelling of his/her audio-visual content consumption. The system is thus transparent to the user, who does not

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

need to take any specific action. Another important characteristic is that the system can produce recommendations for both live and recorded events.

In this paper, they propose a Bayesian-inference-based recommendation system for online social networks. In our system, users share their content ratings with friends. The rating similarity between a pair of friends is measured by a set of conditional probabilities derived from their mutual rating history [2]. A user propagates a content rating query along the social network to his direct and indirect friends. Based on the query responses, a Bayesian network is constructed to infer the rating of the querying user. They develop distributed protocols that can be easily implemented in online social networks. . They further propose to use Prior distribution to cope with cold start and rating sparseness. The proposed algorithm is evaluated using two different online rating data sets of real users. . They show that the proposed Bayesian-inference-based recommendation is better than the existing trust-based recommendations and is comparable to Collaborative Filtering (CF) recommendation. It allows the flexible trade offs between recommendation quality and recommendation quantity.

They propose to conduct personalized travel recommendation by further considering specific user profiles or attributes (e.g., gender, age, race) as well as travel group types (e.g., family, friends, couple)[3]. Instead of mining photo logs only, they exploit the automatically detected people attributes and travel group types in the photo contents. By information-theoretic measures, we demonstrate that such detected user profiles are informative and effective for travel recommendation—especially providing a promising aspect for personalization. The photos gathered from various communities and rich people attributes and travel group types can be automatically detected and provide another important aspects in terms of travel demographics. Rather than the plain travel frequencies from or to certain locations, we can further investigate the demographic distributions in these trips via the statistics of detected people attributes and group types.

Recommender systems for learning try to address these challenge, they attempt to filter content for different learning setting. The authors presented an extensive overview of TEL recommendersystems. In addition, evaluation perspectives on current research in this area and future challenges with respect to the evaluation of TEL recommenders were discussed [4]. The notion of context has started to attract significant attention in this research, as indicated by contributions to a recent special issue on Context-Aware Recommender Systems (CARS). A new set of recommender systems for learning has been developed in recent years to demonstrate the potential of contextual recommendation. From an operational perspective, context is often defined as an aggregate of various categories that describe the setting in which a recommender is deployed, such as the location, current activity, and available time of the learner. A first example of a context-aware recommender system for learning considers the location of the user and the noise level at this location as a basis to suggest learning resources. If the learner is in a cafeteria, the noise level associated to this location might have an impact on her level of concentration and likelihood of interruption.

III. EXISTING SYSTEM

In most existing service recommender systems, such as hotel reservation systems and restaurant guides, the ratings of services and the service recommendation lists presented to users are the same. They have not considered users' different preferences, without meeting users' personalized requirements.Most existing service recommender systems are only based on a single numerical rating to represent a service's utility as a whole. In fact, evaluating a service through multiple criteria and taking into account of user feedback can help to make more effective recommendations for the usersExisting Approaches solve the scalability problem by dividing dataset. But their method doesn't have favourable scalability and efficiency if the amount of data grows.

IV. PROPOSED SYSTEM

A keyword-aware service recommendation method, named KASR, is proposed in this paper, which is based on a user-based Collaborative Filtering algorithm. In KASR, keywords extracted from reviews of previous users are used to indicate their preferences. Moreover, we implement it on a distributed computing platform, Hadoop, which uses MapReduce as its computing framework. In Kasr, keywords are used to indicate both of users' preferences and the quality of candidate services. A user-based CF algorithm is adopted to generate appropriate recommendations. KASR aims at calculating a personalized rating of each candidate service for a user, and then presenting a personalized service recommendation list and recommending the most appropriate services to him/her. Moreover, to improve the scalability

