

Hybrid Recommendation Approach for Big Data Applications

Pradip.K .Patil¹ , Prof.G.A.Patil²

P.G. Student, Department of Computer Science & Engineering, D.Y.Patil College of Engineering, Kolhapur,
Maharashtra India¹

Professor, Department of Computer Science & Engineering, D.Y.Patil College of Engineering, Kolhapur,
Maharashtra, India²

ABSTRACT: Service recommender systems are shown as valuable tools for providing acceptable recommendations to users. Within the last decade, the number of customers, services, and online data has grown speedily, yielding the big data analysis problem for service recommender systems. The previous system uses user based collaborative filtering algorithm, which offers some recommendation to a user on the idea of matches in behavior and functional patterns of users and additionally shows similar fondness and behavior pattern with those users. In Hybrid recommendation method we use the basic two recommendation methods such as keyword aware approach and knowledge aware approach for improving the accuracy of service recommendation. The experimentation can be done on the real world data set used in big data applications.

KEYWORDS: Recommendation Methods, Big Data, Service Recommendation.

I. INTRODUCTION

In recent years, the amount of data in the world has been increasing rapidly. The analyses of large data sets so-called "Big Data "becomes a key basis of competition underpinning new waves of productivity growth, innovation, and consumer surplus. The Big Data comes in large-volume which is complex and growing sets with multiple autonomous sources. In Big Data applications, data collection has grown tremendously and it is beyond the ability of commonly used software to capture, manage, and process that data. The most fundamental challenge for the Big Data applications is to explore the large volumes of data and extract useful information or knowledge for future actions.

The Big data is data that exceeds the processing capacity of conventional database systems. The data is too big, moves too fast and doesn't fit the structures of the database architectures. In order to gain value from this data, it is necessary to choose an alternative way to process it. It is necessary to consider the following three categories: Volume, Velocity, and Variety while categorizing Big Data. Each category is associated with some criterion. The volume could be termed as the total amount of data, the variety which refers to a number of types of data and the velocity which refers to the speed of data processing. Similar to the most big data applications, the big data tendency also possesses serious impacts on service recommender systems. With the growing variety of different services, effectively recommending services that users prefer have become a very important analysis issue. Service Recommender Systems have shown a few valuable tools to help users dealing with overload services and provide appropriate recommendations to them. Recommender systems encounter two main challenges for big data application: 1) to make the decision within acceptable time; and 2) to generate ideal recommendations from so many services.

II. RELATED WORK

Most of the research work is carried out on recommendation system, but in recent years many of researchers have found that recommendation can be useful for the purpose of accurate service provision.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 2, February 2018

Shunmei Meng, Wanchun Dou, Xuyun Zhang, and Jinjun Chen [1] have proposed a Keyword-Aware Service Recommendation method, named KASR. They focused on the same ratings and rankings of services to different users without considering diverse users' preferences, and therefore failed to meet users' personalized requirements. It aims at presenting a personalized service recommendation list and recommending the most appropriate services to the users.

F. Chang, J. Dean, S. Ghemawat, and W.C. Hsieh [2] elaborate Big data- large pools of data that can be captured, communicated, aggregated, stored, analyzed and is now part of every sector and function of the global economy. Many projects at Google store data in Bigtable, including web indexing, Google Earth, and Google Finance. These applications place very different demands on Bigtable, both in terms of data and latency requirements. Despite these varied demands, Bigtable has successfully provided a flexible, high-performance solution for all of these Google products. This paper describes the design and the implementation of Bigtable which is a distributed storage system for managing structured data that is designed to scale in a very large Size.

B. Issac and W.J. Jap [3], have implemented spam detection using Bayesian approach and porter stemming keyword stripping approach for excess usage of traffic bandwidth and results in unnecessary wastage of network resources. The use of stem keywords makes the keyword search more efficient, as the search database would be smaller.

P. Castells, M. Fernandez, and D. Vallet [4] propose the model for the exploitation of ontology-based knowledge for solve information retrieval (IR) problem is reduced to a data retrieval task. It helps to improve search over large document repositories.

G. Adomavicius and A. Tuzhilin [5] provide an overview of new and upcoming applications of recommendation technologies. This overview does not claim to be complete but it is the result of an analysis of work published in recommender systems. This is the beginning to sketch major properties of the next generation of recommendation technologies.

Need of work:-

There is a need to improve the accuracy of service recommendation in big data application and also reduce the human efforts of doing analysis process while searching services online by providing mostly accurate recommendations list to the user.

