A Survey on Smart Service Recommendation System by Applying Map Reduce Techniques

Pallavi R. Desai[1], B. A. Tidke[2]

[1]Department of Computer Network, Flora Institute of Technology, Khopi, Pune, India

[2]Professor, Dept. of Computer Network, Flora Institute of Technology, Khopi, Pune, India

Abstract—*In the era of internet amount of data grown beyond the capacity of storing and processing, This known as Big Data. When Users deals with Big Data it face varies difficulty at the time of needful data extraction. Their for we purpose Smart Service Recommender system is providing appropriate recommendations to users as per their interest. In the past few years, the amount of online web data has increases explosively, yielding the big data processing and analysis problem for recommender systems. Consequently, most of the traditional service recommender systems frequently suffer from scalability and inefficiency problems when processing or analysing such large volume data. Moreover, an existing service recommender system present the same ratings and rankings of items to different users without considering varied users' preferences, and therefore fails to meet users personalized requirements. This project proposes a Smart Service Recommendation system, to address the above challenges. It aims at presenting a personalized recommendation list and recommending the most appropriate items to the users effectively. Specifically, weights of keywords are used to indicate users' preferences, and a user-based Collaborative Filtering algorithm is adopted with Opennlp to generate appropriate recommendations. To improve its scalability and efficiency in big data environment, it is implemented on Hadoop, a widely-adopted distributed computing platform for processes large data using MapReduce parallel processing paradigm. Finally, extensive experiments are conducted on real-world data sets, and results demonstrate that Personalize User-Based Recommendation System significantly improves the accuracy and scalability of service recommender systems over existing approaches.*

Keywords: Big Data, MapReduce, Hadoop, recommender system, preference, keyword

1. Introduction

This chapter is organized as following, introduction to the big data and recommendation analysis which is followed by the Background of proposed model, followed by motivation.

1.1 Background

The amount of structured, semi-structured and unstructured data has been increasing very quickly in our world. These data is nothing but „bid data". Analysis of big data is challenges for companies. To solve these challenges of big data we want to provide hardware and software solutions. With the growing of data challenges may also increases. Every day, people are come with multiple 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. Each of these questions has many alternative solutions. With the growing number of alternative services, and for these types of big data providing an effectively recommending system that users preferred has become an important research issue.

Service recommender systems are subclass of information filtering systems that seek to predict the „rating" and „preferences" that a user would give an item. Recommender systems have been shown as important tools to help users deal with services overload and provide appropriate recommendations to them. Personalized recommendation system is an important trend in the development of information processing technology. With the continuous rapid development and improvement of Internet technology, there has been an explosive growth of information on the Internet.

Although publicly available consumer search engines have become the most effective to search the information on Internet, these tools fail to satisfy all user demands and preferences. Also, if we consider user rating then also customer was unable to get prefect information. Consequently, Keyword aware recommendation system has been developed. Keyword aware service recommendation system provides an automatic function that recommends details by obtaining and analyzing user preferences, predictions based on the analysis and information are made prior to the user launching a search. To achieve the perfect recommendations we apply Opennlp for semantic analysis.

The core value of Keyword aware service recommendation system lies in its recommendation capability. The appropriate use of recommendation system algorithms that improve the accuracy and efficiency of recommendations and algorithms those return results consistent with user interests.

1.2 Motivation

The goal of a recommendation system is to generate meaningful recommendation list to a number of users for select items or products that might interest of them. The traditional algorithms for recommendation system does not give the effective and efficient results when we want to deal with bid data as it has the disadvantages such as the "scalability" and the "accuracy". To overcome these problems of big data it is important to provide advanced filtering algorithm that will solve the above problems and recommend the data efficiently.

Proposed model has provided with the combination of advanced filtering algorithms such as Approximate

Similarity Computation, Exact Similarity Computation, and A basic algorithm of Smart Service that will improve the Recommendation service with accuracy and scalability.

2. Related Work

The authors Shunmei Meng,Wanchun[1] proposes a Keyword-Aware service Recommendation method, to presenting a personalized recommendation list and recommending the most appropriate items to the users effectively. Specifically, keywords are used to indicate users" preferences, and a user-based Collaborative Filtering algorithm is adopted to generate appropriate recommendations. Finally, the experimental results demonstrate that KASR significantly improves the accuracy and scalability of service recommender systems over existing approaches.