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

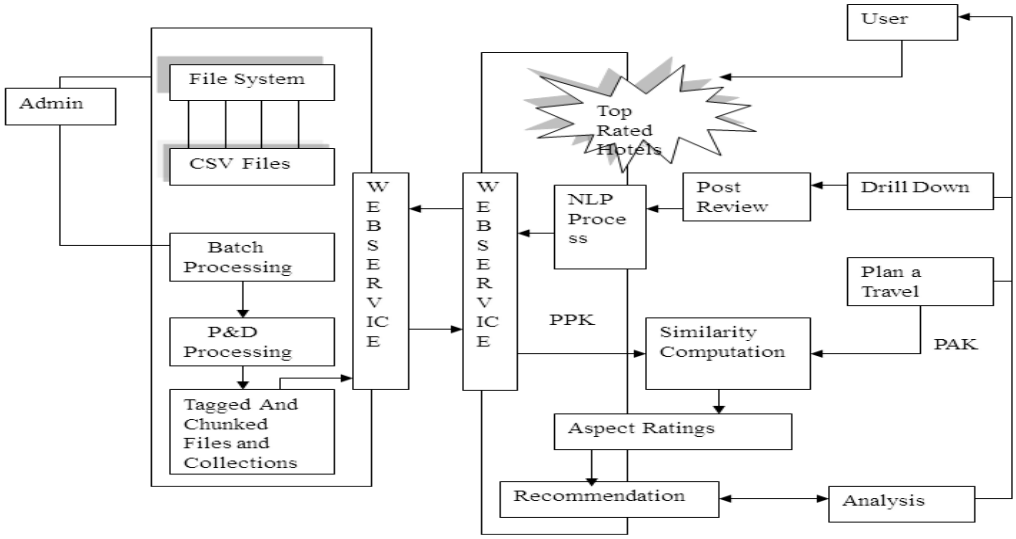
International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol. 4, Issue 2, February 2016

and efficiency of our recommendation method in "Big Data" environment, we implement it in a MapReduce framework on Hadoop by splitting the proposed algorithm into multiple MapReduce phases.

V. ARCHITECTURE DIAGRAM



Data is retrieved from huge amount of data. Comma separated value files were read and manipulated using java API. Traditional views of service recommendation system that show top-k result are displayed. All service rating and review of all hotels for all countries in parallel and distributed manner as batch job. The master job is split into "n" number of small job. Preference of active used and previous user are formalised into their corresponding preference keyword set respectively. Active user can give his/her preference about candidate service by selecting a keyword from the keyword candidate list. Active user select the importance degree of the keyword as "1" represented the general "3" represent importance "5" represent very importance. NLP is implemented to analyse the review of the previous user. NLP process comprises tokenizing a sentence or a word, part of speech tagging, extraction of nouns and verb, synonyms, retrieval and spell check of extraction keyword using wordnet dictionary. Valance and arousal will be implemented for calculating rating of aspect of the hotel. The big data manipulated from comma separated value through our own java API enforces developer friendly access.

MODULES:

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

1. Big Data and Environment

Huge Collection of data is retrieved from open source datasets that are publicly available from major Travel Recommendation Applications. Big Data Schemas were analysed and a Working Rule of the Schema is determined. The CSV(Comma separated values) files were read and manipulated using Java API that itself developed by us which is developer friendly ,light weighted and easily modifiable.

2. Batching and Pre-process

The Traditional View of Service Recommender Systems that shows Top-K Results are displayed with Paginations with which a user can navigate Back and Forth of the Result sets. All Services Ratings and Reviews of Each Hotels are listed. POS(Parts of Speech) Tagger and Chucker Process are done on each and every review of all

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

hotels for all countries in a Parallel and Distributed Manner as Batch jobs. The Master Job is Split up into 'n' no of small Batch jobs based on the slave machines Connected with the Master. POS Tagger tags each words of a review with its tags and the Chunked Process will take POS tagged output as input for Groping the Words based on meaning of the Review.

3. Digging in Big Data & Service Recommender Application

The CSV Files in distributed Systems are invoked through Web Service Running in the Server Machine of the Host Process through a Web Service Client Process in the Recommendation System. The data that Retrieved to the Recommendation Systems are provided with a clean GUI and can be queried on Demand. Each and Every process on the Recommendation Application invokes Web Service which uses light weighted traversal of data using XML. The Users can Review each hotel and can post comments also. The Reviews gets updated to the CSV Files as it get retrieved.

A user can plan or schedule a travel highlighting his requirements in a detailed way that shows the preference keywords set of the active user. A Domain Thesaurus is built depending on the Keyword Candidate List and Candidate Services List. The Domain Thesaurus can be Updated Regularly to get accurate Results of the Recommendation System.

4. MapReduce and Hadoop

(1) Capture user preferences by a keyword-aware approach:

In this step, the preferences of active users and previous users are formalized into their corresponding preference keyword sets respectively. In this paper, an active user refers to a current user needs recommendation.

a) Preferences of an active user.

