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. 3, Issue 8, August 2015

Improved Keyword Aware Service Recommendation System for Big Data Applications

Shakhy.P.S[1], Swapna.H[2]

M.Tech, Dept. of CSE, Marian Engineering College, Trivandrum, India

Assistant Professor, Dept. of CSE, Marian Engineering College, Trivandrum, India

ABSTRACT: Service recommender systems are valuable tools for providing appropriate recommendations to users. In the last decade the rapid growth of the number of customers, services and other online information yields service recommender systems in Big Data environment .In keyword aware service recommendation system ,keywords are used to indicate both of user's preferences and quality of candidate services .A user based collaborative filtering algorithm is adopted to generate appropriate recommendations .In CF based systems ,users receive recommendations based on people who have similar tastes and preferences .The preference of previous users are extracted from their reviews and formalized into a keyword set .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. The previous users who have similar tastes to an active user are found based on similarity of their preferences and recommendations are provided to the active user .But the system only depends on explicit user feedback, and thus is intrusive. It only considers user reviews and does not consider the temporal information about locations of the services .The limitations of the current recommendation system are reduced by possible extensions that provides better recommendation capabilities and usability of the system .These extensions include incorporation of temporal analysis into the recommendation process. This improves the accuracy of the predictions of recommendations.

KEYWORDS: Recommendation,Keyword,Big data,Collaborative Filtering

I. INTRODUCTION

Recommender systems are software applications that attempt to reduce information overload by recommending items of interest to end users based on their preferences .It can be defined as a system that produces individualized recommendations as output or has the effect of guiding the user in a personalized way to interesting or useful services in a large space of possible options. The first recommender system, Tapestry, was designed to recommend documents from newsgroups[11]. Nowadays, the trend "everything as a service" has been creating a Big Services. The big data comprises high volume, high velocity, and high variety information assets , which are difficult to gather, store, and process by using the available technologies.A keyword-aware service recommendation method, named KASR, uses a user-based Collaborative Filtering algorithm. In KASR, keywords extracted from reviews of previous users are used to indicate their preferences and to generate new recommendations.

II. RELATED WORK

Current recommendation methods can be usually 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 other users with similar tastes preferred in the past. Hybrid approaches combine content-based and CF methods in several different ways.

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. 3, Issue 8, August 2015

A. Content-based filtering

Content-based recommender systems work with profiles of users that are created at the beginning. A profile has information about a user and his taste. Taste is based on how the user rated items. Generally, when creating a profile, recommender systems make a survey, to get initial information about a user in order to avoid the new-user problem. The content-based approach to recommendation has its roots in information retrieval and information filtering systems. Because of the significant and early advancements made by the information retrieval and filtering communities and because of the importance of several text-based applications, many current content-based systems focus on recommending items or services containing textual information, such as documents, Web sites (URLs), and Usenet news messages. The profiling information can be elicited from users explicitly, e.g., through questionnaires, or implicitly—learned from their transactional behavior over time[2].

B. Collaborative filtering

Collaborative filtering became one of the most researched techniques of recommender The idea of collaborative filtering is finding users in a community that share appreciations . If two users have same or almost same rated items in common, then they have similar tastes. Such users build a group or a so called neighborhood. A user gets recommendations to those items that he/she hasn't rated before, but that were already positively rated by users in his/her neighborhood. The taste is considered to be constant or at least change slowly[2].

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 the user-based approach the items that were already rated by the user before, play an important role in searching a group that shares appreciations with him.In item-based systems, the predicted rating depends on the ratings of other similar items by the same user.

C. Hybrid recommendation approaches

For better results some recommender systems combine different techniques of collaborative approaches and content based approaches. Using hybrid approaches we can avoid some limitations and problems of pure recommender systems, like the cold-start problem.The recommendation methods described above have performed well in several applications. However, they have certain limitations, described in the TABLE 1.Moreover, in order to provide better recommendations and to be able to use recommender systems in more complex types of applications most of the methods would need significant extensions.

