Diversified and Scalable Recommendation System for the Web Services

[1]Neha Rakeshiya, [2]Nitya khare

[1]*Student, Sagar Institute of Research and Technology excellence Bhopal*
[2]*Prof. Sagar Institute of Research and Technology excellence Bhopal*

Abstract- *Web recommendation got an immense use in current technologies, where an analysis based decision is required in different field such as e-commerce, social media, service providers and other blog platform. In order to perform a competitive analysis various tools are available. Also many research algorithm such as collaborative filtering, HMM, content based filtering, context based filtering is previously performed. Tools which are often using having limitation such as functionality or reporting, as well as they are often costly which cannot be used by every entity. Previous algorithm also finds limitations while computing accuracy and other relevant parameters. In this paper an efficient approach for finding recommendation over web data is done. An experiment performed over a Generated telecom operator dataset using java API. Algorithm computes efficient parameter while comparing with previous approach.*

Keywords: Web recommendation, Collaborative filtering, diversity discovery, data processing.

INTRODUCTION

Web mining is the application of data mining techniques to extract knowledge from web data, including web documents, hyperlinks between documents, and usage logs of web sites.

The World Wide Web has created a colossal quantity of knowledge electronically accessible. The employment of email, news and mark-up languages like markup language permits users to publish and browse documents at a world-wide scale and to speak via chat connections, including data within the sort of pictures and voice records.

Web structure mining [5] is the process of extract patterns from structural components or hyperlinks which used to connect web pages from different locations. But there are a lot of challenges in this process, like collection of real data; it is a lengthy process to collect data from web. Thus it requires collecting metadata by following hyperlinks. Variable nature of the web data on the web changes autonomously and frequently thus, this dynamic nature of the web generate heterogeneity in the web structures. Extraction of the data is also a difficult task which requires because different type of user requires different type of data which fulfill their need, thus an efficient technique is required which enhance the performance of the search and provide data to the user. There is a huge amount of data is introduced into the web by the use of various resources which poses heterogeneity in nature. Thus to extract valuable information from data is a difficult task to do. Generally there are three technique called web content mining, web structure mining, web usage mining are used to provide better performance to the user.

RELATED WORK

There are multiple research is performed by authors for recommendation generation over the period. Although the research performed is having efficient result but still they have low computation parameter value. Some previous work research is discussed here in the field of web mining and recommendation generation.

Guosheng Kang, Mingdong Tang, Jianxun Liu, Xiaoqing (Frank) Liu, and Buqing Cao [1]

This paper contains different approach of diversity finding over the web data. Various algorithms for web data recommendation from large dataset is termed by the authors. The paper depicts diversity content based filtering, collaborative filtering and ranking approach over the retrieved data is done by the paper. The algorithm works on the data history and usage by the users.

Oliver Oechslein, Mario Haim, Andreas Graefe, Thomas Hess, Hans-Bernd Brosius, Anton Koslow [2]

This paper describes the social media content diversity and recommendation generation for the user using the real time available relational data. A hybrid category which aggregate the data and find recommendation over profile. A personalized news aggregation over the dataset is performed and parameters are computed.

T.Nithya [3]
A description over web structure mining is presented. It is the process of extracting useful information from the web to provide better performance to search data. In that hyperlinks are used to be mined to provide better performance to the user. But it is a difficult task to do to mine hyperlinks of web in continuous changing web scenario. Thus a technique is required to provide search mechanism to the user, generally page rank is a technique which used to provide a web structure mining mechanism in web. In that a page rank is assigned on the basis of in-links, out-links or by the use of correlation among the web pages to provide a better mechanism to search data.

B. Rajdeepa and Dr. P. Sumathi [4]
A review over the web mining is presented. A huge amount of data rapidly added to the web which can be introduced by the various resources that generates heterogeneity in that data. Thus to extract useful information from that data is a difficult task to do. There are three type of data presents over the web called web content, web usage data, and web log data. Generally there are three type of techniques called web content mining, web usage mining, web structure mining are used to mine useful information from web. In that there are techniques like web structure mining are used to provide better performance to the user.
These are the previous approaches defined by previous authors to identify the algorithm over web mining dataset.

PROBLEM FORMULATION

As the analysis is performed in previous algorithm but still there are problem formulation with previous algorithm such as:
1. Cold start problem:- It leads to keep data at one outcome and no dynamic evaluation can be performed.
2. Similarity computation problem:- The existing technique persist least strong similarity computation which can further improve using clustering and neural network technique along with the top k-diversify ranking algorithm.
3. Scalability problem :– Scalability over processing large data is still be an issue to work with large dataset and its analysis.