An active user can give his/her preferences about candidate services by selecting keywords from a keyword-candidate list, which reflect the quality criteria of the services he/she is concerned about. Besides, the active user should also select the importance degree of the keywords. The importance degree of the keywords as "1" represents the general, "3" represents important and "5" represents very important.

b) Preferences of previous users.

The preferences of a previous user for a candidate service are extracted from his/her reviews for the service according to the keyword-candidate list and domain thesaurus. And a review of the previous user will be formalized into the preference key-word set of User.

(2)The keyword extraction process is described as follows:

a) Pre-process:

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

b) Keyword extraction:

In this phase, each review will be transformed into a corresponding keyword set according to the keyword-candidate list and domain thesaurus. If the review contains a word in the domain thesaurus, then the corresponding keyword should be extracted into the preference keyword set of the user Known as (PPK).

(3) Similarity computation:

The Third step is to user(PAK) based on the similarity of their preferencesidentify the reviews of previous users(PPK) who have similar tastes to an active user by finding neighbourhoods' of the active. Before similarity computation, the reviews unrelated to the active user's preferences will be filtered out by the intersection concept in set theory. If the intersection of the preference keyword sets of the active user and a previous user is an empty set, then the preference keyword set of the previous user will be filtered out.

5. KASR and Analysis:

The Chunked Reviews of the Similar User List is retrieved and the Keywords corresponding to the User is analysed 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

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

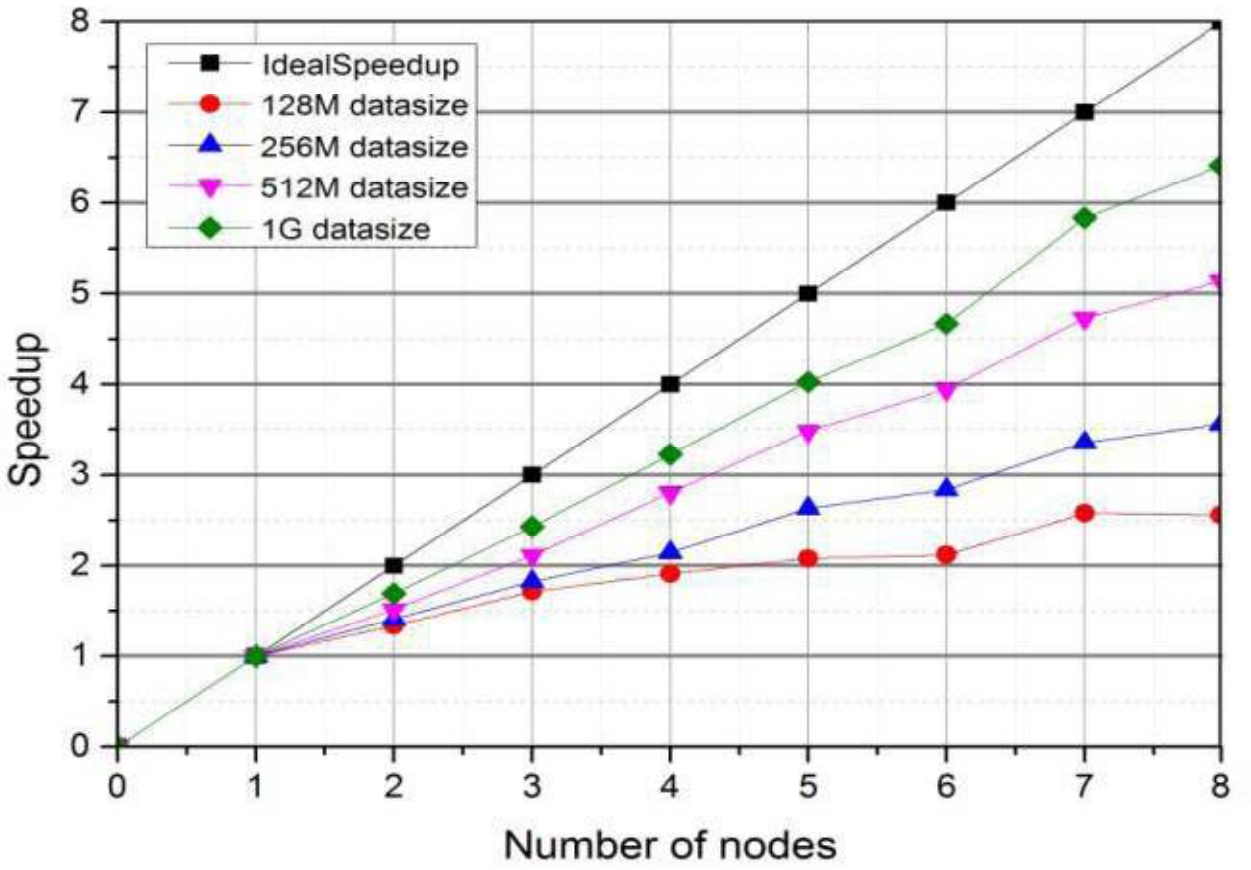
Algorithm to have the Most appropriate personalized Recommendation for the User. The Results will be analysed with Graphical Views so as to understand easier. The Natural Language Processing is implemented to analyse the reviews of the previous user. The NLP Process Comprises Tokenizing a Sentence or a word, POS (Parts of Speech) Tagging, Extraction of Nouns and Verbs, Synonym Retrieval and Spell Check of Extracted 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.