The authors, X. Yang, Y. Guo, and Y. Liu [2], propose a Bayesian-inference-based recommendation system is developed for online social networks. They show that the proposed Bayesian-inference-based recommendation is better and comparable than the existing trust-based recommendations. They are also manages the recommendation quality and quantity. The Prior Distribution is specially used to overcome cold start problem and data sparseness problem.

In [3], Adomavicius and Tuzhilin give an overall of the structure of recommender systems and describe the current generation of recommendation methods. They also describe various drawback and limitations of current service recommendation methods, and discuss possible solutions that can improve recommendation capabilities and make recommender systems applicable to an even broader range of applications.

Most existing service recommender systems are only use a single numerical rating to represent a services utility recommendation as a whole [4]. They are not considering the user"s' preferences. In fact, analysing services through multiple criteria, choices and considering users feedback can help to make more accurate and effective recommendations for the users.

The authors Z D Zhao and M. S. Shang of [5] implement a user-based CF algorithm on Hadoop. They solve the scalability and efficiency problem by dividing dataset into different domains. But their method doesn"t have favourable of scalability and efficiency if the amount of data grows beyond the capacity. It proposes a parallel user profiling approach based on folksonomy information and implements a scalable recommender system by using MapReduce and Cascading techniques.

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, solve a problem untouched by traditional recommender systems by dealing with three types of location-based ratings: spatial ratings for non-spatial items, non-spatial ratings for spatial items, and spatial ratings for spatial items. All above techniques can be applied separately or in concert to support the various types of location-based ratings. It is efficient, scalable, and provides better quality recommendations than techniques used in traditional recommender systems.

The author [16] gives detail about the big data. Today the grow of the data is increase rapidly, it consider as a big data. Big data has now reached every sector in the global economy. The use of big data is becoming a key way for leading companies to outperform their peers. Work for the big data is big issue. For this first we have to know that what actually a bid data and what challenges want to face at the time of operating the big data.

In this paper author [18] 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.

3. System Architecture

The project proposes a novel method of Smart service for recommendation system by applying MapReduce technique. In which Keyword-preference List and Domain Clusters are maintained for particular system. Preferences are taken from reviews of previous users. And similarity between users preference are searched out by keyword extraction method and similarity calculations by applying keyword-aware approach. Then the keywords are classified according their weight and weights of reviews of similar users' preferences are then calculated. Then finally, recommendation list of top-k items is generated.

In proposed method, keywords are used to indicate both of users' preferences and the quality of recommendation services. A Smart Service algorithm is adopted to generate appropriate recommendations. Proposed system aims at calculating a personalized rating of each recommendation services for a user, and then presenting a personalized

Paper ID: NOV152687
366

service recommendation list and recommending the most appropriate services to them.
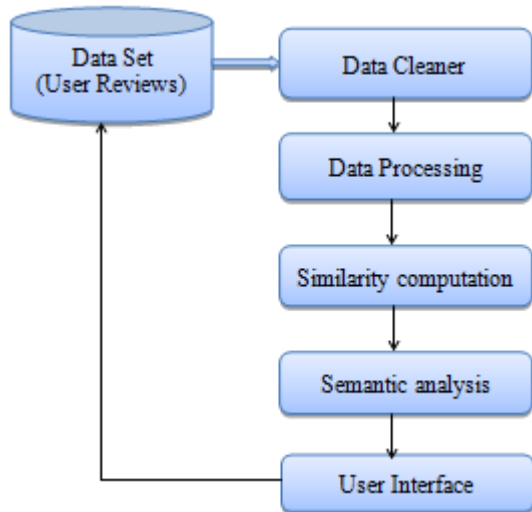

Figure 1: System Architecture of Smart Service Recommendation System

4. Methodology

We propose novel method of personalized user-based recommendation system.

The steps in proposed system are:
- Keyword-preferences List and Domain Clusters are maintained for particular system.
- Preferences are taken from user and similarities of user‟s preferences are examined by keyword extraction method and similarity calculations.
- Then the keywords are classified, and weights of reviews of similar users are calculated.
- Then finally, Generate recommendation list of Top-k items.

A. Keyword- preference List and Domain clusters
In our method, two data structures, "keyword-preference list" and "specialized domain clusters", are introduced to help obtain users' preferences.

Keyword-preference list: The keyword-preference list is a set of keywords about users‟ preferences and multi-criteria of the candidate services[1], which can be denoted as K=n where n is the number of the keywords in the keyword candidate list. An example of a simple keyword preference list of the movies recommendation system is described in Table 1.

