International Journal of Engineering & Technology, 7 (4.29) (2018) 40-44

International Journal of Engineering & Technology

Website: www.sciencepubco.com/index.php/IJET

Research paper

Predicting User Navigation in an Online Newspaper Site Using Association Rules Mining and Markov Model

Husna Sarirah Husin*

Universiti Kuala Lumpur Malaysia Institute of Information Technology (UniKL-MIIT)
**Corresponding author E-mail: sarirah@unikl.edu.my*

Abstract

This paper discusses an approach to predict Web pages from an online newspaper using association rules mining and Markov model decision process. We use a set of Web server logs from an online newspaper, process the logs using Web usage mining methodology, generate transaction files for association mining and predict the web pages using Markov decision model process. We found that users are reading articles from the same section and since majority of users only read one page in a session, it is hard to find associated news articles in a same session. However, the association between section pages are legit and can be used to model the Markov chain for the navigation.

Keywords: *association rules; Markov model; online newspaper; user navigation; Web usage mining*

1. Introduction

The advancement in mobile devices plays major part in driving high traffic to online news. Current research found that 85% of adults in America are reading news on the mobile device [1], [2]. The easy to use interface of the mobile device is one of the reasons why news content is among the most popular mobile content [3]. On the other spectrum, social media is becoming more significant in attracting traffic to the news website compared to other external referrers such as search engine and blogs [4]. However, online newspaper provider is facing fierce competition with blogs and news aggregator in getting users to access their news site. Therefore, understanding of users‘ navigation is vital in order to provide exceptional services to attract the users in browsing more pages in the site.

Today, web applications are motivated to engage their users and provide more personalised experience. Hence, it is important to create an interaction with users and to be in the know of their preferences when it comes to predicting the next accessed pages. For a newspaper company, usage mining using web logs allows the content provider to perform cross navigation strategies and to increase more pageviews. In other respects, usage data is pertinent to develop marketing strategies and promote the online news site to wider audience. Moreover, the information derived from the web logs enable the editorial team to provide best access routes to other pages that may interest the users. Knowledge obtained from web usage mining can be used to enhance web design, introduce personalization service and facilitate more effective browsing [5].

In Web usage mining, there are various data mining, machine learning and statistical analysis to obtain interesting knowledge. Association rules are used to identify the pages that are being accessed together is a session. Association rules can also be a basis to predict future requests by users. The major problem with association rules with large data items sets is the discovery of large number of rules and difficulty to identify the rules that will lead to correct [6].

Markov model is one of the many techniques that are used for modelling and predicting user’s navigational behaviour. To predict the most likely Web page a user will access, we can compare the current access sequence with the previously Web page accesses [7], [8]. However, for Markov models, low order Markov models have the deficiency in accuracy of predicting web pages because of not enough history. On the other hand, high order Markov models have high state complexity [9].

In this work, we propose to use association rules to sample long history for Markov state that provide ambiguous predictions. This is because that association rules can help those states to increase the accuracy rate for predictions. On the other hand, we use low order Markov models to keep low complexity as well as high coverage.

The findings of this study are beneficial because it assists the editorial team to identify the categories of news most read together and later to propose relevant content for the ease of access and reference [10]. Additionally, for news producers the result of the study may be beneficial because they can know the habits and pattern of news consumptions for predicting the target for future subscribers. This is in line with the objective to target specific aim of online readers [11].

2. Related Works

In this section, we review previous works on Web usage mining, and other works that have used association rules mining and Markov model in predicting web pages.

Copyright © 2018 Authors. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

2.1. Web usage mining

Web Usage Mining (WUM) is the use of data mining techniques to discover interesting usage patterns from Web server logs. The information from the usage patterns is important in order to understand the user behaviour. One of the important area in Web mining is mining navigation patterns [12], [13]. Generally, whenever user access a website, the user is interested to navigate the pages in the site. The history of navigation or number of pages within a period of time is referred to as a session. These sessions are extracted from the Web server logs which contains the sequences of pages that users have visited in a specified time [14].
Web usage mining consists of three phases; which are pre-processing, pattern discovery, and pattern analysis [14]. The pre-processing stage is to clean and filter the Web data in order to have a clean data set. This stage includes data cleaning, user identification, session identification and transaction identification [15].
Pattern discovery and analysis applies various methods and algorithms from statistical analysis and data mining. Some of the data mining techniques that can be used in Web usage mining are association rules, classification, clustering, sequential patterns and dependency modelling [13].