PROPOSED METHODOLOGY

As the previous approach exhibit diversity finding and web mining approach but still due to issue in recognition parameters and previous limitation. Here is the proposed methodology is discussed which is performed in current work.
*   Extracting and loading of all the available network provider dataset and their web services entry in the dataset extracted from various resources.
*   Loading the complete data dictionary pair from the dataset.
*   Performing sorting and loading into the framework.
*   Perform the particular algorithm as per selected by the user for further execution.
*   Performing NLP into the dataset and further storing into temporary storage after NLP term extraction and pruning.
*   Applying different level of search such as state, city , operator level etc.
*   Perform Existing Technique and matching operation if any single match is obtained and calculating the parameters.

PROPOSED ALGORITHM WSTHBA

1. Data distribution loading containing operator data (>10,000 records).
2. Finding Existing algorithm diversity CF.
3. Performing proposed WSTHBA, Threshold computation dynamically over dataset processing and NLP library.
4. Mining data over the operator.
5. Performing data diversity.
6. Apply WSTHBA over other domains.
7. Output: Data diversity score (DDS) over Operators.

RESULTS ANALYSIS

In this section a description over the experiment, which perform to analyze the performance of the proposed technique is introduced. This evaluation shows proposed technique performs better result as compare to the other techniques. An experiment Using Web Java Apache server platform is used. NLP library for data recognition and artificial neural network is also used before processing WSTHBA.
Dataset: A dataset and various web service associated with the operator link is taken from the web resources and loaded in WAMP server 2.5x.

A performance evaluation for the proposed technique is presented. To implement proposed technique an Operator datasets are used. Comparison analysis with the existing technique is also presented.

Computation Time:-

In this parameter time taken by the different techniques to retrieve links from domain which are relevant to the search is presented. Statistical analysis and graphical analysis for the proposed technique is shown.

Statistical Analysis:-

A statistical comparison analysis for the proposed technique is presented in this section which shows, proposed technique provides better results as compare to the existing technique. That shows proposed (WSTHBA) takes ample span of time to search data or links in the web.

Table 1 Statistical Analysis For The Different Technique in MP Region.

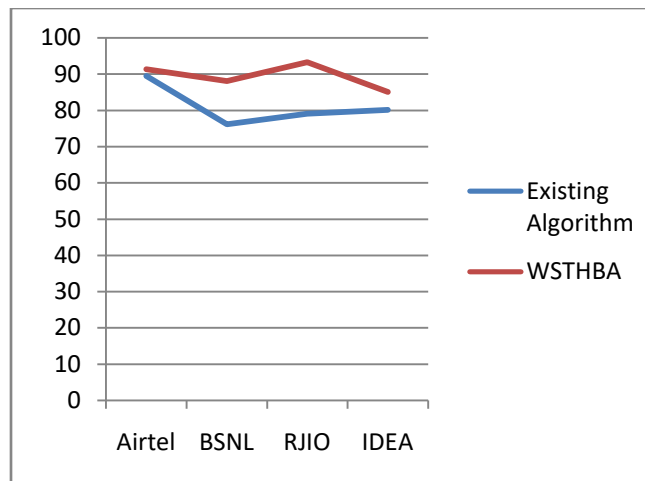| OPERA TORS	EXISTING ALGORITHM PROPOSED WSTHBA					
	Bhopal		Indore		Jabalpur	
Airtel	89.5	91.3	81.4	90.3	89.5	90.4
BSNL	76.12	88	84.21	87.78	89.23	89.9
RJIO	79	90.21	88.78	86.34	87.34	89.2
IDEA	80.1	85	83.32	88.23	87.45	88.8

Graphical Comparison Analysis:

Fig 1: Graphical Analysis for Proposed Method.

A graphical comparison analysis for proposed technique is also presented in Figure 1. It shows proposed technique performs better as compare to the other. In proposed technique consumes less time as compare to the existing technique to assign page rank to the nodes or web pages.

OPERATORS	EXISTING ALGORITHM PROPOSED WSTHBA			
	Nagpur		Mumbai	
Airtel	81.23	83.54	87.34	88.45
BSNL	82.54	87.90	88.56	89.1
RJIO	81.78	85.67	88.3	90.45
IDEA	87.5	87.90	88.1	89.43