Recommendation Approach	Technique Used	Disadvantages
Content based	Information Retrieval and filtering(TF-IDF)	Limited content analysis ,new user problem, Overspecialization
Collaborative	Nearest neighbor	New user and new item problem ,Sparsity
Hybrid	Combine content based and collaborative	Sometimes quality and accuracy gets affected

TABLE1: Comparison of existing recommendation approaches

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. 3, Issue 8, August 2015

D. Modern Recommendation Approaches

Context-aware approaches

Context is the information about the environment of a user and the details of situation he/she is in. Such details may play much more significant role in recommendations than ratings of items, as the ratings alone don't have detailed information about under which circumstances they were given by users. Some recommendations can be more suitable to a user in evening and doesn't match his preferences in the morning at all and he/she would like to do one thing when its cold and completely another when its hot outside. The recommender systems that pay attention and utilize such information in giving recommendations are called context-aware recommender systems. As opposed to content information that is saved in profiles, context changes dynamically and often saved just permanently, as it is more likely to lose its currency after a certain period of time. That is why it is very important to periodically refresh the information. Context-aware recommender systems became much attention, as they noticeably increased the quality of recommendations and the approaches became more specific to use in certain areas[2].

Semantic based approaches

Most of the descriptions of items, users in recommender systems and the rest of the web are presented in the web in a textual form. Using tags and keywords without any semantic meanings doesn't improve the accuracy of recommendations in all cases, as some keywords may be homonyms. That is why understanding and structuring of text is a very significant part recommendation. Traditional text mining approaches that is based on lexical and syntactical analysis show descriptions that can be understood by a user but not a computer or a recommender system. That was a reason of creating new text mining techniques that were based on semantic analysis[8].

Cross-domain based approaches

Finding similar users and building an accurate neighborhood is an important part of recommending process of collaborative recommender systems. Similarities of two users are discovered based on their appreciations of items. But similar appreciations in one domain don't surely mean that in another domain valuations are similar as well. Standard recommender systems based on collaborative filtering compare users without splitting items in different domains. In cross-domain systems similarities of users computed domain-dependent. An engine creates local neighborhoods for each user according to domains. Then, computed similarity values and finite set of nearest-neighbors are sent for overall similarities computation. Recommender system determines the overall similarity, creates overall neighborhoods and makes predictions and recommendations[8].

Peer-to-Peer approaches

The recommender systems with P2P approaches are decentralized. Each peer can relate itself to a group of other peers with same interest and get recommendations from the users of that group. Recommendations can also be given based on the history of a peer. Decentralization of recommender system can solve the scalability problem.

Cross-lingual approaches

The recommender system based on cross-lingual approach lets the users receive recommendations to the items that have descriptions in languages they don't speak and understand .The main idea of cross-lingual based approach is to map both text and keywords in different languages into a single feature space, that is to say a probability distribution over latent topics. From the descriptions of items the system parses keywords than translates them in one defined language using dictionaries. After that, using collaborative or other filtering, the system gives recommendations to users .Cross-lingual recommender systems break the language barrier and gives opportunities to look for items, information, papers or books in other languages[8].

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. 3, Issue 8, August 2015

Keyword based approach

In this method, 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(Keyword Aware Service Recommendation System) 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 .Two data structures use din this method are , “keyword-candidate list” and “specialized domain thesaurus”, introduced to help obtain users' preferences The keyword-candidate list is a set of keywords about users’ preferences and multi-criteria of the candidate services, which can be denoted as K={k1,k2,….,kn} , *n* is the number of the key-words in the keyword-candidate list .The preferences of previous users will be extracted from their reviews for candidate services and formalized into a keyword set. A domain thesaurus is a reference work of the keyword-candidate list that lists words grouped together according to the similarity of key-word meaning, including related and contrasting words and antonyms. Often, domain thesauruses are updated regularly to ensure the timeliness of the words.

The main steps of KASR includes capturing preferences of an active user ,computing the similarity of active user preference to a previous user preference ,and recommending personalized services to the active user by keyword aware approach .The preferences of active users and previous users are formalized into their corresponding preference keyword sets respectively .An active user refers to a current user who needs recommendation. 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. The preference keyword set of the active user can be denoted as APK={ak1,ak2,..,ak*l*} , where *l* is the number of selected keywords .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. A review of the previous user will be formalized into the preference key-word set of him/her, which can be denoted as PPK={pk1,pk2,…,pk*h*},where *h* is the number of extracted keywords .Based on the similarity of the active user and previous users, further filtering will be conducted. Once the set of most similar users are found, the personalized ratings of each candidate service for the active user can be calculated. Finally, a personalized service recommendation list will be presented to the user and the service(s) with the highest rating(s) will be recommended to him/her[1].