2.2. Association rules

Association rule is an established and well-known method to discover interesting relations between variables. The purpose is to identify strong rules discovered in databases using different measures of interestingness [16]. The idea of association rule is to solve the problem of market basket analysis in finding customer patterns from large transactional databases that will generate all significant association rules between items in the database [17]. Association rule learning typically does not consider the order of items either within a transaction or across transactions.
There have been many works that used the association rules mining techniques based on log data or clickstream data. There are work on predicting web pages [18], Web personalization [6], analysing online consumer behaviour pattern [19] and recommend web pages based on association rules of web logs data [20].
An association rule is an expression of $A \rightarrow B$, where A and B are itemsets and $A \cap B = \emptyset$. The inherent meaning of an association rule $A \rightarrow B$ is that transactions which contain the set of items in A tend to contain the set of items in B. The subset of A is the set of antecedent items, or the body of the rule, also known as left-hand side (LHS), while B is the set of consequent items, the head of the rule or the right-hand side (RHS).
For example, the following rule is extracted from the data set of web server logs of an online newspaper. The rule suggests that there is a strong relationship between the article A and article B; as users that view the article A, also view the article B. Each transaction itemset represents the set of pages that are accessed together in a single session. An association rule as shown above means that the session that has itemset of article A, are likely to also have itemset of page article B.

In the context of finding Web pages that are visited together, it is similar as to find associations among items in transaction databases or to find items that are bought together in a basket. A set of session is defined as S = { S1, ..., Sn} where n is the number of sessions in S and a set of URLs or page request is a number of URLs in S defined as U = { U1, ..., Un }.

2.3. Markov model

Markov model is based on a principle of "memorylessness", where the next state of the process only depends on the previous state. Markov models can be expressed in equations or in graphical models. Graphic Markov models typically use circles (each containing states) and directional arrows to indicate possible transitional changes between them. The directional arrows are labelled with the rate or the variable one for the rate. This probabilistic algorithm is applied in many diverse areas such as in the area of process control [21], acoustic surveillance method [22], bioinformatics [23], image processing [24] and many more.
Markov model is one of the many techniques that are used for modelling and predicting user's navigational behaviour. To predict the most likely Web page a user will access, we can compare the current access sequence with the previously Web page accesses [7], [8]. There have been many works performed on predicting the Web pages using Markov model [25]–[29]. Some work combined clustering before using the Markov prediction algorithm to predict the next possible states [30] .
In web prediction, the K^{th}-order Markov model is the probability that a user will visit the k^{th} page provided that she has visited the ordered k-1 pages. For example, in the second order Markov model, prediction of next web page the user visits are computed based only on the two web pages previously visited.
We can model the transition probabilities between discrete states by the aid of matrices. A discrete Markov Chain is a sequence of random variables $X_1, X_2, \ldots, X_n$, characterized by the Markov property, also known as memoryless property. The Markov property states that the distribution of the forthcoming state X_{n+1} depends only on the current state X_n and does not depend on the previous ones $X_{n+1}, X_{n-2}, \ldots, X_1$
The set of possible states $S = s_1, s_2, ..., s_r$ of X_n can be finite or countable and it is named the state space of the chain. The chain moves from one state to another and known as either transition or step. The probability p_{ij} to move from state s_i to state s_j in one step is named transition probability:

$$p_{ij} = P\,r(X_1 = sj|X_0 = si). \quad (1)$$

The probability of moving from state i to j in n steps is denoted by $p_{ij}^{(n)} = \Pr(X_n = s_j \text{—} X_0 = s_i)$.

