Service Recommender System for Big Data Applications
Pratibha Chole, Pradeep Patil

Abstract

Recommender systems help users in finding items of interest. It can be defined as the subclass of information filtering system that presents lists of items which are based on user interest. Big data is all about handling vast amount of data. In modern years, amount of customers, services and online information is increasing rapidly, so the big data analysis problem is generated for service recommender systems. As numbers of alternative services are growing, recommending services that users preferred effectively have become an important research issue. Existing service recommender systems present the same ratings as well as rankings of services to different users without considering various users preferences so fails to meet users personalized requirements. In this paper system uses Keyword based approach and the main goal of a system is to give personalized service recommendation list by considering positive and negative aspect from the review. Particularly, keywords are used to denote user’s priorities and a User-based Collaborative Filtering is selected to generate appropriate recommendations. It is implemented on Hadoop using the MapReduce parallel processing model.

Keywords: Recommender system, preference, big data, Hadoop.

I. INTRODUCTION

Every day people take decisions like which movie should I see?, which city should I visit?, what should I eat? etc. there are many choices and a little time to explore them all. Recommender system helps people to take decisions in these complex information spaces. Recommender systems are a type of information filtering that presents lists of items (films, songs, books, images, videos, products) which are based on user interest. Examples are Netflix, Pandora, etc. The amount of data in real world increases rapidly that cause’s the big data problem. Big data means a large amount of data. Big data refers to a dataset which is difficult to capture, store, manage, process and examine with the available present technology within the endureable speed and time. Volume, Variety and Velocity are the factors of big data. Volume means the quantity of data that is generated: Variety may be in the form of structure, unstructured, semi structure data. Velocity means speed of generation of data. Big data management is a serious problem to IT companies. Also it poses heavy impacts on service recommender systems. In today’s world everybody needs to purchase the best thing not only in sensible cost but also in short time. In such situation, recommender system plays an important role.

II. MOTIVATION

The motivation behind the research work is the number of customers, services and online information is increasing rapidly, so the big data analysis problem is generated for service recommender systems. Also in most existing service recommender systems, for example hotel reservation systems and restaurant guides, the ratings of services and service recommendation list presented to users are same. They have not considered users different preferences and personalized requirements.

III. RELATED WORK

There have been many recommender systems developed not only in academia but also in industry. Current recommendation methods usually can be arranged into three main categories: content-based, collaborative and hybrid recommendation approaches. In [2], Personalized Travel Recommendation System is given by considering automatically mined knowledge from travel photo logs specific user profiles or attribute, group type in photo contents. An information theoretic measure is used to demonstrate that user profiles are informative as well as effective for travel recommendation. In [3], Adomavicius and Tuzhilin give an analysis of the field of recommender systems and describe three current recommendation methods content-based, collaborative and hybrid recommendation as well as they describe different limitations of present service recommendation methods and discuss feasible expansions that can increase recommendation abilities and prepare recommender systems applicable to an even wide range of applications. X. Yang al. [4] presented a Bayesian inference based recommendation system for online social networks. An iterative algorithm is developed to compute most feasible recommendation. In [5], this paper Recommender system for sport videos is given, which is based on the audiovisual consumption. The System can produce recommendations for live as well as recorded events. In [6], to recommend Web services actively authors proposed a novel Web service recommendation method and framework based on users interests as well as preferences according to usage history. AWSR (Active Web Service Recommendation) is
implemented and deployed on the Web. AWSR system can recommend Web services which are uniform with users interests and preferences from usage history as well as it satisfy their functional and non-functional requirements. Jin et al. [7] propose a large-scale video recommendation system. In this system an item-based CF (Collaborative Filtering) algorithm is used. Proposed approach implemented in Qizmt, that is a .Net Map-Reduce framework. The system can be utilized for wide-ranging video sites. In [8], big data background designing and implementation of scalable recommender Systems are possible due to the development of cloud computing software tools like Apache Hadoop, Map-Reduce. CF (Collaborative Filtering) algorithm is executed on Hadoop and scalability problem solved by dividing data set. If the amount of data grows, the proposed method doesn’t have favorable scalability and efficiency. In [9], spam emails means where ones email storage inbox is attacked with emails that make no sense, they are on the rise. The Bayesian spam detection scheme with context matching is evolved by implementing the keyword stripping utilizing the Porter Stemmer algorithm. As the root or stem word is only considered that makes keyword search more efficient.

IV. PROBLEM STATEMENT

Develop a Service Recommendation System which provides appropriate services to users based on various user preferences by considering the positive and negative aspects in review for big data applications.

V. SYSTEM ARCHITECTURE

The proposed system is for the hotel recommendation. Input to the system is keywords with importance degree of the keywords. Output will be the personalized service recommendation list with highest ratings services. For this system services and domain thesaurus are used. Table I represents the services of a Hotel Recommender System.

