

A Model of Dynamic User Pattern Cluster Using Two Level Partitioning Algorithm

Ms. Rashmi Patel
PG Student CE
H.G.C.E., Vahelal

Mrs. Risha Tiwari
Asst. Prof., Computer Department,
H.G.C.E., Vahelal

Dr. Dushyantsinh Rathod
Asst. Prof., CE/IT Department,
ASOIT, Ahmedabad

Abstract -The expanded on-line applications are prompting to exponential development of the web content. The vast majority of the business associations are intrigued to know the web client conduct to improve their business. In this unique circumstance, clients route in static and element web applications assumes a vital part in comprehension client's interests. The static mining procedures may not be appropriate as it is for element web log documents and basic leadership. Conventional web log preprocessing approaches and weblog use designs have confinements to break down the substance association with the perusing history This thing, concentrates on different static web log preprocessing and mining strategies and their material confinements for element web mining using this techniques we can create pattern cluster so we can easily retrieve data from data source. In this paper I have just implemented 1st level algorithm only And in future work I have to create pattern cluster for dataset using 2nd level clustering. This algorithm increase 6% of performance, efficiency and accuracy.

Index Terms—:Data Mining, Web log mining, Web Mining, Clustering, Pattern Clustering.

I. INTRODUCTION

Web mining is a process to analyze the online Web data, navigate between various Web sites and perform transaction of data across the Web. According to the types of data can be mined, web mining is classified into three types. Web Content Mining discovers information or knowledge from millions of sources across the Web. Web structure mining is the technique of finding structure information from the web. Web usage mining is the application of data mining techniques to discover interesting usage patterns from web usage data, in order to understand and better serve the needs of web based applications.

II. TYPES OF WEB LOG FILE FORMAT

- W3C(World Wide Web Consortium) Extended Log file Format Extended log is a customizable ASCII format which has different types of fields.
- Microsoft IIS(Internet Information Services) Log file Format can record more information than the NCSA format.
- NCSA(National Centre for Supercomputing Application) Ordinary Log file Format which is available for Web sites but not for FTP sites.

III. EXISTING ARCHITECTURE

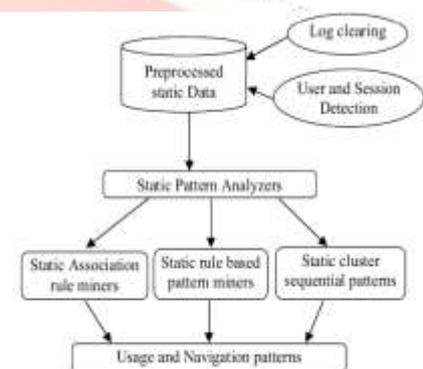


Figure.1:Existing Architecture

IV. EXISTING ALGORITHM

- Read **N** no of records from clean data source **DS**
For $i=1$ to $i \leq N$
- For each records **R** find frequent access data item **F** from data source **DS**
- Read frequency user access item **F**
- If $R = F$ frequent records then
- Save for clustering frequent user access records in frequency access data source **FDS**
- From frequency user access records
- Else not select records
- End If
- Next record

V. DRAWBACKS OF EXISTING

1. It does not provide clustering
2. Does not Cache of visited item
3. It recommended all visited item
4. Doesn't create pattern clustering
5. It gives less performance
6. It consumes time
7. Low efficiency and less accuracy

VI. PROBLEM STATEMENT

Discuss the problem relating to Data cleaning of web log. Web log is generally noisy and ambiguous Web applications are increasing at an enormous speed and its users are increasing at exponential speed. Difficult to find the "right" or "interesting" information, There are a lot of work on data cleaning of web server logs irrelevant items and useless data can not completely removed. Difficulty in specifying the valid data from the log file with unlimited accesses to websites, web requests from multiple clients to multiple web servers.

VII. PROPOSED METHODOLOGY

The Two-level clustering method is improving the quality of data.

1. The onelevel clustering is done in the form of datafrequently user access using clustering method. Remove unwanted or noisy data like .jpg, 404 page not found and any office file.
2. The two level clustering is done by first changing the form of web log data into user access behavior patterns.