3. Methodology

The dataset consists of Web server logs from a Malaysian daily online newspaper called *Berita Harian* Online for a day on Wednesday 5 April 2017. We adapt the Web usage mining and combine it with association rule mining and Markov modelling that is illustrated in Fig. 1.

The Web server log contains information of date and time of access, URL that user requests, the HTTP status code representing that the request is successful. A user agent string determines the platform and device of the user and finally the IP address of the user can be identified in the last part of the web server logs. Fig. 2 shows a sample entry of a *Berita Harian's* web server log.

Before cleaning, there are 657 238 requests for that day. Firstly, we chose the request with 200 status code that denote a successful request. Then, we remove unsuccessful requests (status code 300 ,400 and 500), robots, crawlers, images files, automated requests, and RSS requests. After cleaning process, the total log data set is reduced to 285 516. The Web server logs that we use do not have any data about the users and the users are not a subscriber and not logged-in, makes it almost impossible and very difficult to create a reliable user profile [31]. Therefore, to identify users, we combine the IP address and the user agent string that is type of browser and operating system. If a user has the same IP address, but different type of user agent, it is assumed that it denotes a different user [14].

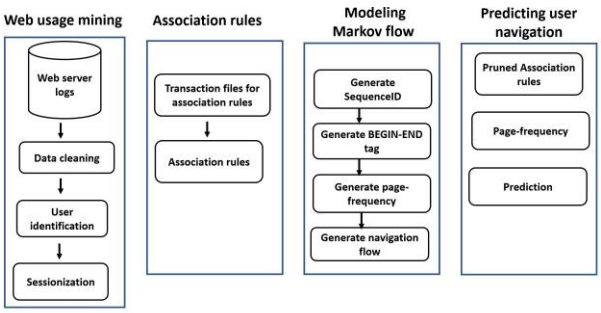


Fig. 1: Methodology of web prediction, consisting of Web usage mining, association rules and Markov modelling

```
[05/Apr/2017:03:36:01 +0000] GET / HTTP/1.1 200 26865
http://www.bharian.com.my/node/267485 Mozilla/5.0 (iPhone; CPU iPhone OS
10_2_1 like Mac OS X) AppleWebKit/602.4.6 (KHTML, like Gecko) Version/10.0
Mobile/14D27 Safari/602.1 183.171.70.83, 23.15.10.108

[05/Apr/2017:03:36:01+0000] GET /node/267777 HTTP/1.1 200 38354
http://m.facebook.com/ Mozilla/5.0 (Linux; Android 6.0.1; SM-J700F
Build/MMB29K; wv) AppleWebKit/537.36 (KHTML, like Gecko) Version/4.0
Chrome/56.0.2924.87 Mobile Safari/537.36 [FB_IAB/FB4A;FBAV/111.0.0.18.69;]
175.144.25.80, 203.106.85.112, 23.15.10.102

[05/Apr/2017:03:36:02+0000] GET /sukan?page=1 HTTP/1.1 200 43573 Mozilla/5.0
(Linux; Android 6.0; PLK-L01 Build/HONORPLK-L01) AppleWebKit/537.36 (KHTML,
like Gecko) Chrome/56.0.2924.87 Mobile Safari/537.36 115.164.190.165,
23.45.232.156
```

Fig. 2: A sample entry of Berita Harian's Web server log

Session identification is intended to divide the page accesses of each user into individual sessions. The literature suggests using a timeout. In our studies, we adapt a five minutes time out. A session is over after five minutes or more and the next page viewed is then assumed to be a new session [32]. We categorize the web pages to three main categories. The first category is the main page or the index page of the *Berita Harian* Online website. Example of the URL for mainpage is *GET/HTTP/1.1* The second category is the section page. There are twelve section pages in the news website. This include National, Sports, Business, Entertainment, World and many more. Example of URL section pages are *GET/duniaHTTP/1.1, GET/sukanHTTP/1.1, GET/jenayahHTTP/1.1*. Finally, the article page is denoted by numbers such as follows *GET/node/268094* and *GET/node/268291*.