III. PROPOSED METHOD

The keyword aware service recommendation system works solely on the basis of explicit user feedbacks and ratings. Thus the system is intrusive in nature. Therefore, the problem of minimizing intrusiveness while maintaining certain levels of accuracy of recommendations needs to be addressed. The existing system only recommends services to users and does not take into consideration about the preference of location of the particular service. Users of these online social services not only encounter the problem of information overload, but also have mutable interests which change fast along with the social information streams. These features pose great challenges to recommender systems, since the purpose of such systems is to provide suitable recommendations that match users' real-time interests, which is quite difficult among massive candidates and fast changing user preferences.To improve the system we can include temporal data. Various limitations of the current recommendation methods discussed in the chapter 2 can be reduced by possible extensions that can provide better recommendation capabilities. These extensions include, the improved modeling of users and service providers by the incorporation of temporal analysis into the recommendation process. The user can also be provided with the choice of selecting a new location if the temporal analysis of the location is not satisfactory or else user can proceed with the same

selection of location .This further improves the usability of the system. Implementation on a Big data platform like MongoDB significantly improves the accuracy and scalability of service recommender systems over existing approaches and performs better with larger datasets.

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. 3, Issue 8, August 2015

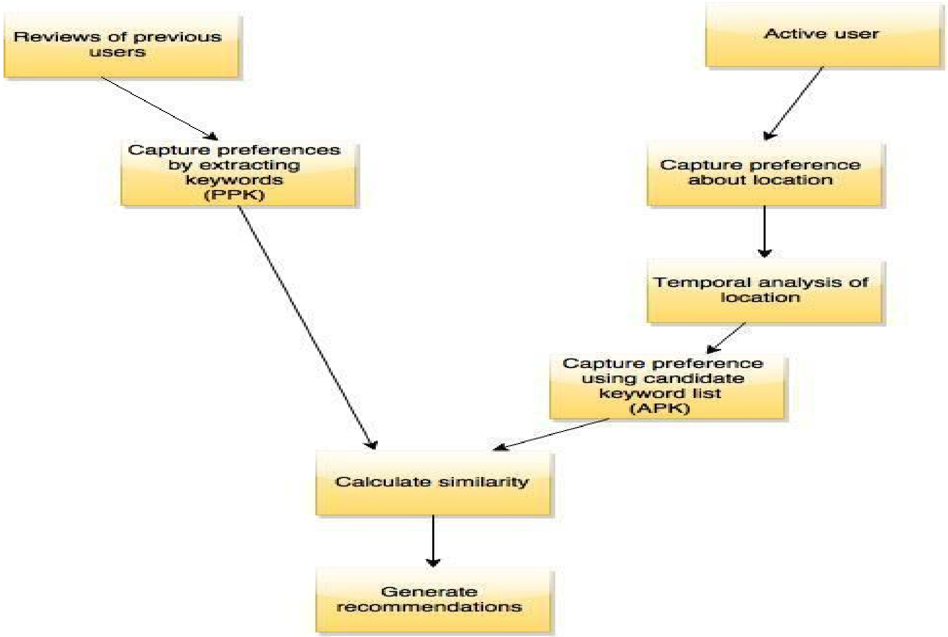

FIG 1:Improved Kasr System With Temporal Analysis

For the hotel recommender system, the reviews of previous users are collected and keywords are extracted to form the PPK(Previous users Preference Keyword set).For an active user, before presenting the keyword list , the preference of the location and the time (in months) of the year during which the user needs the recommendation are collected. An analysis of the location is done using previously collected data to determine whether the location has good, average or poor rating during the selected time of the year .If the user is not concerned about the rating of the location and only needs the recommendation of hotels ,user can proceed with the current selection or else the user can use the analysis result and select another location .The temporal analysis is done using the J48 classifier with the training dataset.