III. PROBLEM STATEMENT

The need is to design and experiment a hybrid approach to overcome the problems that occur in individual recommendation systems. Existing system used individual recommendation systems such as user-based collaborative filtering. Two main problems encountered are-1) Cold-start problem. It is difficult to give recommendations to new users because his profile is almost empty and he has not rated any items so the taste of user remains unknown to the system. Another problem is when the user has not rated before when he is new to the system.2) Data Sparsity problem- Sparsity is the problem of lack of information. Suppose we have a huge amount of users and items but the user has rated only a few items. If a user has evaluated only a few items then it is difficult to determine the taste of the user.

IV. PROPOSED WORK

We propose a hybrid recommendation method, which can be a combination of two or more recommendation techniques in order to increase the overall performance and is adapted to generate appropriate recommendations. The focus is on calculating a personalized rating for each candidate service for a user, then presenting a personalized service recommendation list and recommending the most appropriate services to him/her.

Objectives:-

The objectives of our proposed work are as under.

1. To make a decision within an acceptable time.
2. To generate ideal and accurate recommendations from so many services.
3. To overcome New User/Item Cold Start Problem.
4. To resolve Data Sparsity Problem.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 2, February 2018

Scope of Work:

Individual recommendation systems have problems that can be overcome by novel hybrid approach for recommendation system. It comes with a decision within acceptable time and ideal and accurate recommendations. These individual and hybrid methods will be experimented and will be shown as valuable tools for providing acceptable and accurate recommendations to users.

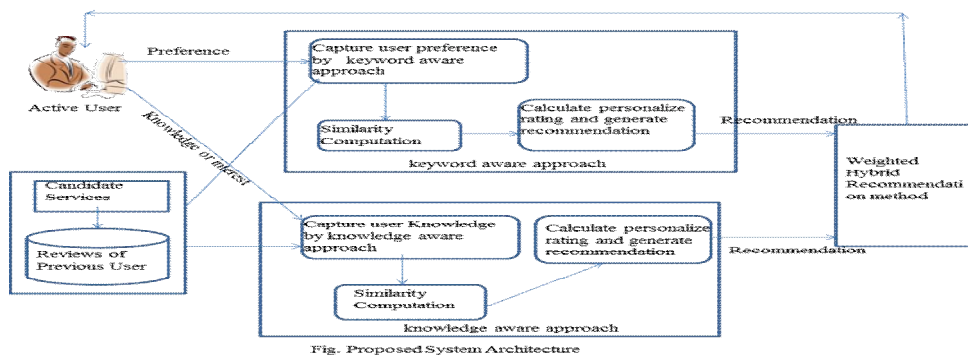


Fig. Proposed System Architecture

Fig 1. System Architecture

Methodology

Following four main modules will be designed and experimented, followed by analysis of those methods.

1. Data Collection and Placing Order
2. Keyword aware Recommendation Method.
3. Knowledge aware Recommendation Method.
4. Hybrid Recommendation Method

1. Data Collection and Placing Order

Data collection is the process of gathering and measuring information. We define data as a collection of product and their attributes, where an attribute is defined as a property or characteristic of a product. User selects the product from the product categories and places order. Data collection and placing order module is consisting of two sub-modules, first is administration and second is user or visitor.

An administrator can do the following activity:

- i. Creates seller and maintains their list.
- ii. Seller can add their products and view the product list
- iii. View user Orders.

User/Visitor can do the following activity:

- i. User Registration.
- ii. View Product list.
- iii. Select Categories and choose the product.
- iv. Place the Order.

2. Keyword aware Recommendation Method

The Keyword aware service recommendation method uses preference about the user to generate a recommendation. In this system, the prediction of the rating of an item for a user depends upon the ratings of the same item rated by similar users.

A) Capture user preferences by a keyword-aware approach:

In this step, the preferences of active users and previous users will be captured and formalized into their corresponding preference keyword sets.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 2, February 2018

i) Preferences of an active user:

An active user can give his/her preferences about candidate services by selecting keywords from a keyword candidate list, which reflect the quality criteria of the services he/she is concerned about. The preference keyword set of

$$\frac{(\vec{W}_{AP} \cdot \vec{W}_{PP})}{\|\vec{W}_{AP}\|_2 + \|\vec{W}_{PP}\|_2}$$

the active user can be denoted as $APK = \{ak1, ak2, \dots, ak_l\}$ where A_{ki} ($1 \leq i \leq l$) is i th keyword selected from the keyword candidate list by the active user, l is the number of selected keywords.

ii) Preferences of previous users:

The preferences of a previous user for a candidate service will be extracted from his/her reviews for the service according to the keyword candidate list and domain thesaurus. A review of the previous user will be formalized into the preference keyword set for him/her, which can be denoted as $PPK = \{pk1, pk2, \dots, pk_h\}$ where p_{ki} ($1 \leq i \leq h$) is the i th keyword extracted from the review, h is the number of extracted keywords.