3.1. Data and generating association rules

When the data is cleaned, we create the transaction file listing all the pages in sequence that are accessed by users. Next, we generate the association rules using the package *arules* for Apriori association rule in R. In reference to transactions on the Web, association rules capture the relationships between the pages. For example, for pages of article A and article B is {A, B} ⇒ {C} [support=0.01, confidence=0.5] represents the relationship of users that access article A and article B pages whom also accessing the page article C with confidence of 50%. The support value represents that itemset {A, B, C} was present in 1% of user sessions.

Table 1: Snippet of sequence ID, begin and end flow

UserID	Session	Page URL	Seq ID	SeqID-Begin	Seq-ID End
1	1	267984	70	BEGIN	
1	1	267982	70		
1	1	268135	70		END
2	1	267918	71	BEGIN	END
3	1	268164	72		
3	1	267977	72		END
4	1	267984	73	BEGIN	END
4	2	267918	74	BEGIN	

3.2. Modelling the data set using Markov flow

From the Web server log records, we build a Markov flow model as a representation of the user's navigation path based on the users' visit recorded in the web server logs. Hence, we can model the transition probabilities between Web pages. We construct a graph from the Web log file, consisting nodes representing the web pages, links representing the hyperlinks and weights on the link representing the probability of traversals on the hyperlinks.
The following explains how Sequence ID are added to the Web server logs after the logs have were cleaned, users identified, and session identified:

- Add sequence ID

The sequence ID is created based on the user and the pages in each individual session. A new sequence ID is generated for every new user. If the same users have a new session, another sequence ID is made. From the Table 1, User 1 has a Sequence ID 70, and the following new user of User 2 has Sequence ID 71. User 4 has two sessions. The first time User 4 appears, he will get Sequence ID 73. Since he has a second session, the next Sequence ID is 74 because it is a new session, although from the same user.

- Add 'BEGIN' and 'END' for a session

We set a RowIDfor each row having a new user. For example, if User 1 three rows of page access, he will have RowID of 1,2,3. A new RowID will be generated for a new user.

- 'BEGIN' tag

A BEGIN tag will be generated for every Sequence ID and an 'End' tag will be created after the session of a particular user ended as shown in Table I.

- Calculate frequency of each flow.

From there, we calculate the total count of the beginning of each type of page and the end navigation of that page. Since most of the access is a single page access, we only select session with at least two pages.

3.3. Integrating association rules and Markov model

After we have obtained the association rules; which are based on a targeted support and confidence value, we use it as a benchmark for the navigation flow using Markov. After we get the total frequency of the pages, we only select those pages that are in the rules mined

4. Results

First, we prepare transaction files for association rules for all the pages; which includes the mainpage, the section page and the article page. We set the support value to 0.01 and Fig. 3 illustrates the item frequency plot for the pages. From Fig. 3, we can see that the highest page is the mainpage, followed by the Sports section page. The subsequent eight top items are also from section pages. The article page of URL 267984, followed by another article of URL 268164. We used *arules* package in R to generate the association rules mining with support 0.005 and confidence of 0.02 and only the associations between the section pages appeared such as shown in Table 2.
The support value that was higher than 0.005 did not yield any results. The rules found are all section pages with the section World (*dunia*) and Entertainment (*hiburan*) being the sections highly associated. The rules found shows that readers are inter-navigating the section pages.

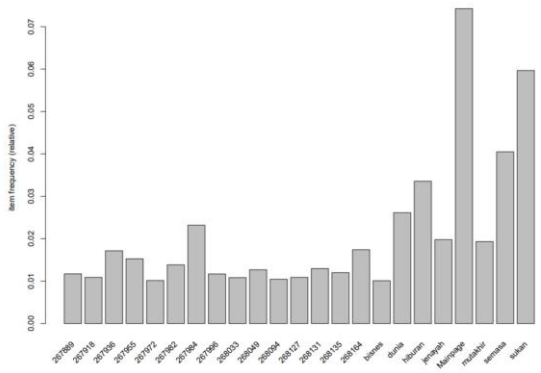


Fig. 3: Item frequency plot of pages of mainpage, section pages and article pages with support value = 0.01

