

A Book Recommendation System for Cloud Computing: A Survey

Dharna Patel
M.Tech (pursuing)
Computer Science Department
Lakshmi Narain College of Technology, Indore
(M.P), India

Jitendra Dangra
Assistant Professor
Information Technology Department
Lakshmi Narain College of Technology, Indore
(M.P), India

ABSTRACT

Data mining is the approach to extract knowledge from raw data facts. Cloud computing is the emerging technique used to share resources and information among nodes. Recommender systems have made significant utility in daily routine life. Online shopping and Social networking sites are playing crucial role in routine life. Over 3.5 billion people use internet for various purposes. Online shopping retail sales are predicted to grow steadily in upcoming years. Book recommendation is one of the major requirements of e-commerce portals for ease of reading and referencing purpose. This feature can help to increase shopping value with minimum shopping time. Logical recommendation not only helps to customer for purchasing but also increases total sales value. This research paper has attempted to explore the importance of book recommendation and existing solutions. It also investigates the existing solutions and scope of improvement to enhance the performance.

Keywords

Book Recommendation, collaborative Filtering, Classification, Cloud Computing, Amazon Dataset

1. INTRODUCTION

Data mining is the technique to extract knowledge from raw facts or previous transactions. The major goal of data mining is to perform analyzing and knowledge extraction process using different knowledge extraction technique. The study observe that large amount data is available in different sectors and can be very useful for different purpose such as business, marketing, lead generation, security analysis etc. This complete data have no sense and useless for industry until it should converted into meaningful facts. Data mining is an approach to extract meaningful facts known as knowledge from this large volume raw facts and data [1].

Subsequently, it has been observe that use of internet to search information is likely to increase with rapid rate. Especially, there is an increase for online shopping and learning purpose. There has been continuous development in the field of data mining and knowledge extraction. One of the techniques used to foster information retrieval and service is Recommender System. This technique is applied to recommend information which yields more relevant results to users' needs resulting in enhancing users' satisfaction when searching for books in the online shopping portals or libraries[2].

Cloud computing is the emerging technique used to share resources and information among nodes. It is the collection of computer nodes connected to share internet based services. Cloud computing is the new utility provides the sharing

phenomena of upgraded resources which physically does not exist. It is the utility which provides parallel and distributed computing with virtualization feature. The complete phenomena provides scalable and robust environment of service provision. [3]

Today most of the portals and web solutions always implemented with the help of cloud services. Simple object Access Protocol [SOAP] or REST service can be used to implement this cloud service model.

As previously mentioned, the researcher is interested in developing book recommendation system to serve individual needs. This research work attempts to investigate the need of book recommendation system for online shopping purpose. Here, data mining approach has been considered to analyze the data material and recommended the relevant books with high level similarity [4].

2. RELATED WORK

Praveena Mathew et. al. [1] proposed that Book Recommendation System (BRS) based on combined features of content based filtering (CBF), collaborative filtering (CF) and association rule mining to produce efficient and effective recommendation. For this we are proposing a hybrid algorithm in which we combine two or more algorithms, so it helps the recommendation system to recommend the book based on the buyer's interest.

Steps involved in the algorithm,

Step 1: Scan the Books

Dataset In this step scan the entire storage server and simultaneously perform the data cleaning, which include removing the irrelevant data and keeping the relevant data for mining.

Step 2: Data Preprocessing

According to our application, it includes the extraction of data that are needed for mining, which means extraction of only book categories and subcategories.

Step 3: Filtering Transactions

For filtering the transactions categorize the book based on category and subcategories.

Step 4: Perform Content based Filtering

In this step we need to perform content based filtering of books according to user preferences. For example, User1 clicked on book B1, Assume that we have some related books B2, B3 and B4 in the dataset. Assume B2 is of different type, but B3 and B4 is of same type of book B1. Now we check the

content of the books B3 and B4, if the contents match with book kB1, then the system will recommend books B3 and B4 for the user. If user clicks on book B1, then the user will get books B3 and B4 as the recommended.