<table>
<thead>
<tr>
<th>Sr. No</th>
<th>Keyword</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Room</td>
</tr>
<tr>
<td>2</td>
<td>Service</td>
</tr>
<tr>
<td>3</td>
<td>Food</td>
</tr>
<tr>
<td>4</td>
<td>Cleanliness</td>
</tr>
<tr>
<td>5</td>
<td>Value</td>
</tr>
<tr>
<td>6</td>
<td>Fitness</td>
</tr>
<tr>
<td>7</td>
<td>Business service</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Keywords</th>
<th>Domain Thesaurus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Room</td>
<td>Bed, Room, Bathroom, size, bedroom, square, shower, space, furniture, comfortable, bath.</td>
</tr>
<tr>
<td>Service</td>
<td>Service, Waiter, Staff, Server.</td>
</tr>
<tr>
<td>Food</td>
<td>Eat, Food, Dishes, Dinner, Breakfast, Delicious, Meal, Restaurant, lunch.</td>
</tr>
<tr>
<td>Cleanliness</td>
<td>Cleanliness ,Clean, Dirty, Grubby, Neat, tidy, musty, spotlessly</td>
</tr>
<tr>
<td>Value</td>
<td>Price, Cheap, Value, Worth, Money, Expensive, Pay ,Experience, cost, accommodate, dollar, choice, overprice</td>
</tr>
<tr>
<td>Fitness</td>
<td>Fitness ,Pool, , Gym, Swimming, Spa</td>
</tr>
<tr>
<td>Business service</td>
<td>Business service ,computer management, internet, speed, laptop, Wi-Fi, broadband, pc, connectivity, high speed, dial-up, speedboats</td>
</tr>
</tbody>
</table>

The system consists of following steps i.e., Preprocessing, Review Classification, Capture user preferences, Similarity Checking and Calculate personalized ratings and generates recommendations. Map Reduce is used to execute the data in a parallel manner. Here the Dataset is added into the database in a parallel manner.

A. Preprocessing

To avoid affecting the quality of the keyword extraction in the next stage HTML tags and stop words in the reviews snippet collection should be removed. And the Porter Stemmer algorithm is utilized to remove the commoner morphological as well as inflexional endings from words in English [1].
B. Review classification

When a keyword is captured, it can be classified into particular aspect. Naive Bayes classifier can be used for classification [10]. It is a supervised as well as Probabilistic learning technique, which concerns about how many times of each word occurred in the document. The probability of a document \( d \) being in class \( c \) is calculated as:

\[
p(c|d) \propto p(c) \prod_{1 \leq k \leq n_d} p(t_k|c)
\]  

(1)

Where \( p(t_k|c) \) is the conditional probability of \( t_k \) occurring in the document of class \( c \), prior probability of a document occurring in class \( c \) is \( p(c) \). \( (t_1, t_2, ..., t_n_d) \) are the tokens in document \( d \) and \( n_d \) is the number of such tokens in document \( d \). In NB (Naive Bayes) classification the most likely or maximum a posteriori (MAP) class is:

\[
c_{\text{map}} = \arg \max_{c \in C} \hat{p}(c) \prod_{1 \leq k \leq n_d} \hat{p}(t_k|c)
\]  

(2)

Where, 

\[
\hat{p}(c) = \frac{N_c}{N}
\]

Where \( N_c \) is the number of documents in class \( c \) and \( N \) is the total number of documents, conditional probability \( \hat{p}(t_k|c) \) is the relative frequency of term \( t \) in documents belonging to class \( c \): 

\[
\hat{p}(t_k|c) = \frac{\sum_{d \in C_l} T_{t_k}}{\sum_{d \in C_l} T_{t_k}}
\]

C. Similarity checking

In this step reviews of previous users who have same tastes like an active user are found out [1].

D. Calculate personalized ratings and generate recommendations

After Similarity checking, for the active user the personalized rating of every candidate service is calculated. Personalized service recommendation list is shown to the user and the services with the top level ratings are recommended to him/her.

VI. RESULTS AND DISCUSSION

The dataset used is the TripAdvisor dataset [11]. This dataset consists of reviews and ratings given by the users for the hotels. To evaluate the performance proposed system we compare system with user-based algorithm using Pearson Correlation Coefficient (UPCC), item-based algorithm using (IPCC) KASR-ESC (Keyword Aware Service Recommendation – Exact Similarity Computation). Mean Absolute Error (MAE) and DCG (Discounted Cumulative Gain) is used for evaluation in this work. Mean Absolute Error (MAE) for proposed system is 0.4037. As compare to existing system it is lower and lower the MAE presents higher accuracy.
VII. CONCLUSION

Recommender system gives recommendations to users according to their tastes. They are applied in variety of applications like books, movies, etc. In this paper, a hotel recommender system is implemented. User gives keyword as a preference, user based Collaborative Filtering and positive aspect in the review is used to produce suitable recommendations. For review classification Naïve Bayes classifier is used. Result shows that as compare to existing system accuracy of proposed system is increased.

VIII. FUTURE SCOPE

As a future work, one can do further research in how to deal with the case where the term appears in different categories of a domain thesaurus from context.

ACKNOWLEDGMENT

We express thank to the anonymous reviewers for their valuable reviews.

REFERENCES