Table 2: Top 5 Association rules for all pages on Wednesday 5 April 2017 with support = 0.005, confidence=0.02

association rules	support	conf	lift
{dunia} ⇒ {hiburan}	0.006102	0.2334	6.9546
{semasa} ⇒ {dunia}	0.007932	0.1958	7.4923
{dunia} ⇒ {sukan}	0.007419	0.2837	4.7555
{hiburan} ⇒ {semasa}	0.007804	0.2325	5.7416
{sukan} ⇒ {hiburan}	0.007162	0.1200	3.5760

4.1. Association rules for articles in Wednesday 5 April 2017

There are no articles pages found in this association rule mining with support of 0.005. Therefore, we further set a lower support value to find the articles are read together in a session. We run the R script again and set the support value lower to 0.0005 and finally found the following rules for article pages as shown in Table 3.

Since the articles are denoted by number, we examine the article titles of the URL found from the association rule mining and list the article titles in Table 4. Fig. 4 shows the item frequency plot of the articles on 5 April 2017. Based on the item frequency plot, we can infer that the highest articles are not in the top association rules. For the articles on 5 January 2017, the top article is a Sports news article, and the second highest article is from National news article. However, the top two association rules are on Entertainment articles. From the association rules found, we can ascertain that users like to read articles of the same section. The article titles are listed in Table 4.

Table 3: Association rules for article page on Wednesday 5 April 2017

association rules	support	conf	lift
{268012} ⇒ {267889}	0.000515	0.1191	6.7255
{267889} ⇒ {267911}	0.000551	0.0311	4.9776
{268033} ⇒ {268049}	0.000508	0.0243	1.4585
{268049} ⇒ {268094}	0.000754	0.0453	2.4861
{268135} ⇒ {268131}	0.00118	0.0518	2.2927
{267982} ⇒ {267984}	0.000646	0.0238	0.3445
{267955} ⇒ {267936}	0.000863	0.0238	0.6362

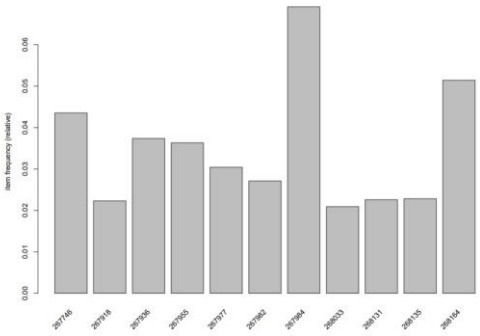


Fig. 4: Item frequency plot of articles that were accessed with support value = 0.0005 and confidence = 0.01

Table 4: Article titles found in association rules for 5 April 2017

No	Article	Title	Total
1	268012	Amelina akur kemunculan semula kurang impak	722
2	267889	Sara Ali buka tudung?	2749
3	267911	Wujud Instagram anak elak penipuan - Stacy	970
4	268033	Harga minyak dunia naik	3169
5	268049	Seksyen 4(5) SOSMA diluluskan	2624
6	268131	United hampir diaibkan Everton	3400
7	268135	Leicester tambah derita Moyes	3449
8	267982	Gomez belum sah jurulatih Harimau Malaya - TMJ	4006
9	267984	FAM perlu RM40 juta	10192
10	267955	Lelaki maut motosikal langgar tembok	5374

4.2. Markov model of articles accessed by users

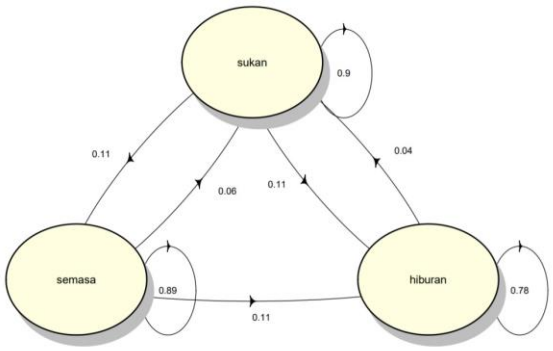


Fig. 5: Markov model of section pages that were found from association rules mining

