International Journal of Computer Science and Mobile Computing

A Monthly Journal of Computer Science and Information Technology

ISSN 2320-088X
IMPACT FACTOR: 7.056

IJCSMC, Vol. 11, Issue. 1, January 2022, pg.194 – 200

User Search Goals with Feedback Sessions Using Map-Reduce Algorithm

Mr. S.RAJA

Assistant Professor, Parisutham Institute of Technology and Science

Mrs. M.A.NIVEDHA

Assistant Professor, Parisutham Institute of Technology and Science

DOI: 10.47760/ijcsmc.2022.v11i01.025

Abstract— *In the existing search results the search engine users search there results for different goals. It will be very useful for improving search engine relevance and user experience. In existing system there are so many Data mining algorithm used for searching relevant information. Rank algorithm, correlation algorithm has been used but that was not effective. In this paper introduces the map reduce algorithm for improving the search results proposed system, the data mining domain introduce the map reduce algorithm. It is effective for cross domain environment and also can use Big data for future work. The search results achieved through using feedback sessions and map reduce technique. This results of the cross domain have been improved using this technique. The experimented results show the efficiency.*

Index Terms— *Feedback sessions, Web search application, Bigdata, Map-reducing algorithm.*

I. INTRODUCTION

In web search applications different users have different search goals. To represent the needs of user information sometimes the search goals does not exactly represent user's specific information. For example we type "Hindu" in search engine users need for the religion Hindu but it shows for the Hindu newspaper. Information searched for search engine should be satisfied for user. User search goals are considered as Big data but they do not collect the information what they want. In such case we use map-reducing concept [2] R.Baeze-Yates, C.Hurtado and M.Mendoza introduces several advantages in improving user experience and the search engine relevance. Firstly, some advantages according to the user search goals web search engine have been restructured the website. Users come with different search goals can easily find what they want. Secondly, users represent the particular search; we use some keywords that utilized in query recommendation-explained in the following section. Thirdly, splitting, mapping and reducing can be useful in web search results and different search goals[4] S.Beitzel, E.Jensen, A.Chowdhury, and O. Frieder.

Data mining is a process of mining the data from the Database. In the day today activities millions of web sites are available. At the same time there will be a millions of users in the world. Users should be satisfied by the web portals, users are satisfied by an accurate results, time minimization, minimum cost [7] C.-K Huang, L.-F Chien, and Y.-J Oyang. Web portals companies have used many new innovation techniques such as web semantic approach, web ontology, Resource Description Language (RDL). Web languages used in the web portals are HTML, XHTML, XML, java, Ajax. In the search engine when keyword is submitted it will split the keyword into different attributes and then links have been generated.

II. ARCHITECTURE DIAGRAM OF FEEDBACK SESSIONS

In this paper we focus on user search goals for particular query. Feedback session is based on single session and it can be extended to the whole session.

A. Feedback sessions

In the proposed feedback session data have been got by the user click event in the search engine links. When a link is clicked for one time it has considered as "1" click event [11]R. Jones and K.L.Klinkner. Unclicked URL in the search engines are considered as "0". Certain mapping has been given to the links while the link is clicked by user. For example when a link is clicked for 1 time the rank is one likewise for 2 and 3 and so on.

B. Representation of feedback sessions

Feedback sessions are to perform ranking to the links clicked by the user in the web search engine. Ranking to the websites are given by direct user by how user friendly to the user and easy to search user's needs. Click event in the search engine is the feedback session [9]T.Joachims. When user click at the web portals ranking would be given to the web portals [13]U.Lee, Z. Liu, and J. Cho. When web portals is clicked by user ranking number of clicks on the web portals ranking will be increase by 1, 2, 3. By giving a feedback sessions to the web portals by user's original feedback session will be given by the users.

Search results Click Sequence

www.thehindu.com	4
www.wikipedia/hindu.com	2
www.hindureligion.com	1

Fig 1: Click Sequence

C. Feedback sessions algorithm

$$T_{ui} = [t_{w1}, t_{w2}, \dots, t_{wn}]^T$$

$$S_{ui} = [s_{w1}, s_{w2}, \dots, s_{wn}]^T$$

Where T_{ui} is the titles of the feedback sessions.