The keyword extraction process is described as follows:

1) Pre-process: Firstly, HTML tags and stop words in the reviews snippet collection will be removed to avoid affecting the quality of the keyword extraction in the next stage. The Porter Stemmer algorithm (keyword stripping) will be used to remove the commoner morphological and in flexional endings from words in English. Its use for term normalization process is usually done when setting up Information Retrieval systems.

2) Keyword extraction: In this phase, each review will be transformed into a corresponding keyword set according to the keyword candidate list and domain thesaurus. If the review contains a word in the domain thesaurus, then the corresponding keyword should be extracted into the preference keyword set of the user

B) Similarity computation:

The second step is to identify the reviews of previous users who have similar tastes to an active user by finding neighborhoods 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:

i) Approximate similarity computation:

The approximate similarity computation method is for the case that the weights of the keywords in the preference keyword set are unavailable. The similarity between the preferences of the active user and a previous user based on Jaccard coefficient is described as follows:

$$\text{Sim}(APK, PPK) = \text{Jaccard}(APK, PPK) = \frac{|APK \cap PPK|}{|APK \cup PPK|}$$

Where APK is the preference keyword set of the active user,
 PPK is the preference keyword set of a previous user.

ii) Exact similarity computation:

The Exact similarity computation method is for the case that the weights of the keywords in the preference keyword set will be available. A cosine-based approach is applied in the exact similarity computation, which is similar to the vector space model (VSM) in information retrieval. The similarity based on the cosine-based approach is defined as follows

$$\text{Sim}(APK, PPK) = \cos(\vec{W}_{AP}, \vec{W}_{PP}) = \frac{(\vec{W}_{AP} \cdot \vec{W}_{PP})}{\|\vec{W}_{AP}\|_2 * \|\vec{W}_{PP}\|_2}$$

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 2, February 2018

We use the Analytic Hierarchy Process (AHP) model to decide the weight of the keywords in the preference keyword set of the active user. The weight vector of the preference keyword set of a previous user can be decided by the term frequency/inverse document frequency (TF-IDF) measure which is one of the best-known measures for specifying the weight of keywords in Information Retrieval.

C) Calculate personalized ratings and generates recommendations:

Based on the similarity of the active user and previous users, further, filtering will be conducted. Given a threshold δ , if $\text{sim}(APK, PPK_j) < \delta$, the preference keyword set of a previous user PPK_j will be filtered out, otherwise PPK_j will be retained.

Finally, a personalized service recommendation list will be presented to the user and the service with the highest rating(s) will be recommended to him/her.

We use a weighted average approach to calculate the personalized rating Pr of a service for the active user.

$$Pr = \bar{r}_k + k \sum_{PPK_j \in \hat{R}} \text{sim}(APK, PPK_j) * (r_j - \bar{r}_k),$$

$$k = \frac{1}{\sum_{PPK_j \in \hat{R}} \text{sim}(APK, PPK_j)}$$

Where, $\text{sim}(APK, PPK_j)$ is the similarity of the preference keyword set of the active user and the preference keyword set of a previous user PPK_j ,

Multiplier k serves as a normalizing factor,

\hat{R} (R cap) denotes the set of the remaining preference keyword sets of previous users after filtering. r_j is the rating of the corresponding review of PPK_j ,

\bar{r}_k is defined as the average ratings of the candidate service.

3. Knowledge aware Recommendation Method

Knowledge-based Collaborative Filtering approach uses knowledge about users and product to generating a recommendation.

The steps included in knowledge aware service recommendation method are as follows.

A) Capture user Knowledge by a Knowledge aware approach:

An active user can give his/her knowledge about how a particular item meets a particular user need.

