ISSN: 2277-9655
Impact Factor: 4.116

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

A SEMANTIC-BASED COLLABORATIVE FILTERING FOR RECOMMENDATION SYSTEMS

D.Jagadish [*], A.Vishnu Kumar, R.Mani Raj

[*] Department of Information Technology, Vel Tech High Tech Dr.Rangarajan Dr.Sakunthala Engineering College,Avadi, India

DOI: 10.5281/zenodo.60956

ABSTRACT

In the present days the web domain is improved with new types of services, with the increase in service and cloud computing. As a result new forms of web content collecting/designing is done based on the numerous openly available web services online. These services are utilized in many ways by different domains and with the exponential growth of these web services users are experiencing difficulties in finding and utilizing a best matching service for their mashup. A collaborative filtering approach is going to filter and recognize the similar services under same cluster and followed by those evaluation recommendations are made. Semantic based collaborative filtering for recommendation system (SCF-RS) is proposed using clustering and to improve the accuracy of the system.

KEYWORDS: big data, recommendation system, collaborative filtering, clustering.

INTRODUCTION

The term big data is defined as large volume of data which needs new technology and architecture to extract value from it by capturing and analysis process. Big data brings a number of benefits to business organizations and has become a necessity because the organizations need to deal with various challenges and issues associated with adding and adapting to new technologies like social networks. Use of Big data [10] has increased due to increase in use of data-intensive technologies. Big data concept involves datasets which continue to grow and become difficult to manage using existing database management concepts and systems. Important problem associated with big data analysis is the lack of coordination between database systems. Future Big data applications could be automatic creation of queries for content creation in websites for promotions or recommendations.

The term Semantic is defined as study of meaning. The study of semantics is closely linked to the subjects of representation, reference and denotation. It figures out the meaning of the respective term (construct meaning representations). Process language to produce common – sense knowledge about the word (extract data and construct models of the word) [8] Semantic analysis is important in some constraints such as power of language, it is generic in nature. Semantic analysis in various areas such as phonetics which study the linguistic sound when compared to morphology, it is meaning of the component of words. eg – Write, Right.

Collaborating filtering (CF) approach is used to filter and recognize the similar services under same cluster and followed by those evaluation recommendations are made. The basic idea of collaborative filtering is to provide the recommendation. A string matching technique [14] is used in collaborative filtering to match and fix the exact data from the user query. Here ABE algorithm is used for security purpose. The main assumption of recommendation system is to filter the required data from the user query through collaborative filtering and how the system is optimized. Attribute-based encryption (ABE) a novel vision designed for public key encryption [7] that allows the user to encrypt and decrypt message based on user feature. In this system data owner store the data in encrypted form. When the user asked for a query it searches the data which is stored in the database if the word matches the encrypted keyword it is decrypted and sends it to the user for recommendation.

RELATED WORK

ClubCF clustering-based collaborative filtering approach is used in big data [12] application to explore large volume of data and extract useful information or knowledge for future actions. Since number of service in a cluster is much less than the total number of services compilation time is reduced by collaborative filtering. Two stages are used such as cluster stage and collaborative stage. Using Agglomerative Collaborative Filtering (AHC), Pearson Correlation Coefficient (PCC) and Cosine Similarity were used. Service Recommendation (SR) is less in traditional CF algorithms. Services are merged into some clustering through AHC algorithm where the number of services in a cluster is much lesser than the whole system. ClubCF costs less online computation time. It is expected to reduce the online execution time of collaborative filtering. In order to overcome this complexity string matching algorithm is used.
Customization of recommendation system using collaborative filtering algorithm on cloud, it is related with recommendation system how people buy things from market in reasonable cost and save time so they prefer online shopping were the items are suggested so that they can take decision for particular items, hence recommendation system plays a vital role. It helps recommending item of a similar type as well as predicting an item. As a result of combining user-based and item-based CF accuracy is improved. Accuracy of the results gets improved, Hadoop has increased throughput because of multiple computer nodes, time taken for the problem solving the problem reduced [14]. Advanced features are used to make it effective it run in Hadoop environment. To handle real-time data randomly, HBASE, HIVE, are the better solutions. Collaborative filtering recommender system is to capture the user's interest. Semantic social recommendation algorithm is used to recommend the product to the costumers.