Where S_{ui} is the snippets of the feedback sessions

$$F_{ui} = w_t T_{ui} + w_s S_{ui} = [f_{w1}, f_{w2}, \dots, f_{wn}]^T$$

Where F_{ui} is the feature representation of i^{th} URL in the Feedback sessions.

Where w_t and w_s are the weights of the title and snippet.

$$F_{fs} = [f_{fs}(w_1), f_{fs}(w_2), \dots, f_{fs}(w_n)]^T$$

$$F_{fs}(w) = \arg \min_{f_{fs}} \left\{ \sum_m [f_{fs}(w) - F_{ucm}(w)]^2 - \lambda \sum [f_{fs}(w) - f_{ucl}(w)]^2 \right\}, f_{fs}(w) \in I_c$$

$$\text{Sim}_{i,j} = \cos(F_{fsi}, F_{fsj})$$

$$= \frac{F_{fsi} \cdot F_{fsj}}{|F_{fsi}| |F_{fsj}|}$$

Distance between the two feedback sessions is

$$\text{Dis}_{i,j} = 1 - \text{Sim}_{i,j}$$

Pseudo documents have been generated by two steps they are

1. Representing the URL in Feedback sessions:

Feedback sessions are collected by various users clicked the web portals in the search engine. By using the feedback sessions pseudo documents are generated by K-means algorithm. Various techniques have been used by using feedback sessions such as various click sequence, providing scalar and vector values to the web portals [14]X. Li, Y.-Y Wang, and A. Acero. Pseudo document is the important method to generate an absolute feedback sessions.

An ambiguous query

Fig 2: Architecture Diagram

In the first step we have to enrich the URL by title and snippet in the web portals. Every web page has title and snippet. Finally title and snippet are implemented by Term Frequency and Inverse term frequency.

By forming a pseudo document based on the URL representation. Pseudo document is generated by the feedback sessions created by users click. Enriched URL is formed by the weighted URL links in the search engine [8]T. Joachims. Enriched URL in pseudo document is generated by both title and snippets used in the URL. Sum of both T_{ui} and S_{ui} .

2. Forming pseudo document based on URL representation:

To obtain the feature representation of the feedback sessions, Optimization method is implemented for both click and unclicked URL. Let F_{fs} is feature representation of feedback sessions and let $f_{fs}(w)$ is the valued to the feature representation feedback session. Let $F_{ucm}(m=1,2,3,...M)$ is the feature representation clicked URL. $F_{ucl}(l=1,2,3,...L)$ is the feature representation of unclicked URL [6]H. Chen and S. Dumais. F_{fs} is the sum of distance between the F_{fs} and each F_{ucm} is minimized and sum of F_{fs} and each F_{ucm} is maximized. Pseudo document is generated by the formula for both clicked and unclicked URL. Similarity between clicked and unclicked URL of generated pseudo document, Cosine score of F_{fsi} and F_{fsj} . When finding the distance between two feedbacks sessions both clicked and unclicked URL in the feedback sessions.

III. MAP-REDUCE TECHNIQUES

Map-reducing algorithm is a mapping algorithm. In this mapping algorithm first mapping the keyword, searching the keyword, grouping the keyword in the certain data warehouse sorting the students alphabetically in the queue. Second, reducing the keyword related to the database. When the keyword is submitted to the search engine keyword is grouped and split into related keyword and grouped to it [10]T. Joachims, L. Granka, B. Pang, H. Hembrooke, and G. Gay. Splitted is having many links and then it will be reduced and produces a number of links in the search engine.

A. Map-reducing Architecture

Fig 3: Map-reducing Architecture

B. Map Reduce Algorithm

```

1: Class MAPPER
2: Method MAP(string t, integer r)
3: EMIT(string t, integer r)

```

```

1: Class REDUCER
2: method REDUCE (string t, integer [r1,r2,.....])
3: sum <- 0
4: cnt <- 0
5: for all integer r ∈ integer [r1,r2,.....] do
6: sum <- sum + r
7: cnt <- cnt + 1
8: ravg <- sum / cnt
9: EMIT (string t, integer ravg)

```

C. Data flow in Map-reduce algorithm

- Input reader.
- Map function.
- Partition function.
- Compare function.
- Reduce function.
- Output reader.