B) Calculate personalized ratings and generate recommendations:

The personalized ratings of each candidate service for the active user can be calculated by using multi-attribute utility theory (MAUT), which evaluates each item with regard to its utility for the user. Each item is evaluated according to a predefined set of dimensions that provide an aggregated view of the basic item properties. The user-specific item utility is calculated on the basis of the following formula.

$$\text{Utility}(p) = \sum_{j=1}^{\#(\text{dimension})} \text{interest}(j) * \text{contribution}(p, j)$$

Where, index j iterates over the number of predefined dimensions.

$\text{interest}(j)$ denotes a user's interest in dimension j . $\text{contribution}(p, j)$ denotes the contribution of item p to the interest dimension j . Finally, a personalized service recommendation list is presented to the user and the service(s) with the highest rating(s) is recommended to him/her.

4. Hybrid Recommendation Method

Hybrid recommender systems combine two or more recommendation techniques to gain better performance with fewer of the drawbacks of any individual one. Most commonly, collaborative filtering is combined with some other technique in an attempt to avoid the ramp-up problem. We used hybrid recommendation method where in the basic two recommendation methods such as keyword aware approach and knowledge aware approach are experimented for improving the accuracy of service recommendation.

International Journal of Innovative Research in Science, Engineering and Technology

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Visit: www.ijirset.com

Vol. 7, Issue 2, February 2018

V. CONCLUSION

In Big Data application recommendation system plays vital role. This paper focuses on various recommendation methods. Some of the recommendation systems have drawbacks like cold start problem and data sparsity. To avoid such kind of problems hybrid recommendation would be good solution. We assume that experimental result will show more accurate and efficient recommendation to active user by combined approach.

REFERENCES

- [1] Shunmei Meng, Wanchun Dou, Xuyun Zhang, and Jinjun Chen "KASR: A Keyword-Aware Service Recommendation Method on MapReduce for Big Data Applications (2014)". IEEE transactions on parallel and distributed systems, vol. 25, no. 12, December 2014
- [2] F. Chang, J. Dean, S. Ghemawat, and W.C. Hsieh, "Bigtable: A Distributed Storage System for Structured Data," ACM Transaction. Computer Systems, vol. 26, no. 2, article 4, 2008.
- [3] B. Issac and W.J. Jap, "Implementing Spam Detection Using Bayesian and Porter Stemmer Keyword Stripping Approaches," Proceeding. IEEE Region 10 Conf. (TENCON '09), pp. 1-5, 2009.
- [4] P. Castells, M. Fernandez, and D. Vallet, "An Adaptation of the Vector-Space Model for Ontology-Based Information Retrieval," IEEE Transaction. Knowledge and Data Eng., vol. 19, no. 2, pp. 261-272, Feb. 2007.
- [5] G. Adomavicius and A. Tuzhilin, "Toward the Next Generation of Recommender Systems: A Survey of the State of-the-Art and Possible Extensions," IEEE Transaction on Knowledge and Data Engineering, vol. 17, no. 6, pp. 734-749, June 2005.
- [6] R. Burke, "Hybrid Recommender Systems: Survey and Experiments," User Modeling and User-Adapted Interaction, Springer vol. 12, no. 4, pp. 331-370, 2002
- [7] Y. Zhu and Y. Hu, "Enhancing Search Performance on Gnutella-like P2P Systems," IEEE Transaction. Parallel and Distributed Systems, vol. 17, no. 12, pp. 1482-1495, Dec. 2006.
- [8] G. Salton, Automatic Text Processing. Addison-Wesley, 1989
- [9] A. Chu, R. Kalaba, and K. Spingarn, "A Comparison of two Methods for Determining the Weights of Belonging to Fuzzy Sets," Optimization Theory and Applications, Springer vol. 27, no. 4, pp. 531-538, 1979.
- [10] 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.
- [11] 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.
- [12] K. Lakiotaki, N.F. Matsatsinis, and A. Tsoukis, "Multi-Criteria User Modeling in Recommender Systems," IEEE Intelligent Systems, vol. 26, no. 2, pp. 64-76, Mar./Apr. 2011.
- [13] Z.D. Zhao and M.S. Shang, "User-Based Collaborative-Filtering Recommendation Algorithms on Hadoop," Proc. Third Int'l Workshop Knowledge Discovery and Data Mining, pp. 478-481, 2010.
- [14] 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.
- [15] M. Alduan, F. Alvarez, J. Menendez, and O. Baez, "Recommender System for Sport Videos Based on User Audiovisual Consumption," IEEE Trans. Multimedia, vol. 14, no. 6, pp. 1546-1557, Dec. 2012.