Step 5: Perform Collaborative Filtering

Here we consider the quality of the book content. In our example, recommending the books B3 and B4.

This will perform based on the registered user's opinions and rating.

Step 6: Final Recommendations

In the final recommendation, the order of the book is consider based on the book which holds highest rating compare to remaining books and arrange in descending order.

This book recommendation system is recommending books to the buyers that suits according to their interest and stores recommendations in the buyer's web profile. This system will store the details of the books which users have bought earlier and find the category of book from users buying history. It using content based filtering and collaborative filtering and find out the list of books based on content and ratings.

The system actually evaluate the quality of the recommending books dependent on the rating given by the existing users also use association rule mining algorithm to finds interesting association and relationship among large data set of books and provide an efficient recommendation for the book.

This system may helpful for lots of people as well as students who need the best books available from the database for both general and academic purpose. Pijitra Jomsr[2] observe that due to wide application of management system, information data grows rapidly. On one hand, people have a large number of information resources. On the other hand, the time cost and difficulty of people finding the proper information increases. To tackle the problems, book recommendation is one of the solutions for university libraries which possess huge volumes of books and reading-intensive users. This paper proposes a library book recommendation system based on user profile loaning and apply association rule to create model. The result shows that new association rule algorithm suitable to apply for recommender book in library. Salil Kanetkar[3] Recommender Systems have been around for more than a decade now. Choosing what book to read next has always been a question for many. Even for students, deciding which textbook or reference book to read on a topic unknown to them is a big question. In this paper, we try to present a model for a web-based personalized hybrid book recommender system which exploits varied aspects of giving recommendations apart from the regular collaborative and content-based filtering approaches. Temporal aspects for the recommendations are incorporated. Also for users of different age, gender and country, personalized recommendations can be made on these demographic parameters. Scraping information from the web and using the information obtained from this process can be equally useful in making recommendations.

Suhasini Parvatikar[4] recommendation systems is used for the purpose of suggesting items to purchase or to see. They direct users towards those items which can meet their needs through cutting down large database of Information. A various techniques have been introduced for recommending items i.e. content, collaborative and association mining techniques are used. This paper solves the problem of data sparsity problem by combining the collaborative-based

filtering and association rule mining to achieve better performance. The results obtained are demonstrated and the proposed recommendation algorithms perform better and solve the challenges such as data sparsity and scalability.

3. CLOUD COMPUTING

Cloud computing can be defined “ A model for enabling ubiquitous, convenient, on –demand network access to a shared pool of resources (e.g., networks, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction”.

Study observes that cloud computing model comprises five essential components can be listed as below:

1. Data
2. Storage
3. Client Networks
4. Application Computing
5. Virtualization

A block representation of component architecture of cloud computing is shown in figure 1.

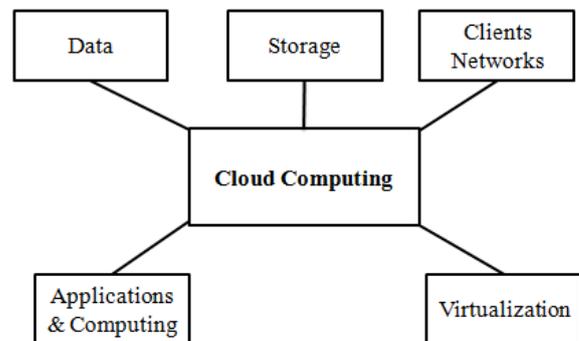


Figure 1: Components of Cloud Computing

Broadly cloud computing environment has been classified into three service model which is illustrated below [2];

1. Software as a Service [SaaS]:

This service model works to share access of software solutions through cloud networks. It provides facility to access devoted software through browser based on “pay us you use” basis. Google docs are the best example of this service model where we can access any document file online through browser support.