Table 2: Result Comparison in city in Maharashtra

In the table 2 above, a comparison between the different city of Maharashtra is shown.
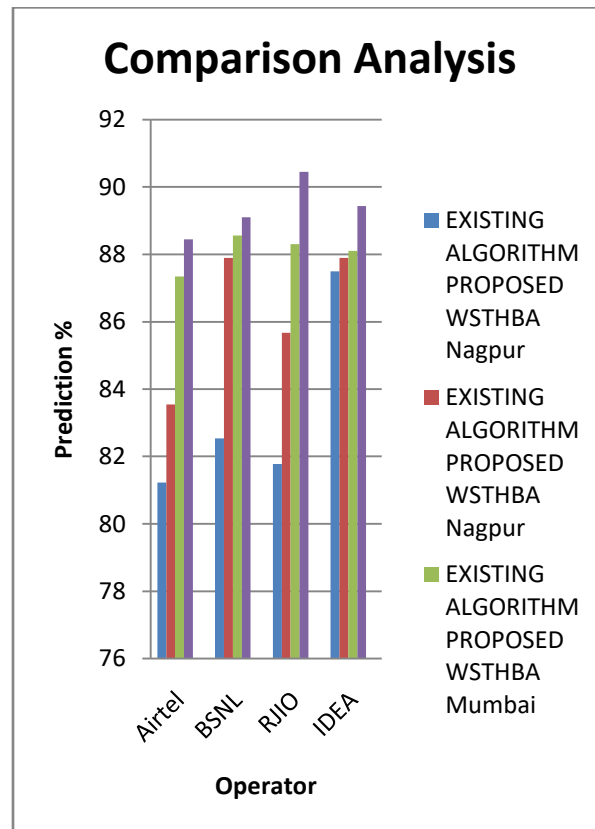

Fig 2: Graphical analysis of city in Maharashtra

In the figure 2, above the Maharashtra city analysis is performed.

SYSTEM ADVANTAGE

Although recommendation system exhibit advantage in its our criteria , but still there are more benefit which can be stated:

1. It recommend user to take decision from large data availability.
2. It helps in process data in low computation , which saves time in processing as well as in decision making.
3. It helps in attribute selection and usage, according to particular area, interest and more specific input parameter.

CONCLUSION

Web service recommendation take an important part in every segment, where the data process , recognition and its usage is highly important with correct data and services. Recommendation system helps to recognise a proper system with its usage along with risk analysis. various problems are also associate with such data that to understand inputs , deal with cold start problem , also to identify new generated data for the training set and inputs. In proposed algorithm an THBA approach is proposed by us which describe how the recommendation can be generated for the large data with high accuracy and low computation time. The work done by us performed on dataset which is related to operator and recommendation generation in different areas are computed. Result parameter compared with existing technique which shows the efficient of proposed algorithm over existing scenario. As the work is performed on recommendation generation and applied in specific category. A further work is going to perform in these following criteria such as Finding generic structure which can get apply to any data set availability of any genre. Also cloud integration with proposed algorithm can be implemented to find efficiency in cloud environment with more availability.

 In future , more parameter can be observed such that more accuracy and efficiency can be monitored.

REFERENCES

[1]. Guosheng Kang, Mingdong Tang, Jianxun Liu, Xiaoqing (Frank) Liu, and Buqing Cao "Diversifying We Service Recommendation Results via Exploring Service Usage History" IEEE, 2015

[2]. Oliver Oechslein, Mario Haim, Andreas Graefe, Thomas Hess, Hans-Bernd Brosius, Anton Koslow "The Digitization of News Aggregation: Experimental Evidence on Intention to Use and Willingness to Pay for Personalized News Aggregators" IEEE, 2015

[3]. Charu Virmani, Anuradha Pillai, Dimple Juneja "Study and Analysis of Social Network Aggregator" IEEE2014.

[4]. B. Rajdeepa and Dr. P. Sumathi, "Web Mining and Its Methods", International Journal of Scientific & Engineering Research June-2013.

[5]. Monika Yadav and Mr. Pradeep Mittal, "Web Mining An Introduction" International Journal of Advanced Research in Computer Science and Software Engineering March 2013.

[6]. Cem Tekin, Mehaela van Der Schaar "contextual online learning for multimedia content aggregation" IEEE Transactions on multimedia Vol. 17 No. 4, April 2015.

[7]. Eric Bruno, Stephane Marchand Mailet "multimodel preference aggregation for multimedia content aggregation" journal of multimedia vol. 4 No.5, October 2009.

[8]. Reoland Ordelman, Fransiciska De Jong, Martha Larson "Enhanced Multimedia content access and Exploitation using semantic speech retrieval" IEEE conference on Semantic Computing, 2009.

[9]. Alberto Messina, Maurizio Montagnuolo "A Generalized cross model clustering method applied to multimedia news semantic indexing and retrieval" April 2009

[10]. Joao Rodrigues, Paulo Salvador and   Antonio Nogueira "Multimedia Content Aggregator applied to an IPTV Content-Zapping Service" IEEE, 2011.

[11]. Rashmi Sharma, Kamaljit Kaur," Review of Web Structure Mininng Techniques using Clustering and Ranking Algorithm" International Journal of Research in and Communication Technology, 6 June- 2014.

[12]. T.Nithya Link Analysis Algorithm for Web Structure Mining "International Journal of Advanced Research in Computer and Communication Engineering " 8 August 2013.