Heterogeneous, Autonomous source with distributed and decentralized control, and seeks to explore Complex and Evolving relationships among data (HACE) theorem which characterized the feature of big data revolution, and processing model from data mining [5]. This technique is used in various organizations and provides essential information for designing big data.

In fuzzy keyword search the straightforward approach provides fuzzy keyword search over the encrypted files while achieving search privacy using the technique of secure trapdoors [7]. However, this approaches serious efficiency disadvantages. The simple enumeration method in constructing fuzzy keyword sets would introduce, which leads to large data storage areas, accuracy is less instead semantic search can be performed for future enhancement. Social tagging system for recommendation, where the information is gathered quickly where people cope with it, the recommendation system is used to over this problem, this paper is mainly used for extending the capabilities of recommendation system. It evaluation in industries, entertainment, and many other purposes hence it is expected for the upcoming technologies to improve the recommendation system. Stability of collaborative filtering algorithm is to perform new measure of recommendation system. They have proposed stable and unstable recommendation Techniques by reviews and ratings. In this paper represents study of stability-related issues in the recommendation system the work should be little more explored. A distributed storage system for structured data, which manages structured data that is designed to scale to a very large size. This paper provide simple data model by designing and implementing big data gives dynamic control over data layout and format. As a result, Substantial amount of flexibility is achieved from designing own data model for big table can be implemented.

Service-generated big data and Big Data-as-a-Service generate in huge volume of data. Big data it is employed to provide common big data related services in order to enhance the efficiency and reduce the cost they provide services for big data and quality of services. It may provide more service scope generated to big data.

SYSTEM DESIGN

In this system the query is received from the user for searching the relevant keyword. The query is in the form of selecting a particular keyword from user request. Semantic search is done in order to find unique keyword of users prerequisite. For searching the relevant keyword string matching function is processed, it is divided into pattern P and text T through this process filtering is done. Likewise to get accurate keyword wildcard-based and gram-based technique is used. The clustered data is analyzed through these techniques. As a result, the query is solved and the key term for search is alone taken from the query after solving it. Optimized recommendation using hit count [10] similar services can be optimized using hit count. When the user purchase the product impulsively the count is getting incremented so that the system analyze the query, which product hits the highest count will be recommended to user, in case if the user ask for particular product brand then it display the specific brand with its hit count another criteria if user does not have any information about the product, it will recommend the highest

hits of the product general recommendation will be displayed to user. This is how the product is recommended to the user. System architecture for SCF-RS is illustrated in fig: 2.

String matching it is a sequence of character, given a pattern P[1….m] and text T[1…n], which find all occurrence of P in T. P occurs with shift s (beginning at s+1). P[1]=T[s+1], P[2]=T[S+2]…….P[m]=T[s+m] which is illustrated in fig: 1. So called s is valid shift or invalid shift. P =abab, T=cbbababaccb, P occurs at s=3 and s=5.

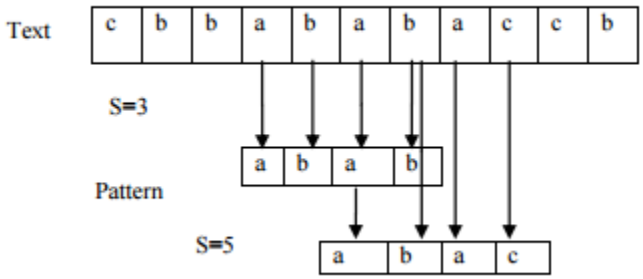


Fig: 1 String matching technique

The objective is to find all occurrence of pattern P=abab in the text T= cbbababaccb. The pattern occurs only once in text, at shift s=3, the shift s=3 is said to be valid shift.

Wildcard are characters that are used to assist into searching for information, it searches the words to know the exact Keyword or it searches for variations of a word. Edit distance process is carried on in three ways substitution, deletion, insertion. Wildcard especially used in situation when the user is not sure of the correct spelling of the word. If want to search for a keyword o*perate it retrieve all records that contain *operate, op*erate, ope*rate, opera*te……..operat.
Gram based is effective technique which is used for matching approximate function. It involve with only one specific character of the keyword either prefix or suffix

eg: operate →operate, oerate, oprate, opeate, operte, operae, operat.