Similarity Computation

There are different algorithms of measuring similarities among items or services in data base and those in users profile The simplest and most common example of a distance measure is the Euclidean distance .Other methods include , the Minkowski Distance, Mahalanob distance,Cosine based method, Pearson correlation, the Simple Matching coefficient, the Jaccard coefficient,etc..The Keyword based approach uses Approximate similarity computation and Exact similarity computation[1].

Approximate similarity computation

A frequently used method for comparing the similarity and diversity of sample sets, Jaccard coefficient, is applied in the approximate similarity computation. Jaccard coefficient is measurement of asymmetric information on binary (and non-binary) variables, and it is useful when negative values give no information. The similarity between the preferences of the active user and a previous user based on Jaccard coefficient is described as follows

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

where *APK* is the preference keyword set of the active user, *PPK* is the preference keyword set of a previous 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. 3, Issue 8, August 2015

Exact similarity computation

A cosine-based approach is applied in the exact similarity computation, which is similar to the Vector Space Model (VSM) in information retrieval. The preference keyword sets of the active user and previous users will be transformed into *n*-dimensional weight vectors respectively. The weight vector of the preference keyword set of a previous user can be decided by the term frequency/inverse document frequency (TF-IDF) measure[1].

IV.TOOLS AND TECHNIQUES

Mahout

Apache Mahout learning library is written in Java that is designed to be scalable, i.e. run over very large data sets. It achieves this by ensuring that most of its algorithms are parallelizable (map-reduce paradigm on Hadoop. The Mahout project was started by people involved in the Apache Lucene project with an active interest in machine learning and a desire for robust, well-documented, scalable implementations of common machine-learning algorithms for clustering and categorization.Mahout is designed mainly for the following use cases. Recommendation mining takes user's behavior and from that tries to find items users might like.

MongoDB

MongoDB is a document database that provides high performance, high availability, and easy scalability. A MongoDB deployment hosts a number of databases. A database holds a set of collections. A collection holds a set of documents. A document is a set of key-value pairs. Documents have dynamic schema. Dynamic schema means that documents in the same collection do not need to have the same set of fields or structure, and common fields in a collection’s documents may hold different types of data.

V. SIMULATION RESULTS

To evaluate the performance of the system in accuracy , it can be compared with the existing keyword aware service recommender system. The metric used to evaluate the accuracy is MAE, Mean Absolute Error. MAE is a statistical accuracy metric often used in CF methods to measure the prediction quality. It is defined as the average absolute deviation between a predicted rating and the real rating. The lower the MAE presents the more accurate predictions.

Experiment Setup and Datasets

The front end is a web application for hotel recommendation which collects user reviews and ratings .It captures user preferences using keywords and importance degree.it also captures the location preference and time (in months) for which recommendation is needed .The datasets used are Tripadvisor datasets. All approaches are implemented in java. The

backend used is MongoDB which is a document database that provides high performance, high availability, and easy scalability. It is a NoSQL database used for data applications.

Experimental Evaluation

Experiments are conducted to evaluate the accuracy of the system. The figure shows the comparison of MAE measure .The MAE for the existing was 0.4753 which is around 0.5.The temporal analysis of the proposed system has 0.4421 as the MAE measure for the same conditions. The lower the MAE presents the more accurate predictions. Thus the accuracy of the system is increased.