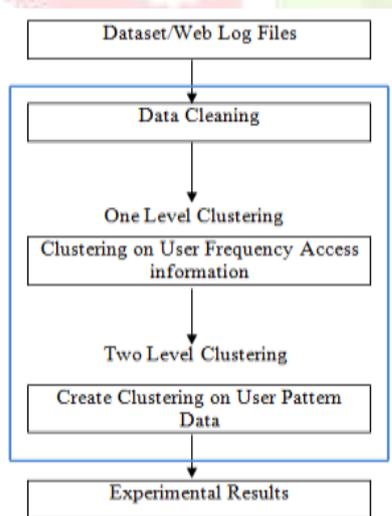


Figure.2:Proposed Clustering Process

VIII. PROPOSED ALGORITHM

1) One Level Algorithm

1. Read N no of records from clean data source DS
For i = 1 to i <= N
Next
2. For each records R find frequent access data item F from data source DS
3. Remove unwanted or noisy data like .jpg, 404 page not found and any office file.
4. Read frequency user access item F
5. If R = F frequent records then
6. Save for clustering frequent user access records in frequency access data source FDS
7. Make cluster from frequency user access records
8. Else not select records
9. End If
10. Next record

2) Two Level Algorithm

1. Read N no of records from clean data source FDS
For i = 1 to i <= N
Next
2. For each records R from data source FDS find pattern data.
3. Read pattern data using specified address from data source FDS.
4. If requested records from frequent data source FDS with specified pattern then
5. Collect and Save in pattern data source PDS.
6. Make two level cluster in pattern data source PDS.
7. Else not select that records.
8. End If
9. Next record

IX. PERFORMANCE AND ACCURACY

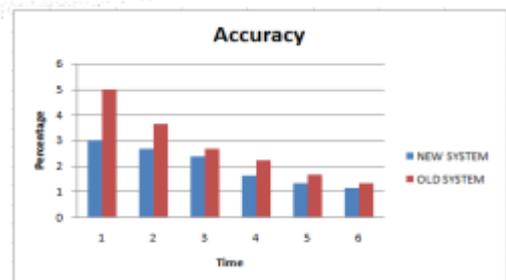


Figure.3: Accuracy Chart

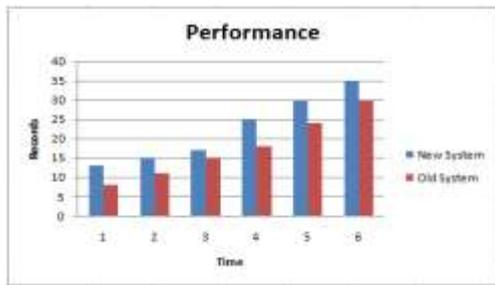


Figure.4: Performance Chart

Sr No.	Old Algo (Records in thousands)	Proposed work (Records in thousands)	Old Algo Accuracy(%)	Propose Algo Accuracy(%)	Time (MS)
1	8	13	5	3	1
2	11	15	3.63	2.66	2
3	15	17	2.66	2.35	3
4	18	25	2.22	1.6	4
5	24	30	1.66	1.33	5
6	30	35	1.33	1.14	6

Figure.5: Accuracy Table

Sr No.	Existing work (Records)	Proposed work (Records)	Time(MS)
1	8	13	1
2	11	15	2
3	15	17	3
4	18	25	4
5	24	30	5
6	30	35	6