2. Platform as a Service [PaaS]:

PaaS service model provide facility to share collection of libraries or environment variable throughout the network. Virtual JVM and Database server is the best practices to implement this service model

3. Infrastructure as a Service [IaaS]:

This service model provides facility to share access of various resources such storage or processing capability out of the network. Cloud server or virtual machine share access is the example of this service model

To implement the cloud service model planning of deployment model is required. Cloud computing classify deployment model into four ways based on access policy. A brief review of types of cloud is cited below.

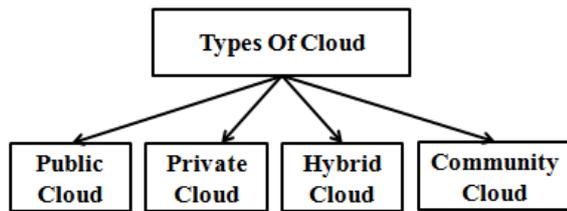


Figure 2: Cloud Computing Service model.

4. DATA MINING

Data mining is sorting through data to identify patterns and establish relationships.

Data mining parameters include:

- Association - looking for patterns where one event is connected to another event
- Sequence or path analysis - looking for patterns where one event leads to another later event
- Classification - looking for new patterns (May result in a change in the way the data is organized but that's ok)
- Clustering - finding and visually documenting groups of facts not previously known
- Forecasting - discovering patterns in data that can lead to reasonable predictions about the future (This area of data mining is known as predictive analytics.)

5. PROBLEM STATEMENT

Online data is the collection of structured and unstructured data organized into relational or non-relational ways. It comes with very large volume and velocity with variety of types. Massive nature makes it complex to analyze and derive relevant conclusions. . Data is becoming one of the most important technology trends that have the potential for dramatically changing the way organizations. Subsequently, growing of e-commerce portals creates a huge attention of users on online shopping and it change the trend of business. Big investors are trying to get into this market and launching their online shopping apps and web applications.

E-Commerce based online shopping portals concentrate to allow huge range of product variety and simplify the way of purchasing. Consumers do not need to worry to visit various shops, compare and buy a product. Although, this sites provide wide range of shopping feature but also increase the consumer expectations. Now a day's consumer wants to purchase lots of product in minimum shopping time. This generates the demand of personalization of shopping portals and recommends products or customized product view based of consumer interest. Customer behavior can play very crucial role in this dimension. Subsequently, previous shopping transactions can also help to observe the shopping demand and consumer interest.

The survey on customer analytics revealed that various algorithms are derived to analyze the customer behavior for Discovering Hidden Relationships of customers, Cause/ Effect analysis, predictive and data visualization behavior. In the same line various product recommendation plug-in are developed for ease of product view and smooth shopping phenomena. Collaborative filtering, Content Mining, Ranking algorithms are integrated with mining solutions to recommend more accurate and perfect results.

The study of existing solutions observes that most of the product recommendations implement ranking and popularity index value estimated by consumer view and ranking algorithms. Analysis of product category and classification of shopping is still a big challenge. Subsequently, e-portals provide web personalization based on product sale not the product nature. Although, limited work has been done in this field and a wide range of improvement is expected. The most important challenge into existing field is small scale data analysis. All the traditional recommendation work has been evaluated and created for small data size. Huge data size not only change the data nature but also increase lots of hurdle during mining work. Another ways, integration of customer behavior analysis and product nature is also absent into existing solutions.

The complete problem statement concludes into following points:

1. Enhancement in product recommendation algorithm for large volume dataset.
2. Integration of customer behavior with Product nature

A solution based on based on similarity, popularity and customer nature index for product recommendation is big challenge to overcome all the above situations.

6. SOLUTION STATEMENT

Recommendation system is a method of filtering that use rating, similarity score or preference score to predict the frequency between item and elements. Recommendation systems have become increasingly popular in recent years due to wide area of applications and use into movies, books, articles etc. prediction and suggestions. Here, a customized recommendation algorithm has been proposed using DBSCAN algorithm to recommend Book based on user nature. Following Points are listed to describe the approach of proposed solution.

