

Review Based Recommendation System Using Big Data

Ajay Kumar , Bhumika Makkar , Harshit Sinha , Naman Madharia

*School of Computer Science and Technology, MIT Academy Of
Engineering, Alandi (D), Pune INDIA*

ajay-kumar@mitaoe.ac.in
hssinha@mitaoe.ac.in

bhumikamakkar@mitaoe.ac.in
nbmadharia@mitaoe.ac.in

Abstract - In the past few years, the increase in a wide range of options and reviews has led to the need of filtering the data. Many times, the user gets irrelevant or unwanted recommendations due to the use of unfiltered data that are present on the internet. The information overloaded with various and numerous reviews has helped us with the idea of filtering the irrelevant data and using the important data for building a recommendation system.

The current system faces scalability problems due to expansion in Service providers in this domain. This paper explores the big data domain and also resolves the problem of handling big data. We can also observe that using the combination of sentimental analysis with a collaborative filtering approach can give more efficient results.

Keywords— *Big Data , Review Based Recommendation , Sentimental Analysis , Collaborative Filtering Approach*

I. Introduction

In 1990 onwards big data analysis was the challenge in front of the IT industry. Big data term is used when the amount of data exceeds the current capacity of the system to capture and analyse data. Data on social networking sites such as Facebook and twitter, call centre logs, stock exchange data etc is increasing exponentially. This data is in unstructured or semi structured form. Thus, analysis of this big data is a major problem. Due to this huge and unstructured data the recommendation systems are either lagging or facing issues due to unstructured data.

To solve this problem, we are implementing a system that will recommend the desired service using reviews of other customers, using Sentiment Analysis and collaborative filtering for this huge Big Data according to user's preferences.

Nowadays customers' review and satisfaction are very crucial for any enterprise or for any product for enhancing their services. Before purchasing any service or a product customer first search for reviews and satisfaction provided by other customers. Sometimes those reviews may not be true or not sufficient enough, therefore we find it very difficult to judge that particular product or service. Hence, we are trying to implement a system that will extract the customers' review for different sources and will evaluate those reviews and finally the system will recommend the best service or product to the end user.

In our project, specifically we are working on a hotel recommendation system which will be based on customers' reviews.

II. OVERVIEW OF EXISTING WORK

Background

In early 21st century the source of data were not in abundance, therefore, the systems that were working back then faced many problems. There were various kind of upgrades that were provided in order to work in this domain. The steep rise in growth of data source from various platforms in 21st century gave rise to the need of having better platform to handle and analyze it.

So the systems that were earlier being used were comparatively slow and less accurate. The initial release of Hadoop platform in 2006 and Spark platform in 2014 by Apache helped the people gaining more control in big data domain over period of time. The subsequent upgrades in these platforms made the system more easy to use. The language was easily understandable and the system was fast in Handling big data.

Analysis of unstructured data and its processing was part of bottleneck that was occurring lately. This problem was solved with the advent of Hadoop and subsequently spark platform. The approach like user based collaborative filtering overcome the shortcoming of handling such data to greater extent. The primary idea behind using this approach is to help people recommending items that matches their set of preferred items. The proposed system is said to be accurate as the items that are being recommended has high confidence level.

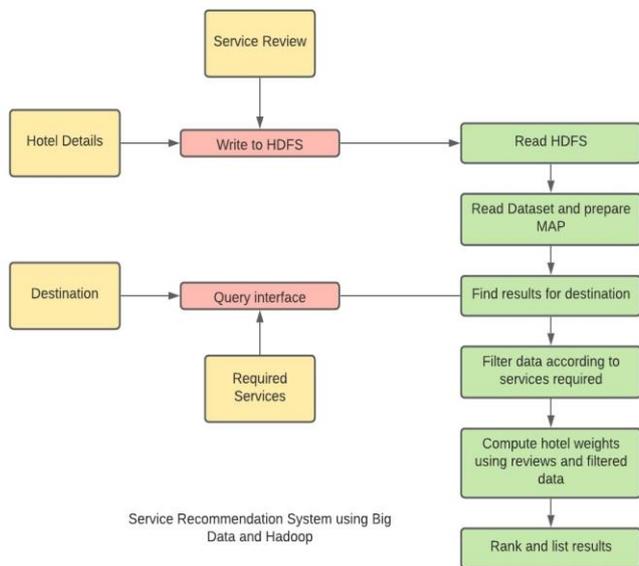


Figure 1 System Architecture

III. PROPOSED METHODOLOGY

- A. **Dataset Cleaning and Filtering :** HTML tags and stop words should be removed so that quality of keywords extraction must be maintained. In this method all keywords are extracted. Porter stemmer algorithm is used for normalisation of words. The data should be filtered before going for action in spark according to the requirements of the user.
- B. **Creating Dictionary:** Now we have a list of Keywords. For rating those words we use word net dictionary. This dictionary contains a keyword and a rating value given to that keyword. In our paper, we are creating our own dictionary in java which will contain a keyword and a value given to that keyword separated by a tab.
- C. **Rating Short Comments:** Calculation of all the rating values is done using Sentiment Analysis. Highest scoring hotel is ranked one and recommended first.
- D. **Keyword Extraction from Long Comments:** Long comments include stop words, spaces, words in ing form so we have to remove all these things. To obtain keywords in root form Porter Stemmer algorithm is used. Stop words and spaces are removed using our own programming logic. Term Frequency (TF) is calculated. In cases where the same keyword is repeated many times, the weight of the keyword is calculated.
- E. **Recommendation using both sentiments and collaborative filtering:** Keywords extracted from short comments, long comments and users preferences are stored in a dictionary. Rate all keywords. Calculate the overall score of hotels which are matching with the user's preferences using sentiment analysis.