ISSN: 2277-9655
Impact Factor: 4.116

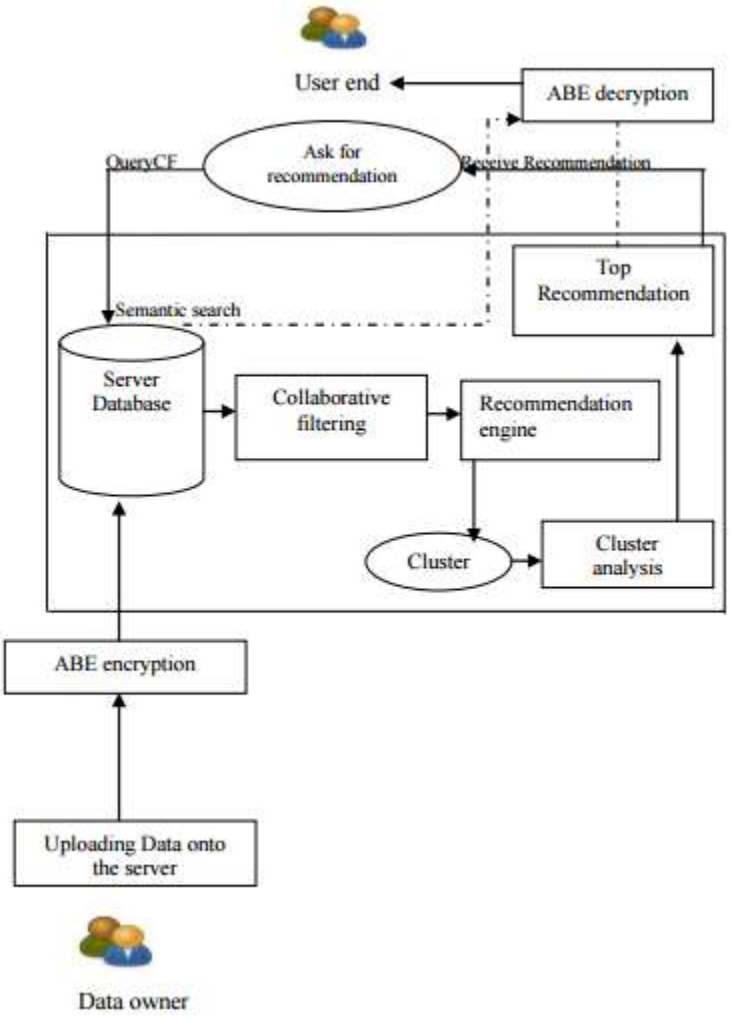


Fig: 2 System Architecture for SCF-RS

Algorithm 1: String Matching Algorithm (SMA)

Input: k- user keyword, ks- set of keywords in DB, Cd- set of character in DB

Output: subset of matching keywords

1. User keyword is given as input.
2. Decrypt the keyword and split the keyword string By inserting a space between characters based on The length of the string
3. Store the character separately.
4. Encrypt the stored character
5. Match the encrypted character with encrypted Keyword stored in the database.
6. If matches

   Construct the keyword k, at edit distance d

   Pre-set distance 1 created within the keyword

   Insert *, is used to represent all other character Sequences in the keyword

ISSN: 2277-9655
Impact Factor: 4.116

Delete the unmatched word
Total number of variants is constructed in the Keyword
Display all the matching keyword related to user keyword
Construct keyword k, to generate distance d
Set the character for approximate string search
Remove the matches which are incorrect

7. else
   Display no result found
8. end if

PERFORMANCE EVALUATION

According to the recommendation system, semantic analysis is performed. To evaluate the service recommendation in collaborative filtering string matching technique is used to find the accuracy of exact keyword.

Consider a keyword "operate" each shift is measured. If K=2 keyword is divided as 'op' whereas it searches for the word 'op' which is linked to other data similarly when K=3,

K=4 accuracy is measured. Accuracy of the keyword is illustrated in fig: 3. Through algorithm (1), keyword is retrieved. In existing system, ClubCF is a revised version of traditional item-basedCF. It is expected to reduce execution time of collaborative filtering. A comparative study is taken for non-SemanticCF and SemanticCF. Computation complex analysis in fig: 4 illustrates the experimental results, it suggests that SemanticCF increase the scalability and services through a recommendation system. By evaluating the hit count true user-specified ratings is taken so that false ratings will be avoided. In semantic collaborative filtering applies less computation time than non-semantic collaborative filtering