Figure.6: Performance Table

X. OUTPUT

ID	Products	Price	Product_Request
1	COOLPAD MEGA	6999	www.amazon.in/Coolpad-Mega-2-5D-Royal-Gold
2	MOTO G	12000	www.flipkart.com/moto-g
3	SONY BRAVIA 32 INCH LED	25000	www.flipkart.com/sony-80cm-32-hd-ready-led-tv
4	LENOVO VIBE NOTE 5	20000	www.amazon.in/lenovo-vibe-note-5
5	SAMSUNG 253 L FROST	40000	www.flipkart.com/samsung-253-direct-cool-single-door-refrigerator
6	ONIDA 42 INCH LED	55000	www.flipkart.com/onida-42-hd-ready-led-tv
7	MOTO G ADVANCE	15000	www.amazon.in/moto-g-advance
8	MICROSOFT LUMIA 640 XL	48000	www.amazon.in/microsoft-lumia-640-xl
9	SONY 40 INCH FULL HD	45000	www.flipkart.com/sony-40-hd-ready-led-tv
10	LG 190 L DIRECT COOL	46000	www.flipkart.com/lg-190-direct-cool-single-door-refrigerator
11	XIAOMI REDMI NOTE 3	18000	http://www.amazon.in/xiaomi-redmi-note-3
12	MICROMAX 32 INCH	30000	www.flipkart.com/micromax-80cm-32-hd-ready-led-tv
13	GODREJ 185 L DIRECT COOL	48000	www.flipkart.com/godrej-253-direct-cool-single-door-refrigerator
14	MICROMAX 40 INCH FULL HD	36000	www.flipkart.com/micromax-40-hd-ready-led-tv