1. Initially, Amazon Book Dataset will be used as the input source for user and book information.
2. A customized classification method will be used to extract the nature and important credential from dataset.
3. A parsing module has been proposed to clean the data and remove unnecessary information and attributes.
4. Afterwards, Stop word removal and lemmatization approach is proposed to extract the nature of user and book from user review and other information. This activity is required to convert review text into meaningful word format.
5. A DBSCAN clustering algorithm is proposed to create set of relevant user and optimize the recommendation process. Before doing this clustering, a binary matrix generation is expected to map matching factor between book and users.
6. In the last module a separate book recommendation approach has been proposed. The best part of this module is all existing algorithm calculate the similarity factor for whole dataset which create huge overhead on processor and degrade the performance. This approach will only applicable for relevant cluster where desire user characteristics lie. So it will help to save the extra overhead of existing solutions and recommend the book in best way.

7. A similarity weight calculation has been proposed based on user and book characteristics.
8. Similarity threshold value has been used to filter out the retrieved user id and recommend most close user references.
9. High threshold would represent filter with high strength and it will recommend users with most close values. Subsequently, low threshold represents low filter strength and high numbers of users.

A block diagram to demonstrate the complete solution is shown in figure 3.

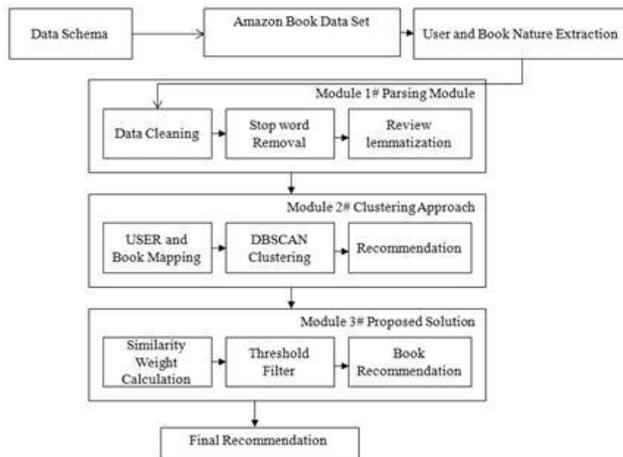


Figure 3: Proposed System Architecture

7. CONCLUSION

The complete work expects an integrated solution based on book characteristics and customer behavior for book recommendation. This paper has attempted to explore the importance and need of book recommendation and also try to

propose solution for same. The complete study proposes to implement this solution in future endeavor for standard dataset and evaluate the performance on basis of computation time and accuracy factor.

8. REFERENCES

- [1] Ms. Praveena Mathew, Ms. Bincy Kuriakose, Mr.Vinayak Hegde, "Book Recommendation System through Content Based and Collaborative Filtering Method" published in IEEE International Conference on Data Mining and Advanced Computing (SAPIENCE), 16-18 March 2016
- [2] Pijitra Jomsri, "Book Recommendation System for Digital Library Based on User Profiles by Using Association Rule" published in IEEE Fourth International Conference on Innovative Computing Technology (INTECH), 2014
- [3] Salil Kanetkar, Akshay Nayak, Sridhar Swamy, Gresha Bhatia, "Web-based Personalized Hybrid Book Recommendation System" IEEE International Conference on Advances in Engineering & Technology Research (ICAETR - 2014)
- [4] Nursultan Kurmashov, "Online Book Recommendation System" Twelve International Conference on Electronics Computer and Computation (ICECCO), 2015
- [5] T. Mather, S. Kumaraswamy and S. Latif, "Cloud Security and Privacy: An Enterprise Perspective on Risks and Compliance," Sebastopol, CA: O'Reilly Media, Inc., 2009.
- [6] Y. Park and K. Chang, "Individual and Group Behavior-based Customer Profile Model for Personalized Product Recommendation.," Expert Syst. with Appli- cations, vol. 36, no. 2, pp. 1932–1939, 2009.