Volume 6, Issue 3, March 2016 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

Service Recommendation System With Efficient Filtering Techniques for Big Data

E. Kodhai, P. Saraswathy, H. Juliana, K. Suruthi
Department of Computer Science, Sri Manakula Vinayagar Engineering College,
Puducherry, India

Abstract: ***The Service Recommendation system provides best recommendation services to the users. The service recommender application became popular among customers. In the past the service recommendation system concentrated on user needs with the help of user based collaborative filtering and the filtering technique which was used in the past does not give much efficiency while recommending. The proposed service recommendation system concentrated on both item based and user based perception with efficient filtering techniques. Such techniques are namely Euclidean distance, Pearson Correlation, Tanimoto Similarity and Uncentered Cosine these techniques are unique by their functionality.***

Keywords: Service Recommendation, Collaborative filtering , Item and User based, Datasets.

I. INTRODUCTION

The term Big Data became popular among the people because of storing and processing large amount of data. Big data mainly refer to datasets, it was related to database software tool in which the data will be captured, stored, managed and analyzed. When the technology increases the data will also increase. In many sectors big data will range from terabytes to petabytes [1]. The internet provides a huge amount of data for business executive. Big data contain structured, unstructured and semi-structured data. The formatted data are Structured data , the unformatted data are Semi-structured and unstructured data. Hadoop uses both unstructured and semi-structured big data, which uses mapreduce to locate the related data directly by answering the query. The available data's can be now saved, retrieved and processed [2].

Big data provide many challenges and opportunities to industries and business [3][4]. Big data have many impacts on service recommendation system. Every day people are dealing with some options and ratings, so these may arise to many solutions and one among them is the recommendation system.

1.1 Recommender System

The content of data is growing day by day in world wide web, so users will get the amount of data as response to their questions. But mining these data to provide an appropriate answer to the user is a tedious work. To overcome this the recommendation systems integrate with web applications which will help the users to find the data easily . The recommendation system is the one which has been used widely in the industry [5].

The most important recommendation systems are content based and collaborative based recommendation. Content based is a continuous process for filtering information. Collaborative recommendation is the most dominant one when compared to other technologies. It is used to recognize the common things between the users based on their ratings and it will produce a new recommendation by comparing item and user. The recommendation system will learn about the profile of the user interest based on the item which the user has rated, this is known as "item-to-item correlation". Collaborative system will work for complex objects like music and movies in which the variation in tastes is responsible for the variation in preferences. This is known as "people-to-people correlation"[5].

1.2 Collaborative Filtering

Collaborative filtering is a technique used in data mining, by considering the rating information it can predict the need of active users from previous users . It is mainly based on the assumption of active user and previous users are the same. This technique is a simple and effective one [5]. A collaborative filtering algorithm is a personalized recommendation technique and it is used in mainly commercial recommender system. The similar tastes and preferences of people will provide a recommendation to other users in collaborative filtering. It is divided into two types item based collaborative filtering and user based collaborative filtering [6]. Item based is used to predict the similarity between items which is calculated by using people rating of those items [7]. On user based system the rating of the item for a user depends upon the rating of the same item rated by similar users [8].

II. RELATED WORK

Xiaoyuan Su and Taghi M. Khoshgoftaar [9] stated that the most successful technique to build a recommender system is collaborative filtering which makes use of the known preferences of a set of users to recommend the unknown

preferences to the other set of users. GhadhAlzamzmi, Yan Albright, NavaneethaHanumanthagowdaa [10] stated that the collaborative filtering technique predict the user preferred products by using the user's history as well as the other user's known history to recommend the product between the user. J.S. Breese, D. Heckerman, and C. Kadie [11] described that the memory based collaborative filtering technique is a kind of user based filtering technique its focus on finding the similarity between the neighbors of the same locality. It combines the preference of user by two categories that is either by collecting users those who rated same rating for a product or by users those who purchased similar products. Joonseok Lee, Mingxuan Sun, Guy Lebanon [12] proposed a comparative study of collaborative filtering technique that the similarity accuracy varies whenever the density of data and the user increases, so each technique suits best in different condition recommending only by considering user preferences does not appropriate in certain conditions.