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. 3, Issue 8, August 2015

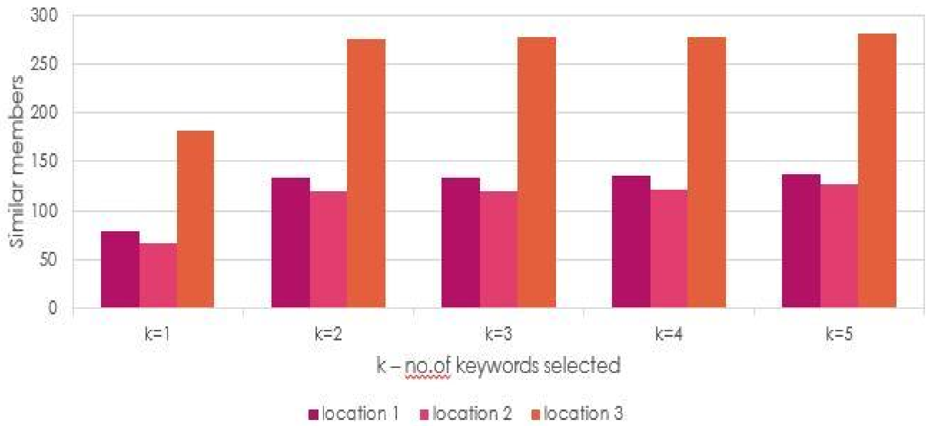

FIG 2.Analysis of the system

The figure shows the analysis of the proposed system in terms of number of similar members generated for the number of keywords selected. Analysis of proposed system is done with the help of WEKA tool. Three locations were selected for the analysis. Along the X- axis k represents the number of keywords selected and Y- axis represents the number of similar members generated. For k= 1, the similar members generated are few as compared to k=2 , for the three locations. But from k=2 onwards there is not much increase in the number of similar members generated .For example consider location 1, when k=1 the number

of similar members generated is 79, when k=2 and k=3, it is 134, for k=4 it is 135 and for k=5 it is 137.So even though the number of keywords selected increase , it does not affect the performance of the system much.

VI.CONCLUSION

An improved Keyword Aware Service Recommendation System has been proposed in which apart from getting personalized recommendation about services, the user also gets an analysis of the location during a particular time of the year. The proposed method eliminate the intrusiveness of Keyword Aware Recommendation System by incorporating temporal constraints. Finally, the experimental results demonstrate that the proposed system significantly improves the accuracy and usability of service recommender systems over existing approaches.

REFERENCES

1. Shunmei Meng, Wanchun Dou, Xuyun Zhang, Jinjun Chen, “ KASR: A Keyword Aware Service Recommendation Method on MapReduce for Big Data Applications” IEEE Transactions on Parallel and Distributed Systems, TPDS-2013-12-1141.
2. Osman Khalid Muhammed Usman Shahid Khan, Samee U Khan , Albert Y Zomaya “Omnisuggest: A Ubiquitous Cloud Based Context Aware Recommendation System for Mobile Social Networks ” IEEE Transactions on Services Computing ,2013.
3. Recommender Systems , Daniel Rodriguez, University of Alcala ,Windsor ,August 5,2013
4. K. Lakiotaki, N. F. Matsatsinis, and A. Tsoukis ,“Multi-Criteria User Modeling in Recommender Systems”, IEEE Intelligent Systems, Vol. 26, No. 2, 64-76, 2011.
5. Sang Hyun Choi, Young-Seon Jeong, and Myong K. Jeong “A Hybrid Recommendation Method with Reduced Data for Large-Scale Application”
6. IEEE Transactions on systems, man, and cybernetics—Part C: Applications and Reviews, Vol. 40, No. 5, September 2010
7. Neal Lathiax, Stephen Hailesx, Licia Caprax, Xavier Amatriainy” Temporal Diversity in Recommender Systems”, ACM 978-1-60558-896-4/10/07,July 2010
8. Yehuda Koren “Collaborative Filtering with Temporal Dynamics”, ACM 978-1-60558-495-9/09/06, July 2009
9. Gediminas Adomavicius, and Alexander Tuzhilin “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, June 2005.
10. J. Dean, and S. Ghemawat, “MapReduce: Simplified data processing on large clusters,” Communications of the ACM, Vol. 51, No.1, pp. 107-113, 2005.
11. G. Linden, B. Smith, and J. York, “Amazon.com Recommendations: Item-to-Item Collaborative Filtering,” IEEE Internet Computing, Vol. 7, No.1, 76-80, 2003.
12. Daniar Asanov “Algorithms and Methods in Recommender Systems” Berlin Institute of Technology Berlin, Germany.
13. Jayant Shekhar ,”Building Recommendation Platforms with Hadoop”, O’Reilly Strata Conference Making Data Work.
14. Xavier Amatriain, Alejandro Jaimes, Nuria Oliver, and Josep M. Pujol ,“Recommender Systems Handbook” Springer Science+Business Media, LLC 2011,pp 39-67.