Implementation of Collaborative Filtering Approach in Preference Aware Service Recommendation

Ms. Sneha.V
Student (M Tech) of Dept. of Computer science & Engineering
Sai Vidya Institute of Technology, Bangalore
Bangalore, 560064, India
snehayadav.sn@gmail.com

Mr. S Deepak Raj
Associate professor, Dept. of Computer science & Engineering
Sai Vidya Institute of Technology, Bangalore,
Bangalore,560064 ,India
Deepak.raj@saividya.ac.in

Abstract- Service recommendations are shown as remarkable tools for providing recommendations to users in an appropriate way. In the last few years, the number of customers, online information and services has grown very rapidly, resulting in the big data analysis problem for service recommendation system. Consequently, there is scalability and inefficiency problems associated with the traditional service recommendation system which suffers in processing or analyzing large-scale data. Moreover, most of available service recommendation system gives the same rankings and ratings of services to different users without any considerations of many user's preferences, and hence it fails to reach user's personalized requirements. In this paper, we have proposed a Preference-Aware Service Recommendation method, to overcome the above challenges. It aims at recommending the most appropriate and preferred services to the users and provide a personalized service recommendation list in an effective way. Here, users' preferences are captured as keywords, and a user-based Collaborative filtering approach is adopted to create appropriate recommendations. A widely-adopted distributed computing platform, Hadoop is used for the implementation of this approach, which improves its efficiency and scalability in big data environment, using the MapReduce parallel processing method.

Keywords-*recommender system, preference, keyword, service, efficiency, scalability, collaborative filtering, big data, Map-Reduce, Hadoop.*

*****

I. INTRODUCTION

In the last few years, the number of customers, online information and services has grown very rapidly, resulting in the big data analysis problem for service recommendation system. This has lead to scalability and inefficiency problems which is associated with the traditional service recommendation system which suffers in processing or analyzing large-scale data.

A. Big Data

Big data usually contains sets of data with sizes which are beyond the ability of software tools that are commonly used to capture, curate, manage, and process data within a tolerable time limit [1]. Big data is a set of technologies and techniques which requires new forms of integration to uncover large hidden values from huge datasets that are of a massive scale, complex and diverse.

The challenges with big data include capture, analysis, sharing, search, transfer, storage, privacy violations and visualization. Big data is difficult to work with using most relational database management systems and desktop statistics.

B. Cloud Computing

Cloud computing focuses on computations over a scalable network of nodes and sharing the data. The important goal of cloud computing is to share the resources, such as platform, infrastructure, business process and software. There are many tools available for cloud computing, such as Mahout (http:// mahout.apache.org/), Hadoop (http://hadoop.apache.org/), the Dynamo of Amazon.com, MapReduce of Google.

Nowadays, a Big Services era has been created by the trend "everything as a service" due to the foundational architecture of services computing. Thus the efficiency and scalability of service recommendation methods can be improved in the Big Data environment using the above mentioned tools of cloud computing.

C. Hadoop

The Hadoop, a distributed computing platform is a batch processing system for a groups of nodes that provides the most Big Data activities of analytics because it groups two sets of functionality mostly needed to deal with unstructured large datasets like, Map-Reduce modeling and distributed file system (DFS) It is a project written in Java by the Apache Software Foundation to support data intensive distributed applications. Hadoop enables applications to work with petabytes of data and thousands of nodes. The biggest contributor of Hadoop has been the search g engine Yahoo, where it is extensively used widely in the business platform.

The rest of the paper is organised as follows: section II explains basic knowledge about recommendation systems and existing recommender systems, section III about the implementation and section IV conclusion.

II. RELATED WORK

Recommendation system was developed as an independent research field in the mid 1990s when the focus of recommendation problems started on rating models.

The main objective of the recommendation system is to predict the active user's ratings for the items which he has not yet expressed his rating on and then find a list of n recommended items for the active user [2].

The present recommendation methods can be usually classified into three main categories:, collaborative, content-based and hybrid recommendation approaches. Content-based methods recommend similar services to the users which were preferred in the past by the user. Collaborative filtering (CF) approach is adopted which recommends services to the user that users with similar tastes preferred in the past. In CF based systems, recommendations are received by the users based on people who have similar preferences and tastes.

A. User-Item Predictive Model for Collaborative Filtering Recommendation

The approach first determines similarities between the items, and subsequently identifies the confidence of the items, indicating the relevance of prior predictions.

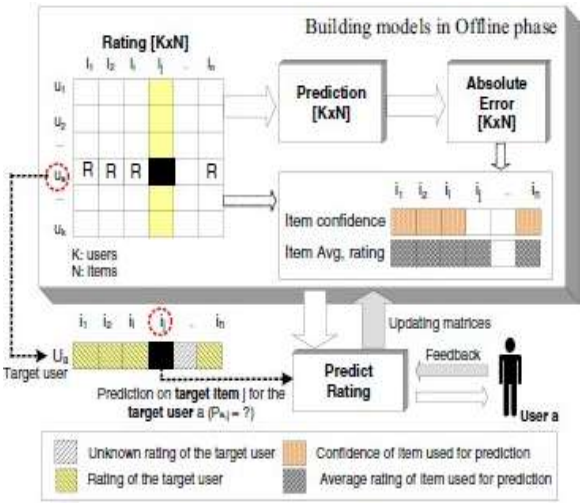

Figure 1. Applying the model to a collaborative filtering recommendation

The major advantage of this approach is that it supports updating of the model instantaneously, even when users present explicit feedback. The experimental results demonstrate that the prediction quality improves with growth of items used for the prediction generation. Moreover, this approach offers reasonably good quality although the prediction quality is slightly worse than the best quality of a user-based CF. However, there still remains a defect that this prediction strategy depends on the error threshold.