III. PROPOSED SYSTEM

In the proposed system the filtering technique carried out based on both the item and user. For that the datasets have been gathered from grouplens are used. It consists of a user identification number, movie identification number, rating of the movie, which has been given by the user and the timing. The different filtering techniques incorporated in this system are Euclidean distance, Pearson correlation, Tanimoto similarity and Uncentered cosine similarity. These techniques calculate the similarity and dissimilarity based on user as well as the item, but they are different by their accuracy of calculation.

To know the implementation flow of the service recommendation system on both item and user perception with efficient filtering techniques refers figure 1. The implementation is carried out by five modules. They are

- Collecting datasets
- Pearson Correlation
- Tanimoto similarity
- Euclidean distance
- Uncentered Cosine

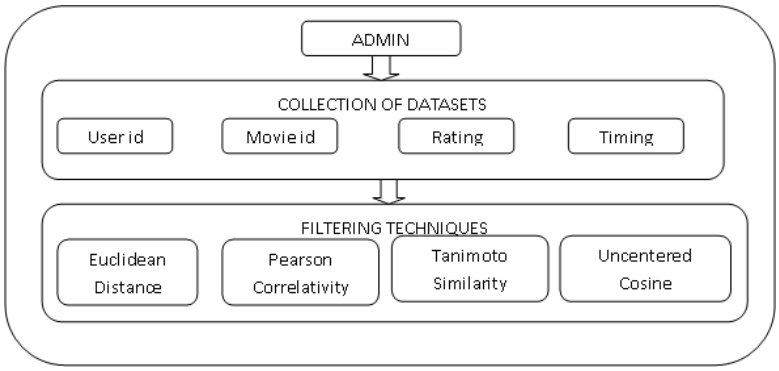

Fig. 1 Service recommendation method with efficient filtering techniques

3.1 Collecting Datasets

The datasets of movies gather from the website known as grouplens. The datasets contain a list of user identification, movie identification number, ratings and the timing for an example 1, 16, 4.0, 1217897793.In which 1 is user id, 16 is movie id, 4.0 is a rating and 1217897793 is timings. The admin will provide the user id and movie id by referring the movie datasets and then the item based and user based method can be performed with the filtering technique.

3.2 Pearson Correlation

Correlation is a kind of technique for analyzing the relationship between two quantitative continuous variables, for example, blood pressure and age. Pearson's correlation coefficient (r) is a measure of the strength of the association. In the field of statistics the Pearson correlation formula is a useful method to measure the strength between variables and relationships. The scale of measurement should be interval or ratio [13].

$$R=\frac{\sum_i(xi-\bar{x})(yi-\bar{y})}{\sqrt{\sum_i(xi-\bar{x})^2}\sqrt{\sum_i(yi-\bar{y})^2}} \quad (1)$$

3.3 Tanimoto Similarity

This method is the ratio of dissimilarity to similarity. Consider to two set A and B with a set of things and ratings. Thus its value equal to zero , then there is no intersection between the set if it's equal to 1 then entire set is intersected[14].

$$T=\frac{N_c}{N_a+N_b-N_c} \quad (2)$$

Where

Na – number of elements in set A

Nb – number of elements in set B

Nc – number of elements intersected.

3.4 Euclidean Distance

The mathematical formula for the Euclidean distance is really simple. An example for two dimensional Euclidean distance calculation is considering 2 points, A and B, with their associated coordinates, the distance is

$$\text{Distance (a, b)} = \sqrt{(a1 - b1)^2 + (a2 - b2)^2 + \cdots + (an - bn)^2} \quad (3)$$

Euclidean distance is a collaborative filtering algorithm which is used to find the dissimilarity between the users and items. It uses a simple mathematical Pythagorean metric for calculation[15].

