Preference Based Book Recommendation System

A. Rashmi, Y. Ramachandra and Dr.U.P. Kulkarni

Abstract--- Preference based Recommendation system deals with providing the service list to the users based on the preferences of the diverse users. Preferences play an important role in providing the appropriate services. User preferences are indicated by keywords and user-based collaborative algorithm is applied to generate the recommendations. In the current scenario, most of the recommendation systems suffer from issues such as scalability and efficiency in order to overcome these issues, the system is implemented on Hadoop. Hadoop is distributed computing platform that makes use of MapReduce framework. In this project, Hadoop is implemented on single node cluster. This system improves the scalability and accuracy over the existing systems.

Keywords--- Hadoop, MapReduce, Preferences.

I. INTRODUCTION

In the present electronic world the amount and size of the data is increasing at the fast rate hence its becoming difficult to process and analyze the data. This has given enormous opportunities in the field of technology to discover new ways to handle such data. Basically if the size of data exceeds present data size then it is called big data and in the present world Big Data management poses opportunities for software professionals to discover new ways for processing such data. Efforts are made to provide software solutions.

With the invention of Web 2.0, most of the industries have adopted electronic methods to capture information about their employees, customers and their business operations. Service recommender systems play a vital role in helping consumers in selecting the appropriate services. Existing recommender systems provide same recommendations to the users irrespective of the users' preferences. In this project preference based recommendation system is used to provide the appropriate services by considering the preferences of the users.

II. LITERATURE SURVEY

Title: Bayesian-inference Based Recommendation in Online Social Networks

Author: Xiwang Yang, Yang Guo, and Yong Liu

The phenomenally popular online social networks, such as Facebook, twitter, and YouTube. In literature it has been proposed that a Bayesian-inference based recommendation system for online social networks. In the 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. 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. In literature it has been developed distributed protocols that can be easily implemented in online social networks. In literature it has been proposed 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. It has been shown that the proposed Bayesian-inference based recommendation is more accurate than the traditional Collaborative Filtering (CF) recommendation and the existing trust-based recommendations. It allows the flexible trade-offs between recommendation quality and recommendation quantity.

Title: User based collaborative Filtering Recommendation Algorithms on Hadoop

Author: Zhi-Dan Zhao, Ming-Sheng Shang

Collaborative Filtering (CF) algorithm is a widely used personalized recommendation technique in commercial recommendation systems, and many works have been down in this field to improve the performance. However, a big problem of CF is its scalability. Collaborative Filtering (CF) algorithms are widely used in a lot of recommender systems; however, the computational complexity of CF is high thus hinder their use in large scale systems. In literature it has been implemented that user-based CF algorithm on a cloud computing platform, namely Hadoop, to solve the scalability problem of CF. It has been resulted that that a simple method that partition users into groups according to two basic principles, i.e., tidy arrangement of mapper number to overcome the initiation of mapper and partition task equally such that all processors finish task at the same time, can achieve linear speedup.

III. METHODOLOGY

In this approach, preferences of the previous users are collected via reviews given by the previous users, while preferences of the current users are obtained via allowing the user to choose the services from the enlisted service list. Since the preferences of the old users are collected via reviews, it may contain HTML tags and linking words so reviews are preprocessed and then preferences are collected to form keyword set.

Once the preferences are identified, next step is to find the similar previous users having the similar taste as of current user. Unrelated reviews are eliminated. Similarity is computed using two methods namely,
1) Approx commonality computation
2) Exact commonality computation.

Further based on the similarity, we assume certain threshold value. Comparing the similarity between both the users against this threshold value leads to filtering some more values. Finally personalized ratings for each service of the candidates are computed and the one with the highest rating is suggested for the active user.

Book recommendation system is implemented using Hadoop on single node cluster by making use of MapReduce framework.

Candidate services considered in this paper are science, novel, devotional, story, poetry, arts, general. These form a set of keywords. Reviews are used for extracting the preferences of the old buyers, reviews reviews may/ may not contain exact service word so related words are grouped together according to the similarity to form domain thesaurus. This is stored in the database. Once the keywords are extracted for previous and active users, similarity computation is performed. There are two approaches followed to find the similarity between the previous and the current buyer.

In the first method of commonality computation, Jaccard Coefficient is used to calculate the commonality and diversity of data sets. Jaccard coefficient is used in the absence of required information.

\[ J(A, B) = \frac{|A \cap B|}{|A \cup B|} \]

Here A represents preferences of the old buyer of the book. B represents current buyer. In the second method, cosine based approach is used. Sets indicating the preferences of current and old buyers are converted into n-dimensional weight vectors. In this approach, reviews by the old buyer is collected and these reviews are known as 'all reviews' for the enlisted set of preferences. Along with these common list, that are not present in the enlisted services are also collected. These reviews must be converted into keywords based on domain thesaurus and keyword candidate set.

Once the commonality computation is done next step is generation of recommendation by eliminating the irrelevant entries. Personalized rating is done by considering the certain threshold value and if the similarity factor of current and old reader is less than the threshold then that particular old reader is eliminated.

Finally to enhance the scalability of the algorithm, we implement the algorithm on map reduce framework on Hadoop.

Main steps of the method is shown below

---

IV. SYSTEM ARCHITECTURE

The Book Recommendation system consists of four modules namely—Collecting the preferences of the users based on the enlisted keywords Commonality Computation Recommendation Generation and Computation of Personalised Rating Implementation of the system on HadoopMapreduce

In the first module for collecting the preferences of the active reader, we make use of GUI having the enlisted preference keyword set from which reader has to select the preferences. For collecting the preferences from the reviews of the previous reader porter stemmer algorithm and stop word removal algorithm is used.

In the second module, as discussed commonality computation is carried out by Jaccard coefficient method for approximate calculation and Cosine based method for exact commonality computation.

In the third module, we will compare the commonality between the current reader and old reader, based on the threshold value if the commonality factor of both readers is less than the specified threshold value, specific old reader is eliminated.

In the final module, recommendation system is implemented on hadoopmapreduce framework.

---

Figure 1: KASR’s Mainsteps

Figure 2: Architectural Design of Book Recommendation System
V. RELATED WORK

This paper is extension of the recommender system by [1]. The authors have implemented the recommender system for hotels. In [5], the authors have used Bayesian method for designing the recommendation system for online social networks. In [4] the authors have given an overview of the current generation of recommender systems. But most of the existing system recommend the services based on the single rating. In order to have the better scalability and efficiency, we make use of hadoop map reduce framework for implementation.

VI. CONCLUSION AND FUTURE WORK

In this paper we have made use of KASR[1] for implementing the preference based Book recommender system. We have used user based collaborative filter algorithm for generating the recommendations. Our system aims at providing the appropriate services to the readers.

In our future work, we will be considering the different reviews of the readers. In this paper only positive aspects of the old readers are taken into account in the future scope negative comments are also considered.

ACKNOWLEDGEMENT

I would like to thank Prof. Ramachandra and Dr. U.P. Kulkarni for guiding me in this process of paper presentation. This paper is extension of paper [1].

REFERENCES