Figure.7: Existing Output

Index No.	Date	Client IP	Server IP	Category	Company	Product	URI_Stream	Status Code
1	2017-04-13	10.8.0.13	202.71.129.26	Electronics	LG	LED	http://www.amazon.in/LG-Inches-D32N190ARJLX	200
2	2017-04-13	10.8.0.13	202.71.129.26	Mobiles	SAMSUNG	Mobile	http://www.amazon.in/Samsung-Galaxy-C7-Navy-Blue	200
3	2017-04-13	10.8.0.13	202.71.129.26	Electronics	LG	LED	http://www.amazon.in/LG-Inches-D32N190ARJLX	200
4	2017-04-13	10.8.0.13	202.71.129.26	Mobiles	LG	Mobile	http://www.amazon.in/LG-G6-FullBody	200
5	2017-04-13	10.8.0.20	67.218.96.251	Mobiles	SAMSUNG	Mobile	http://www.amazon.in/Samsung-Galaxy-S6-Gold	200
6	2017-04-13	10.8.0.22	67.218.96.251	Electronics	SAMSUNG	LED	http://www.amazon.in/Samsung-Inches-UA22K500ARJLX	200
7	2017-04-13	10.8.0.17	67.218.96.251	Mobiles	SAMSUNG	Mobile	http://www.amazon.in/Samsung-Galaxy-S6-Gold	200
8	2017-04-13	10.8.0.18	202.71.129.26	Mobiles	LG	Mobile	http://www.amazon.in/LG-G6-FullBody	200
9	2017-04-13	10.8.0.10	172.16.251.251	Electronics	SAMSUNG	LED	http://www.amazon.in/Samsung-Inches-UA42K500ARJLX	200
10	2017-04-13	10.8.0.26	209.85.135.109	Mobiles	SAMSUNG	Mobile	http://www.amazon.in/Samsung-Galaxy-C7-Navy-Blue	200
11	2017-04-13	10.8.0.19	67.218.96.251	Mobiles	LG	Mobile	http://www.amazon.in/LG-G4-Luster-Red	200
12	2017-04-13	10.8.0.17	202.71.129.26	Electronics	SAMSUNG	LED	http://www.amazon.in/Samsung-Inches-UA22K500ARJLX	200
13	2017-04-13	10.8.0.18	202.71.129.26	Mobiles	LG	Mobile	http://www.amazon.in/LG-G4-Luster-Red	200
14	2017-04-13	10.8.0.18	202.71.129.26	Mobiles	SAMSUNG	Mobile	http://www.amazon.in/Samsung-Galaxy-C7-Navy-Blue	200
15	2017-04-13	10.8.0.10	172.16.251.251	Electronics	SAMSUNG	LED	http://www.amazon.in/Samsung-Inches-UA42K500ARJLX	200
16	2017-04-13	10.8.0.26	209.85.135.109	Mobiles	SAMSUNG	Mobile	http://www.amazon.in/Samsung-Galaxy-C7-Navy-Blue	200
17	2017-04-13	10.8.0.19	67.218.96.251	Mobiles	LG	Mobile	http://www.amazon.in/LG-G4-Luster-Red	200
18	2017-04-13	10.8.0.17	202.71.129.26	Electronics	SAMSUNG	LED	http://www.amazon.in/Samsung-Inches-UA22K500ARJLX	200
19	2017-04-13	10.8.0.18	202.71.129.26	Mobiles	LG	Mobile	http://www.amazon.in/LG-G4-Luster-Red	200
20	2017-04-13	10.8.0.18	202.71.129.26	Mobiles	SAMSUNG	Mobile	http://www.amazon.in/Samsung-Galaxy-C7-Navy-Blue	200
21	2017-04-13	10.8.0.13	172.16.251.251	Mobiles	LG	Mobile	http://www.amazon.in/LG-G6-FullBody	200
22	2017-04-13	10.8.0.26	209.85.135.109	Mobiles	SAMSUNG	Mobile	http://www.amazon.in/Samsung-Galaxy-C7-Navy-Blue	200
23	2017-04-13	10.8.0.19	67.218.96.251	Electronics	SAMSUNG	LED	http://www.amazon.in/Samsung-Inches-UA22K500ARJLX	200
24	2017-04-13	10.8.0.18	202.71.129.26	Mobiles	SAMSUNG	Mobile	http://www.amazon.in/Samsung-Galaxy-S6-Gold	200
25	2017-04-13	10.8.0.13	202.71.129.26	Electronics	LG	LED	http://www.amazon.in/LG-Inches-D32N190ARJLX	200
26	2017-04-13	10.8.0.18	202.71.129.26	Mobiles	SAMSUNG	Mobile	http://www.amazon.in/Samsung-Galaxy-S6-Gold	200
27	2017-04-13	10.8.0.11	202.71.129.26	Mobiles	LG	Mobile	http://www.amazon.in/LG-G6-FullBody	200
28	2017-04-13	10.8.0.15	209.85.135.109	Electronics	SAMSUNG	LED	http://www.amazon.in/Samsung-Inches-UA22K500ARJLX	200
29	2017-04-13	10.8.0.21	67.218.96.251	Electronics	LG	LED	http://www.amazon.in/LG-Inches-D32N190ARJLX	200
30	2017-04-13	10.8.0.29	67.218.96.251	Mobiles	SAMSUNG	Mobile	http://www.amazon.in/Samsung-Note-5-Ds	200
31	2017-04-13	10.8.0.12	202.71.129.26	Mobiles	LG	Mobile	http://www.amazon.in/LG-G6-FullBody	200
32	2017-04-13	10.8.0.22	209.85.135.109	Mobiles	SAMSUNG	Mobile	http://www.amazon.in/Samsung-Note-5-Ds	200
33	2017-04-13	10.8.0.5	202.71.129.26	Mobiles	LG	Mobile	http://www.amazon.in/LG-G6-FullBody	200
34	2017-04-13	10.8.0.18	202.71.129.26	Mobiles	LG	Mobile	http://www.amazon.in/LG-G6-FullBody	200
35	2017-04-13	10.8.0.13	202.71.129.26	Electronics	SAMSUNG	LED	http://www.amazon.in/Samsung-Inches-UA22K500ARJLX	200
36	2017-04-13	10.8.0.13	202.71.129.26	Mobiles	LG	Mobile	http://www.amazon.in/LG-V20	200
37	2017-04-13	10.8.0.12	202.71.129.26	Electronics	SAMSUNG	LED	http://www.amazon.in/Samsung-Inches-UA22K500ARJLX	200
38	2017-04-13	10.8.0.13	202.71.129.26	Mobiles	LG	Mobile	http://www.amazon.in/LG-V20	200