B. Prediction by hierarchy process

This research provides a new recommendation method related to high-tech products based on consumer's preference with less complexity and more effectiveness. It aims to make the new products equivalent to the old ones during the computing process therefore improve the effectiveness of new product recommendation. SOM Model utilizes the supervised learning here to train SOM rating "OLD" instances' criteria by classifying instances and make predictions of criteria ratings for "NEW" ones. The SOM is employed to deal with the internal part. It is utilized to explore the links between products and forecast the rating for the new products therefore make them equivalent to old ones inside the computing of recommendation system [4].

The AHP is employed to collect the user preference and, which is more importantly, connect the internal and the external parts together into computing. This method is capable of reducing the buying complexity of high-tech products and improving the effectiveness of new product recommendation.

C. "Recommender System for Sport Videos Based on User Audiovisual Consumption

This paper describes a recommender system for transmitting and broadcasting over the Internet and sport videos in the context of the large-scale events, which were tested for the Olympic Games. The recommender does not depend on the number of users and is based on audiovisual consumption running on the client side only. This avoids privacy problems of central server approaches, computation and the concurrence in scenarios where large numbers of users are present such as the Olympic Games [5]. This system is made available to take advantage of the information which is available in the videos and which is used along with the user's implicit information and the designing of their consumption of audiovisual content. Thus the system is made transparent to the user, who does not require to take any specific action.

III. PROPOSED APPROACH

A. Preference Aware service recommendation system

In this method, keywords are used to indicate the quality of candidate services and user's preferences. Two data structures, "keyword-candidate list" which is a set of keywords about multi-criteria of the candidate services and preferences of users. And A domain thesaurus is created as

keyword-candidate list's reference work that lists words that are grouped together as per the similarity of keyword.

In the first step the preferences of previous users and active users are formalized into their corresponding preference keyword sets respectively. An active user selects keywords from a keyword-candidate list to give preferences about candidate services, as APK= {ak1, ak2,....akn}. Similarly previous user's preferences for a candidate service are extracted from their reviews according to the domain thesaurus.

The second step is the identification of previous user's reviews of those who have similar tastes compared to an active user by finding neighbourhoods of the active user based on the similarity of their preferences.

Jaccard coefficient is a measurement of asymmetric information on binary (and non-binary) variables and used to find a similarity between the PPK and APK. Exact similarity computation is calculated using cosine-based approach which is similar to the(VSM) Vector Space Model used for information retrieval. The preference keyword sets of the previous users and active user will be transformed as weight vectors of *n*-dimensional respectively, namely as preference weight vector.

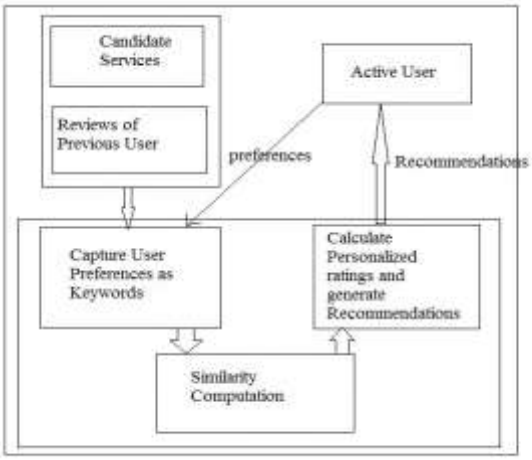

Figure 2. The Architecture of the proposed service recommendation system

The preference keyword set's weight vector of a previous user is considered by *the term frequency/inverse document frequency* (*TF-IDF*) measure, which is one of the well-known measures in Information Retrieval for specifying the weight of the keywords.

In the last step, further filtering will be conducted based on the similarity of the previous users and active user. Once the most similar set of users are found, each candidate service's personalized ratings can be calculated for the active user. And Finally, a personalized recommendation list of services will be presented to the active user and the services with the highest ratings will be recommended to him/her.

IV. CONCLUSION

In this proposed system, keywords shows user preferences and collaborative filtering approach is used to provide appropriate recommendation lists of services to the user. For faster calculations and scalability, it is implemented on Hadoop Map-Reduce framework which gives better scalability and accuracy.

REFERENCES

[1] Manyika, M. Chui, B. Brown, et al, "Big Data: The next frontier for innovation, competition, and productivity," 2011.

[2] K. Abhishek, S. Kulkarni, V. Kumar, N. Archana and P. Kumar, "A Review on Personalized Information Recommendation System Using Collaborative Filtering," *International Journal of Computer Science and Information Technologies (IJCSIT)*, vol. 2, no. 3, pp. 1272-1278, 2011.

[3] Heung-Nam Kim1, Ae-Ttie Ji1, Cheol Yeon1, and Geun-Sik Jo2,"] A User-Item Predictive Model for Collaborative Filtering Recommendation" 2008.

[4] Xuecong ZENG, Longsheng CAI and Tomohiro MURATA, "Preference based Recommendation Method by a hierarchy process and a Case Study of Smartphone", 2011.

[5] M. Bjelica, "Towards TV Recommender System Experiments with User Modeling," IEEE Transactions on Consumer Electronics, Vol. 56, No.3, pp. 1763-1769, 2010.