3.5 Uncentered Cosine

This method is very similar to the previous one, but it gives slightly different results, because this method measures similarity instead of dissimilarity. At the visual the 2 axis and 2 points, the cosine of the theta angle ares needed between the vectors associated with the 2 points. It gives better results for the sample. The values will range between -1 and 1. -1 means that two items are total opposites, 0 means that the two items are independent of each other and 1 means that the two items are very similar [16].

$$\text{Similarity} = \cos\theta = \frac{A.B}{\|A\|\,\|B\|} = \frac{\sum_{i=1}^{n} AiBi}{\sqrt{\sum_{i=1}^{n} Ai^2}\sqrt{\sum_{i=1}^{n} Bi^2}} \quad (4)$$

IV. CONCLUSIONS

The Service Recommendation system based on both item and user deals with efficient filtering techniques like Euclidean distance, Pearson Correlation, Tanimoto Similarity and Uncentered Cosine. These techniques will provide better result when compared to the previously used filtering technique. Each filtering technique described in this paper will provide unique functionality . Hence the proposed system will be implemented on both item and user perception with the help of movie datasets.

REFERENCES

[1] James Manyika, Michael Chui, Brad Brown, Jacques Bughin, Richard Dobbs, Charles Roxburgh, Angela Hung Byers: "Big data: The next frontier for innovation, competition and productivity", 2011.

[2] The emergence of “big data” technology and analytics Bernice Purcell Holy Family University.

[3] F. Chang, J. Dean, S. Ghemawat, W. C. Hsieh, D. A. Wallach, M. Burrows, T. Chandra, A. Fikes, R. E. Gruber:" Bigtable : A Distributed Storage System For Structured Data",2011.

[4] X. Zhang, J. Liu, J. Chen W. Dou, "HireSome-II: Towards Privacy-Aware cross Cloud Service Composition for Big Data Applications," IEEE Transactions on Parallel and Distributed Systems, 2013.

[5] B. Smith, and J. York G. Linden, "Amazon.comRecommendations: Item-to-Item collaborative Filtering," IEEE Internet Computing, vol. 7, no. 1, pp.76-80, 2003.

[6] G. Adomavicius, and A. Tuzhilin, “Toward the Next Generation of Recommender Systems: A Survey of the Stateof- the-Art and Possi-ble Extensions,” IEEE Transactions on Knowledge and Data Engi-neering, Vol.17, No.6 pp. 734-749, 2005.

[7] Sarwar, Badrul; Karypis, George; Konstan, Joseph; Riedl, John (2001). "Item-based collaborative filtering recommendation algorithms".

[8] "Service recommender system for user's personalized requirements" S. Nandhini , S .Pavithra , N. Poojitha , M. Sangeetha-2015.

[9] "A Survey of Collaborative Filtering Techniques", Xiaoyuan Su and Taghi M. Khoshgoftaar-2009.

[10] Comprehensive Survey of Collaborative Filtering Technique, GhadhAlzamzmi, Yan Albright, Navaneetha Hanumanthagowdaa.

[11] J.S. Breese, D.Heckerman, and C.Kadie. Empirical analysis of predictive algorithms for collaborative filtering. In Proceedings of the Fourteenth Conference on Uncertainty in Artifical Intelligence, 1998.

[12] "A Comparative Study of Collaborative Filtering Algorithms" , Joonseok Lee, Mingxuan Sun, Guy Lebanon-2012.

[13] "The relation between Pearson's correlation coefficient *r* and Salton's cosine measure", Leo Egghe and Loet Leydesdorff-2009

[14] Unsupervised Data Base Clustering Based on Daylight's Fingerprint and Tanimoto Similarity: A Fast and Automated Way To Cluster Small and Large Data Sets.

[15] "A New Similarity Measure Based on Adjusted Euclidean Distance for Memory-based Collaborative Filtering" Huifeng Sun, Yong Peng, Junliang Chen, Chuanchang Liu, Yuzhuo Sun 2011.

[16] Sidorov, Grigori; Gelbukh, Alexander; Gómez-Adorno, Helena; Pinto, David. "Soft Similarity and Soft Cosine Measure: Similarity of Features in Vector Space Model" 2014.