Figure.8: Clean Data

Select Category	Mobiles		
Company	SAMSUNG		
<input type="button" value="Create"/>			
Index No.	Client IP	Server IP	URI_Stream
1	10.8.0.13	202.71.129.26	http://www.amazon.in/Samsung-Galaxy-C7-Navy-Blue
4	10.6.0.20	67.218.96.251	http://www.amazon.in/Samsung-Galaxy-S6-Gold
6	10.6.0.27	67.218.96.251	http://www.amazon.in/Samsung-Galaxy-S6-Gold
9	10.6.0.26	209.85.135.109	http://www.amazon.in/Samsung-Galaxy-C7-Navy-Blue
13	10.8.0.18	202.71.129.26	http://www.amazon.in/Samsung-Galaxy-C7-Navy-Blue
15	10.5.0.51	209.85.135.109	http://www.amazon.in/Samsung-Galaxy-C7-Navy-Blue
22	10.6.0.28	209.85.135.109	http://www.amazon.in/Samsung-Galaxy-C7-Navy-Blue
24	10.8.0.16	202.71.129.26	http://www.amazon.in/Samsung-Galaxy-S6-Gold
26	10.8.0.18	202.71.129.26	http://www.amazon.in/Samsung-Galaxy-S6-Gold
30	10.6.0.29	67.218.96.251	http://www.amazon.in/Samsung-Note-5-Ds
33	10.5.0.5	202.71.129.26	http://www.amazon.in/Samsung-Note-5-Ds
37	10.8.0.12	202.71.129.26	http://www.amazon.in/Samsung-Note-5-Ds
39	10.8.0.10	202.71.129.26	http://www.amazon.in/Samsung-Note-5-Ds
40	10.8.0.13	202.71.129.26	http://www.amazon.in/Samsung-Galaxy-C7-Navy-Blue
43	10.6.0.41	67.218.96.251	http://www.amazon.in/Samsung-Galaxy-C7-Navy-Blue
45	10.6.0.47	67.218.96.251	http://www.amazon.in/Samsung-Galaxy-C7-Navy-Blue

Figure.9: Pattern Data Clustering 1

Index_No	Client_IP	Server_IP	URI_Steam
3	10.5.0.12	59.162.23.130	http://www.amazon.in/LG-G6-Fulvission
7	10.5.0.5	202.71.129.26	http://www.amazon.in/LG-G6-Fulvission
10	10.8.0.15	67.218.96.251	http://www.amazon.in/LG-G4-Leather-led
12	10.8.0.15	59.162.23.130	http://www.amazon.in/LG-G4-Leather-led
16	10.5.0.13	59.162.23.130	http://www.amazon.in/LG-G2-D08
18	10.6.0.23	67.218.96.251	http://www.amazon.in/LG-G2-D08
19	10.6.0.28	67.218.96.251	http://www.amazon.in/LG-G6-Fulvission
21	10.5.0.13	172.30.255.255	http://www.amazon.in/LG-G6-Fulvission
27	10.8.0.11	202.71.129.26	http://www.amazon.in/LG-G6-Fulvission
31	10.6.0.32	67.218.96.251	http://www.amazon.in/LG-G6-Fulvission
34	10.5.0.16	172.30.255.255	http://www.amazon.in/LG-V20
36	10.8.0.53	67.218.96.251	http://www.amazon.in/LG-V20
42	10.5.0.11	59.162.23.130	http://www.amazon.in/LG-G6-Fulvission
46	10.5.0.51	202.71.129.26	http://www.amazon.in/LG-G6-Fulvission
48	10.6.0.21	209.85.135.109	http://www.amazon.in/LG-G6-Fulvission
49	10.8.0.55	67.218.96.251	http://www.amazon.in/LG-V20

Figure.10: Pattern Data Clustering2

Index_No	Client_IP	Server_IP	URI_Steam
5	10.6.0.22	67.218.96.251	http://www.amazon.in/Samsung-inches-UA32K5100ARLXL
8	10.5.0.20	172.30.255.255	http://www.amazon.in/Samsung-inches-UA40K4100ARLXL
11	10.8.0.17	202.71.129.26	http://www.amazon.in/Samsung-inches-UA43K5200ARLXL
14	10.8.0.14	202.71.129.26	http://www.amazon.in/Samsung-inches-UA46K5300ARLXL
17	10.6.0.21	67.218.96.251	http://www.amazon.in/Samsung-inches-UA55K5400ARLXL
20	10.5.0.5	202.71.129.26	http://www.amazon.in/Samsung-inches-UA60K6100ARLXL
23	10.8.0.19	67.218.96.251	http://www.amazon.in/Samsung-inches-UA65K6500ARLXL
28	10.5.0.55	209.85.135.109	http://www.amazon.in/Samsung-inches-UA70K7100ARLXL
32	10.6.0.37	67.218.96.251	http://www.amazon.in/Samsung-inches-UA72K7200ARLXL
35	10.6.0.29	209.85.135.109	http://www.amazon.in/Samsung-inches-UA80K8100ARLXL
38	10.8.0.51	59.162.23.130	http://www.amazon.in/Samsung-inches-UA32K5100ARLXL
41	10.5.0.12	209.85.135.109	http://www.amazon.in/Samsung-inches-UA32K5100ARLXL
44	10.6.0.42	67.218.96.251	http://www.amazon.in/Samsung-inches-UA32K5100ARLXL
47	10.5.0.18	172.30.255.255	http://www.amazon.in/Samsung-inches-UA82K8200ARLXL
50	10.8.0.53	202.71.129.26	http://www.amazon.in/Samsung-inches-UA32K5100ARLXL