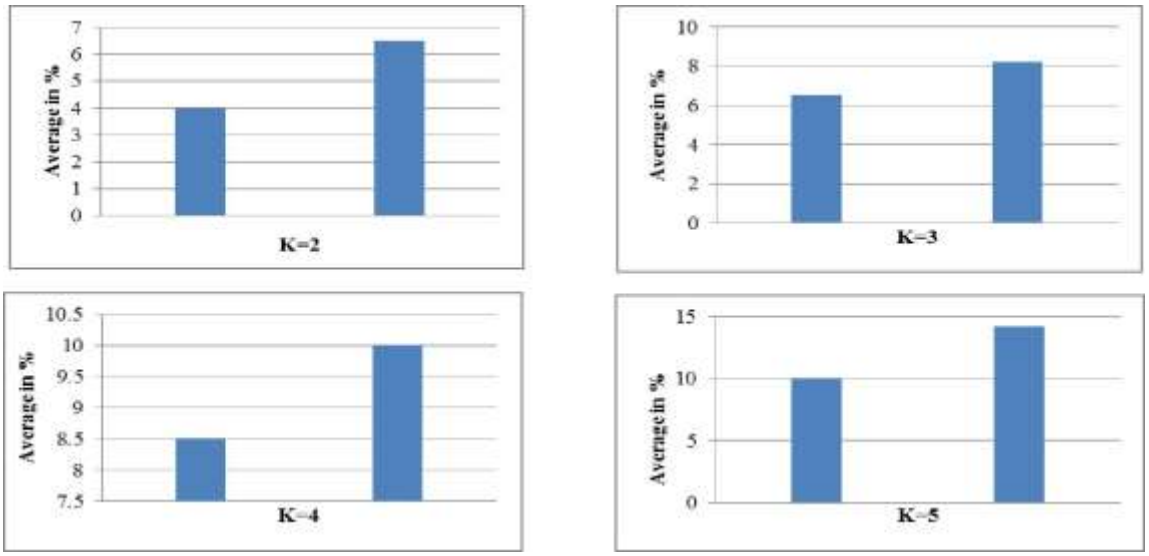
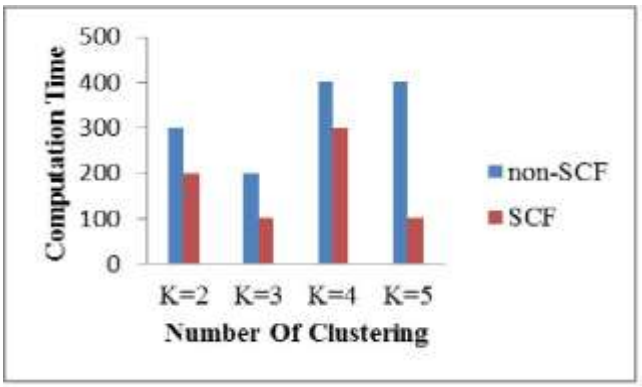


Fig: 3 Keyword accuracy in string matching technique

ISSN: 2277-9655
Impact Factor: 4.116



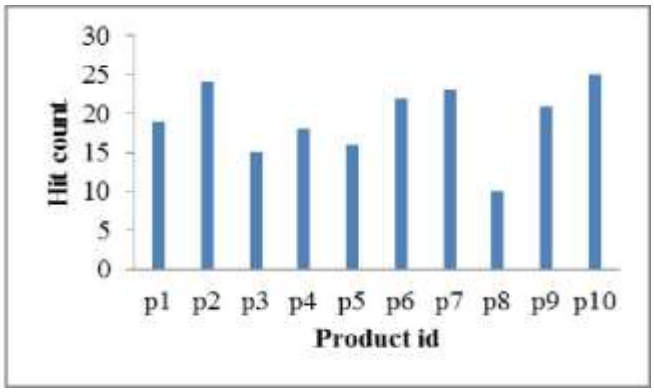


Fig : 4 Comparision of Computation Time with non-SCF and SCF *Fig : 5 Hit count Based Recommendation Accuuracy Estimation*

Since through hit count [17] more accurate computing rating similarity is taken, the predicted ratings of the target services are more accurate than that of clubCF. Through semantic search, precise data is extracted from the server so similar data is also available. Instead of rating services hit count is considered for a recommendation process so that it can avoid false positive and false negative. Once it hit the highest count, it is suggested to the user. Rating accuracy is illustrated in fig: 5. Where p2-23, p7-22, p10-24 hits the highest count.