Input reader is the initial step in the Map-reducing algorithm. Input reader is nothing but getting a keyword from the user in the search engine. When the keyword is submitted to the search engine it will search for the desired web links for the keyword. Given keyword should be meaningful.

Map function is the important function in the map reducing algorithm. Map function is to perform mapping the user clicked keyword to the database. Mapped keyword is analyzed performed and produce a web links in the web portals.

Partition function is to perform partitioning the mapped function keywords. Partitioning is nothing but grouping the data by the keyword submitted to the search engine. Partitioning is the important phase in the map reducing algorithm.

Compare function in the map reducing algorithm is to perform the keyword and compare the keyword to the Big data. When it found in the Big data it will compare to the keyword and produce a desired reduced function.

When partition function is completed from the partition function results it reduced the data gathered from the database. Reducing function is important function in the map reducing algorithm.

IV. EXPERIMENTAL RESULTS AND DISCOVERABLE ANALYSIS

Result analysis for our paper is done by using the data mining tool called as weka tools. Weka was developed at the University of Waikato in New Zealand. Weka - open source DM tool developed in Java. Used for Research, Education, and Applications.

A. Input to weka tool

Sl:no	Web portals	Click sequence
1	www.annauniv.edu	10
2	www.facebook.com	25
3	www.twitter.com	17
4	www.gmail.com	30
5	www.google.com	61
6	www.yahoo.com	36
7	www.sony.com	8
8	www.dinakaran.com	23
9	www.dinamani.com	12
10	www.thehindu.in	36
11	www.youtube.com	29
12	www.linkedin.com	13
13	www.ebay.com	19
14	www.twitter.com	27
15	www.microsoft.com	35
16	www.ask.com	27
17	www.paypal.com	61
18	www.apple.com	34
19	www.pinterest.com	26

Table: 1

Sl:no	Web portals	Click sequence
20	www.babylon.com	43
21	www.flickr.com	9
22	www.mywebsearch.com	19
23	www.blogspot.in	45
24	www.instagram.com	69
25	www.adobe.com	28
26	www.bbc.co.uk	54
27	www.avg.com	25
28	www.bing.com	19
29	www.msn.com	23
30	www.sony.co.in	39
31	www.nokia.com	43
32	www.micromaxinfo.com	43
33	www.lenovo.com	43
34	www.dailythanthi.com	29
35	www.tnpsc.gov.in	63
36	www.ibps.in	35
37	www.jabong.com	29
38	www.amazon.in	27

Table: 2

Sl:no	Web portals	Click sequence
39	www.yebhi.com	21
40	www.cognizant.com	47
41	www.snapdeal.com	43
42	www.surfexcelin	43
43	www.raymondindia.com	36
44	www.bhel.com	19
45	www.ongcindia.com	34
46	www.intel.in	23
47	www.avast.com	39
48	www.kaspersky.co.in	16
49	www.lg.com	43
50	www.dinamalar.com	85
51	www.indianrail.gov.in	23
52	www.indianbank.in	81
53	www.hdfc.com	63
54	www.canarabank.in	61
55	www.axisbank.com	25
56	www.cityunionbank.com	40
57	www.icicibank.com	86

Table: 3

Sl:no	Web portals	Click sequence
58	www.iob.in	85
59	www.pnbindia.in	45
60	www.obcindia.co.in	99
61	www.sbi.co.in	89
62	www.kvb.co.in	63
63	www.tmb.in	56
64	www.sunnetwork.in	102
65	www.javanetwork.in	25
66	www.aiwa.com	86
67	www.vijaytv.in	46
68	www.polimertv.com	77
69	www.megatv.com	32
70	www.captainmedia.in	12
71	www.sundirect.in	55
72	www.indianexpress.com	86
73	www.airtel.in	87
74	www.aircel.com	46
75	www.tatadocomo.com	81
76	www.vodafone.in	74

Table: 4

Table 1: link that is searched in the search engine.

Table 2: link that is searched in the search engine.

Table 3: link that is searched in the search engine.

Table 4: link that is searched in the search engine.