Figure.11: Pattern Data Clustering3

Index_No	Client_IP	Server_IP	URI_Steam
0	10.8.0.15	202.71.129.26	http://www.amazon.in/LG-inches-UA32K5100ARLXL
2	10.5.0.54	209.85.135.109	http://www.amazon.in/LG-inches-UA40K4200ARLXL
25	10.8.0.16	59.162.23.130	http://www.amazon.in/LG-inches-UA32K5100ARLXL
29	10.5.0.21	59.162.23.130	http://www.amazon.in/LG-inches-UA80K8100ARLXL

Figure.12: Pattern Data Clustering4

XI CONCLUSION AND FUTURE WORK

Data filtering perform by removing unwanted patterns from each record in the database. Since the pre-processing techniques performed is to mine the interesting patterns, the data end with *.jpg, *.gif, *.bmp be removed. In this paper I have implemented two level partitioning algorithms to create pattern cluster for dataset. This algorithm increase 6% of

performance and accuracy for pattern cluster. In future work I am going to implement security portion.

XII REFERENCES

[1]. P.Dhanalakshmi, Dr. K.Ramani, Dr.B.EswaraReddy,“The Research of Preprocessing and Pattern Discovery Techniques on Web Log files” -978-1-4673-8286-1/16 \$31.00, 2016 IEEE DOI 10.1109/IACC.2016.35

[2]. ShashiMehrotra, ShrutiKohli, “Comparative Analysis of K-Means with other Clustering Algorithms to Improve Search Result” - 978-1-4673-7910-6/15/\$31.00, 2015 IEEE

[3]. Dilip Singh Sisodia, ShrishVerma, “Web Usage Pattern Analysis Through Web Logs: A Review” - 978-1-4673-1921-8/12/\$31.00, 2012 IEEE

[4]. TingZhong Wang, D. Jin and S. Lin (Eds.), “The Development of Web Log Mining Based on ImproveK-Means Clustering Analysis” - Advances in CSIE, Vol. 2, AISC 169, pp. 613–618. Springer-Verlag Berlin Heidelberg 2012

[5] S.C. Satapathy et al. (Eds.), “Design and Implementation of an Effective Web Server Log Preprocessing System” - Proceedings of the InConINDIA 2012, AISC 132, pp. 897–905. Springer-Verlag Berlin Heidelberg 2012

[6]. J.MonishaPrivthyJeba,M.S.Bhuvanewari, K.Muneeswaran, “Extracting Usage Patterns from Web Server” - 978-1-4673-6615-1/16/\$31.00 © 2016 IEEE

[7]. Supinder Singh, SukhpreetKaur, “Web Log File Data Clustering Using K-Means and Decision Tree” - 2013, IJARCSSE All Rights Reserved

[8]. Ripal Patel, Mr. KrunalPanchal, Mr. DushyantsinhRathod, “Efficient Log Mining from Web Server Using Clustering Technique” - JJETIR1512016 - 2015

[9]. Mr. Dushyant B. Rathod, “Customizable Web Log Mining from Web Server Log” - IJEDR1302021

[10]. CiroGracia, Xavier Anguera, Xavier Binefa, “Two-level clustering towards unsupervised discovery of acoustic classes” – IEEE - 2013

[11]. Mr. DushyantsinhRathod, Dr. SamratKhanna, Mr. VijaykumarGadhavi “A Survey on Different Efficient Clustering Techniques used in Web” ISJRD - 2016

[12]. <http://www.w3.org/TR/WD-logfile.html>

[13]. <http://www.extratrend.com>