CONCLUSION

A semantic-based collaborative filtering for recommendation systems has been proposed. An optimal decision has been made for recommendation within an acceptable time using collaborative filtering. In this proposed system rating is evaluated by hit count which gains accurate prediction. Rating similarity results is based on highest hit count products. Through this logic, computation time is reduced when compared with non-semantic CF. The result set is generated based on top recommendation. Future research can be done with respect to user's interest. Mining their implicit interests from usage records may be a complement to the explicit interests. It can also be done in distributed environment.

REFERENCES

[1] Anand Rajaraman , Jure Leskovec and Jeffrey D.Ullman, "Mining of massive datasets,"pp-1-513,2010, 11
[2] Aayad Aabbas and Jaun Liu " Designing an Intelligent Recommender System Using Partial Credit Model and Bayesian Rough Set" Vol.9, No.2, pp-179-187, March 2012
[3] Felix Hernandez del Olmo, Elena Gaudiaoso. "Evaluation of Recommender systems: A new approach"pp-790-804, 2008.Advances in Social Networks Analysis and Mining, 2013, pp-642-647.
[4] Florian Daniel, Federico Michel E Facca, "Current Trends in Web Engineering", pp-115-119 ICWE 2010.
[5] Gediminas Adomavicius, and Alexander Tuzhilin "Towards the next generation of Recommender Systems: A Survey of the state–of–the –Art and Possible Extensions". Vol.17, No.6, pp- 734-749, June 2005.
[6] Gediminas Adomavicius, andYoungok Kwon working paper: "Improving Recommendation Diversity Using Ranking-Based Techniques"pp-1-33,Aug 2009.
[7] Jingnan Xu, Xiaolin Zheng, and Weifeng Ding."Personalized Recommendatoin Based on Reviews and Ratings Alleviating the Sparsity Problem of Collaborative Filtering".pp-9-16, 2012 Ninth IEEE International Conference.
[8] Jin Li, Qian Wang, Ning Cao, Kui Ren and Wenjing Lou"Enabling Efficcient Fuzzy Keyword Search over Encrypted Data in Cloud Computing", pp-1-5, March 2010.
[9] Khaled Sellami, Mohamed Ahmed-Nacer, Pierre Tiako, Computer Science Department, A/Mira University of Bejaia "From Social Network to Semantic Social Network in Recommender System",july 2012.
[10] Kaairo, K, Wikstrom, M, Hamalainen, T "Method for improving Cache hit-rates in Qos-aware load balancing algorithm (QoS-LB) Vol.4, 2001, pp-2321-2325.
[11] M. A. Beyer and D. Laney, "The importance of "big data": definition, Gartner, Tech. Rep., 2012.

[12] Nathaniel Good, J.Ben Schafer, Joseph A. Konstan, AI Borchers, Badrul Sarwar, Jon Herlocker, and John Riedl. Combining"Collaborative Filtering with Personal Agents For Better Recommendations".pp-.1-8, 1999.
[13] Rong Hu, IEE, Wanchun Dou, IEE,Jianun Liu,,IEE.Cl F: "A Clustering-based Collaborative Filtering Approach for Big Data Application".Volume:2,pp-.302-313,11 March 2014 .
[14] Robin Burke, "Knowledge-based recommender systems"pp- 1-43.
[15] Robert Cantu, Gerry Gioia, Kevin Guskiewicz, Blaine Hoshizaki, David Hovda, Ann McKee, Chris Nowinski, William Meehan, Kelly Sarmiento, "Hit Count Threshold White Paper" pp-1-3,2013.
[16] Swati Pandey and T.Senthil Kumar"Customization of Recommendation System Using Collaborative Filtering Algorithm On Cloud Using Mahout", Volume.2,pp-39-43, 2015.
[17] Thaler, D.G Ravishankar, C.V "Using name-based mappings to increase hit rates" Vol-15, 1997, pp-291-303.
[18] Zan Huang, Daniel Zeng and Hsinchun Chen "A Comparative Study of Recommendation Algorithms in Ecommerce Applications"pp-1-23.
[19] X. Wu and X. Zhu, "Data mining with big data," IEEE Transaction Knowledge and Data Engineering, Vol. 26, No. 1, pp. 97-107, January 2014.
[20] Tian.T, Geller, J.Soon Ae Chun, "Predicting Web Search Hit Count" Vol-1, pp-162-166, Aug. 31 2010-Sept. 3 2010.