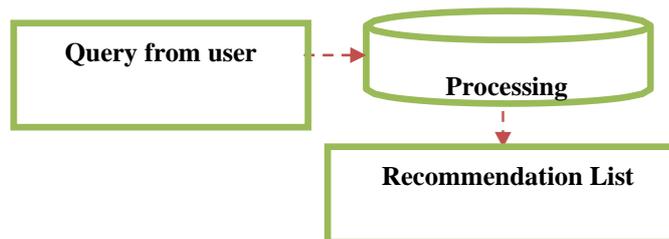


Figure 2 General Concept of Recommendation Method

IV. RESULT

The result obtained from our recommendation system is quite satisfactory. The system is able to rate the keywords from -1 to 1 value. Where -1 represents the worst and 1 represents the best stock works (good, bad, satisfactory etc.).

V. CONCLUSION

In this project a review-based service recommendation will be implemented to recommend services to users. User based collaborative filtering algorithm and sentiment analysis is used to generate appropriate recommendations. Users can give more than one keyword as a preference or can say a user can filter or select their choices like type of payment, cuisine, category of hotel, city etc . We have a huge dataset of hotels with more than 3,00,000 rows.

First dataset cleaning is done. Stop words, spaces are removed then we will filter the dataset according to the required keywords and a new comparatively small but accurate dataset for the process is generated. We have formed the ratings and have given rating values from -1 to +1 after applying sentiment analysis. Sentiment Analysis is used for calculating hotels with the highest rating value and is ranked one and recommended first. This ranking is updated according to the user or the keywords/ filters that user choose.

Along with this to make the system more accurate and dependent we will use ALS/ matrix factorisation algorithm of collaborative filtering. This is already an algorithm which runs parallel and using spark we can run this in a different cluster for faster computation and processing. Collaborative filtering is the prior choice of most recommendation services. The main concept behind collaborative filtering is that users who are having similar taste or opinion for some item will also have same match for other items or services. Finally using both the methods the top-n recommended list of hotels will be generated and this is the recommended list by our system for that particular user.

VI. REFERENCES

1. *Bushra Ramzan, Imran Sarwar Bajwa, Noreen Jamil, Riaz Ul Amin, Shabana Ramzan, Farhan Mirza, Nadeem Sarwar, "An Intelligent Data Analysis for Recommendation Systems Using Machine Learning" Hindwai Scientific Programming Volume 2019 Article ID 5941096*
2. *J.G. Enríquez, Leticia Morales, Fernando Calle-Alonso, Francisco José Domínguez Mayo, "Recommendation and Classification Systems: A Systematic Mapping Study" Hindwai Scientific Programming Volume 2019 Article ID 8043905*
3. *Kavita Mandlik, Dr. Pramod S. Nair and D. Srinivas Rao, "Enhance keyword Based Service Recommendation Using Big Data Analytics" 2017 2nd International Journal of Engineering Trends and Application (IJETA)- Volume 4 Issue 5*
4. *Khushboo R. Shrote and A.V. Deorankar, "Review Based Service Recommendation for Big Data" 2016 2nd International Conference on Advances in Electrical, Electronics, Information, Communication and Bio-Informatics (AEEICB)*
5. *Yifan Gao, Wenzhe Yu, Pingfu Chao, Rong Zhang, Aoying Zhou, Xiaoyan Yang, "A Restaurant Recommendation System by Analyzing Ratings and Aspects in Reviews" : International Conference on Database Systems for Advanced Applications, April 2015 DOI: 10.1007/978-3-319-18123-3_33*
6. *social networks," IEEE Transactions on Parallel and Distributed Systems, Vol. 24, No. 4, pp. 642-651, 2013.*
7. *Yan-Ying Chen, An-Jung Cheng, "Travel Recommendation by Mining People Attributes and Travel Group Types From Community- Contributed Photos" IEEE Transactions on Multimedia, Vol. 15, No. 6, October 2013.*
8. *M. Alduan, F. Alvarez, J. Menendez, and O. Baez, "Recommender System for Sport Videos Based on User Audiovisual Consumption," IEEE Transactions on Multimedia, Vol. 14, No.6, pp. 1546-1557, 2013.*
9. *Zibin Zheng, Xinmiao Wu, Yilei Zhang, Michael R. Lyu, Fellow, and Jianmin Wang, "QoS Ranking Prediction for Cloud Services" IEEE Transactions On Parallel And Distributed Systems, Vol. 24, No. 6, June 2013.*
10. *G. Kang, J. Liu, M. Tang, X. Liu and B. cao, "AWSR: Active Web Service Recommendation Based on Usage History," 2012 IEEE 19th International Conference on Web Services (ICWS), pp. 186-193, 2012.*