V. CONCLUSION

In this paper, we propose a new approach for mining a data from the search engine. We have using several techniques to perform mining concepts such as Feedback sessions, pseudo document and map reducing algorithm. By using these techniques we will produce an optimal solution to the search engine. Sample input and outputs are generated by using the weka tools .Output is generated by the weka tools have various methods. We have using the k means clustering for generating the desired output. By using the paper we proposed the approach for mining a web portal from the search engine using the related keywords. By using the approach we have framed an optimal solution to the users.

REFERENCES

- [1] Zheng Lu, HongyuanZha, Xiaokang Yang, Weiyao Lin, and ZhaohuiZheng,"A New Algorithm for Inferring User Search Goals with Feedback Sessions" IEEE Transactions On Knowledge And Data Engineering, VOL. 25, NO.3, MARCH 2013.
- [2] R. Baeza-Yates, C. Hurtado, and M. Mendoza, "Query Recommendation Using Query Logs in Search Engines," Proc. Int'l Conf. Current Trends in Database Technology (EDBT '04), pp. 588-596, 2004.
- [3] D. Beeferman and A. Berger, "Agglomerative Clustering of a Search Engine Query Log," Proc. Sixth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '00), pp. 407-416, 2000.
- [4] S. Beitzel, E. Jensen, A. Chowdhury, and O. Frieder, "Varying Approaches to Topical Web Query Classification," Proc. 30th Ann. Int'l ACM SIGIR Conf. Research and Development (SIGIR '07), pp. 783-784, 2007.
- [5] H. Cao, D. Jiang, J. Pei, Q. He, Z. Liao, E. Chen, and H. Li, "Context-Aware Query Suggestion by Mining Click-Through," Proc. 14th ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '08), pp. 875-883, 2008.
- [6] H. Chen and S. Dumais, "Bringing Order to the Web: Automatically Categorizing Search Results," Proc. SIGCHI Conf. Human Factors in Computing Systems (SIGCHI '00), pp. 145-152, 2000.
- [7] C.-K Huang, L.-F Chien, and Y.-J Oyang, "Relevant Term Suggestion in Interactive Web Search Based on Contextual Information in Query Session Logs," J. Am. Soc. for Information Science and Technology, vol. 54, no. 7, pp. 638-649, 2003.
- [8] T. Joachims, "Evaluating Retrieval Performance Using Clickthrough Data," Text Mining, J. Franke, G. Nakhaeizadeh, and I. Renz, eds., pp. 79-96, Physica/Springer Verlag, 2003.
- [9] T. Joachims, "Optimizing Search Engines Using Clickthrough Data," Proc. Eighth ACM SIGKDD Int'l Conf. Knowledge Discovery and Data Mining (SIGKDD '02), pp. 133-142, 2002.
- [10] T. Joachims, L. Granka, B. Pang, H. Hembrooke, and G. Gay, "Accurately Interpreting Clickthrough Data as Implicit Feedback," Proc. 28th Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '05), pp. 154-161, 2005.
- [11] R. Jones and K.L. Klinkner, "Beyond the Session Timeout: Automatic Hierarchical Segmentation of Search Topics in Query Logs," Proc. 17th ACM Conf. Information and Knowledge Management (CIKM '08), pp. 699-708, 2008.
- [12] R. Jones, B. Rey, O. Madani, and W. Greiner, "Generating Query Substitutions," Proc. 15th Int'l Conf. World Wide Web (WWW '06), pp. 387-396, 2006.
- [13] U. Lee, Z. Liu, and J. Cho, "Automatic Identification of User Goals in Web Search," Proc. 14th Int'l Conf. World Wide Web (WWW '05), pp. 391-400, 2005.
- [14] X. Li, Y.-Y Wang, and A. Acero, "Learning Query Intent from Regularized Click Graphs," Proc. 31st Ann. Int'l ACM SIGIR Conf. Research and Development in Information Retrieval (SIGIR '08), pp. 339-346, 2008.
- [15] M. Pasca and B.-V Durme, "What You Seek Is what You Get: Extraction of Class Attributes from Query Logs," Proc. 20th Int'l Joint Conf. Artificial Intelligence (IJCAI '07), pp. 2832-2837, 2007.