VI. IMPLEMENTATION

To improve the scalability and efficiency of KASR in "Big Data" environment, we implement it in a MapReduce framework on Hadoop platform. Then the user login to process through the authorised user name and password. After login to process the authentication process begins. The pre-processor is been done. After pre-processor output is been taggered.

VII. EXPERIMENT EVALUATION

Two groups of experiments conducted to evaluate the accuracy and scalability of KASR. In the first one, we compare KASR with UPCC and IPCC in MAE, MAP and DCG to evaluate the accuracy of KASR.



VIII. CONCLUSION AND FUTURE WORK

In this paper keyword is extracted using natural language processing.Identify the reviews of previous users(PPK) who have similar tastes to an active user by finding neighbourhoods' of the active. Before similarity computation, the

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

reviews unrelated to the active user's preferences will be filtered out by the intersection concept in set theory.The NLP Process Comprises Tokenizing a Sentence or a word, POS (Parts of Speech) Tagging, Extraction of Nouns and Verbs, Synonym Retrieval and Spell Check of Extracted 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.Moreover, to improve the scalability and efficiency of KASR in "Big Data" environment, we have implemented it on a MapReduce framework in Hadoop platform. Finally, the experimental results demonstrate that KASR significantly improves the accuracy and scalability of service recommender systems over existing approaches.

REFERENCES

[1] Faustino sánchez, maríaalduán, federicoálvarez, *member* "recommender system for sport videos based on user audiovisual consumption" ieee transactions on multimedia, vol. 14, no. 6, DECEMBER 2012
[2] Xiwang Yang, Yang Guo, and Yong Liu, "Bayesian-Inference-Based Recommendation
in Online Social Networks" IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 24, NO. 4, APRIL 2013
[3] Yan-Ying Chen, An-Jung Cheng, and Winston H. Hsu "Travel Recommendation by Mining People
Attributes and Travel Group Types From Community-Contributed Photos "IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 15, NO. 6, OCTOBER 2013
[4] KatrienVerbert, Nikos Manouselis, Context-Aware Recommender Systems for Learning: A Survey and Future Challenges IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES, VOL. 5, NO. 4, OCTOBER-DECEMBER 2012
[5] ShunmeiMeng, Wanchun Dou, Xuyun Zhang "KASR:AKeyword-AwareService Recommendation
Method on MapReduce for Big Data Applications" IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS, VOL. 25, NO. 12, DECEMBER 2014.
[6]A. Ansari, S. Essegaier, and R. Kohli, "Internet Recommendations Systems," J. Marketing Research, vol. 37, no. 3, pp. 363-375, Aug. 2000, doi: 10.1509/jmkr.37.3.363.18779.
[7] Y. Zhang and J. Koren, "Efficient Bayesian Hierarchical User Modeling for Recommendation System," Proc. 30th Ann.In'l ACM SIGIR Conf. Research and Development in Information Retrieval,
pp. 47-54, 2007, doi:10.1145/1277741.1277752.
[8] A. Narayanan and V. Shmatikov, "Robust de-Anonymization of Large Sparse Data Sets," Proc. IEEE Symp.Security and Privacy, 2008.