Domain clusters: A specialized domain clusters are built to support the keyword extraction from reviews of previous users. A domain clusters is a collection of reference work of the keyword-preferences list that lists words grouped together according to the similarity of keyword meaning, including related and contrasting words and antonyms.

Table 1: sample keyword-preferences list of movie recommendation

No	Keyword	No	Keyword
1	Action	6	Crime
2	Adventure	7	Drama
3	Animation	8	Horror
4	Children's	9	Romance
5	Comedy	10	Mystery

B. User Preferences/ Choices
In this step, the preferences of active user (current user) and previous users are grouped into their corresponding preference keyword sets respectively. First, HTML tags start and stop words in the reviews keyword sets should be removed to avoid affecting the quality of the keyword extraction process in the next stage. And the Porter Stemmer algorithm (keyword stripping) is used to calculate user preferences/choices. The ratio of preferences/choices of previous users and preference/choice of active user are calculated to observe accuracy in keyword-preference list. Its main use is as part of a term normalization process that is usually done when setting up Information Retrieval systems.

C. Keyword Extraction
In this phase, each review will be transformed into a corresponding keyword set according to the keyword-preference list and domain clusters. If the review contains a word in the domain clusters, then the corresponding keyword should be extracted into the preference keyword set of the user. For example, if a review of a previous user for a movie has the word "Children's", which is corresponding to the keyword ``Comedy" in the domain clusters, then the keyword ``Comedy" should be contained in the preference keyword set of the previous user. The list contains movies which referencing children‟s as well as comedy. If a keyword appears multiple times more than once in reviews, the times of repetitions will be recorded. In this method, it is regarded that keywords appearing multiple times are more important. The times of repetitions will be used to calculate as the weight of the keyword in preference keyword set in the next step.

D. Classify Keywords
When a keyword is captured, it is classified into positive or negative keyword with respect to the meaning of particular word in sentence. The weights of keyword are used to classify the items. For the classification, naive bayes algorithm will be used.

E. Similarity Calculation
The next step is to identify the reviews of previous users who have similar tastes to an active user by finding neighbourhoods of the active user based on the similarity of their preferences. 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.

Two similarity computation methods are introduced in our recommendation method:
1) Approximate similarity computation method.
2) Exact similarity computation method.

1) Approximate Similarity Computation:

The approximate similarity computation method is appropriate for to compute similarity between the preferences of the active user and a previous user for the case that the weights of the keywords in the preference keyword set are unavailable. It is often used method for comparing the similarity and diversity of sample sets. The Jaccard coefficient is measurement of asymmetric information on binary (and non-binary) variables, and it is useful when negative values give to information.

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.

This is shown in figure 3.

2) Exact similarity computation:

The exact similarity computation method is appropriate for to compute similarity between the preferences of the active user and a previous user for the case that the weights of the keywords are available. A cosine-based approach is applied in the exact similarity computation, which is similar to the vector space model (VSM) in information retrieval.

Calculate the weight by the following function:

$$Wi = \frac{1}{m} \sum_{j=1}^{m} \frac{aij}{\sum_{k=1}^{m} akj}$$

F. Generate Personalized Recommendation List

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. The thresholds given in two similarity computation methods are different, which are both empirical values. 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.

5. Conclusions

Proposed model overcomes the limitations of the traditional filtering algorithm such as,
- Scalability
- Poor accuracy

Moreover, to improve the scalability and efficiency in "Big Data" environment, we have implemented it on a MapReduce framework in Hadoop platform. In our future work, we will do further research in how to deal with the case where term appears in different categories of a domain thesaurus from context and how to distinguish the positive and negative preferences of the users from their reviews to make the predictions more accurate.The proposed system is more efficient in terms of complexity. And the system gives more accurate results or recommendations to the users.

References

[1] Shunmei Meng, 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] Magdalini Eirinaki, Malamati D. Louta, Member, IEEE, and Iraklis Varlamis, 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. IEEE 19th 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 Joint Computer 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. Data Mining 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] J. Manyika et al., "Big Data: The Next Frontier for Innovation,
Competition, and Productivity," 2011.

[17] C. Lynch, "Big Data: How Do Your Data Grow?" Nature, vol. 455,
no. 7209, pp. 28-29, 2008.

[18] An Effective Recommendation Framework for Personal Learning Environments Using a Learner Preference Tree and a GA Mojtaba Salehi, Isa Nakhai Kamalabadi, and Mohammad B. Ghaznavi Ghoushchi, Member, IEEE IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES,VOL. 6, NO. 4, OCTOBER-DECEMBER,2013

Paper ID: NOV152687

369