Fig. 5 shows the Markov model of all the three section pages that were mined from the association rules. We are only taking in the section pages because the support value for article pages were very low; which means that the association between two or more article pages is almost to none. This may be due to the reason that users only read one article page in a session. Thus, it is very difficult to find strong associated relationship between articles. From Fig. 5, we can see very strong association between the same section pages with itself. This means, that users have very high probability to read the same sections in the same session. For example, there is a high chance that users reading Sports (*sukan*) section, will read the same section page within a session.

4.3. Simulating the Markov model

To create a Markov chain, we set the initial state to the first row as found in the transition matrix from the association rules; which are *sukan*=0.9, *semasa*=0.06, *hiburan*=0.04 as shown in Fig. 6. For the first order Markov Chain the case is different because the current state depends only on the previous state. Given that points clear, a second order Markov Model will be a model that reflects that the current state only depends on the previous two states before it.

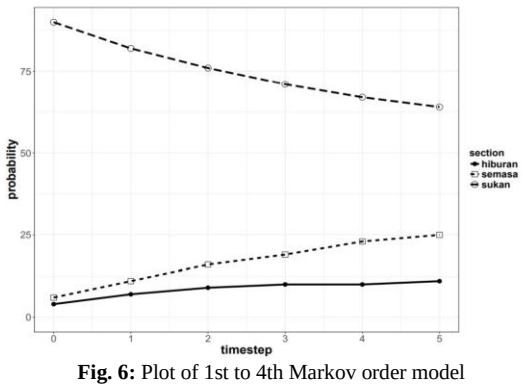


Fig. 6: Plot of 1st to 4th Markov order model

5. Conclusions and Future Work

In this paper, we present a usage navigation pattern prediction system using the web server logs collected from an online newspaper website. We combined the Web usage mining methodology with association rules mining and Markov modelling. In the first stage of web usage mining, the web server logs that were collected were cleaned and formatted, identify users based on user-agent string, perform sessionization and lastly identify the page transaction for each user. In the association rules mining, we mined the transaction file using R *arules* package. On the other hand, we also processed the web server logs and assigned sequence ID and BEGIN-END tag for each of the sessions. From this, we calculated the frequencies of users' traversal path from one page to another in order to generate the Markov modelling.

Based on the association rules and the Markov model, we integrate these two methods to predict Web page accesses. This integration is based on a low order Markov model. The association rules mined are used as a basis to generate the Markov model for resolving ambiguous predictions by the use of long history data. The integration will avoid complexity in high order Markov model, as well as the limitation of short history. In addition, this model reduces the large number of association rules.

From the association rules, we can prune redundant rules and only use the rules that satisfy the support and confidence value that were set. The support value that was set to 0.005 and confidence of 0.02 only resulted to association rules with the section pages. The highly association sections are between the World (*dunia*) and Entertainment (*hiburan*) sections, followed by Current (*semasa*) and World (*dunia*). The third pair of section pages of strong association is between Sports (*sukan)* and World (*dunia*) sections. We can infer that the World (*dunia*) section is particularly popular among the news reader as it is strongly related with other sections as well. To find the association between articles, we went further down to 0.0005. We found that the strongly associated articles are within the same section, with news articles from Entertainment (*hiburan*) being the most strongly related.

As for Markov model, we based on the section pages because the articles change every day; as compared to the section pages that stay the same.

Acknowledgements

Appreciation and many thanks to Aswandi Yassin from *Berita Harian* for assistance in data collection and much appreciation and acknowledgment to Universiti Kuala Lumpur for the conference grant.

References

[1] S. Wang, "85 percent of americans use mobile devices to access news and seniors are driving that number up," Nieman Lab, 2017. [Online]. Available: http://www.niemanlab.org/2017/06/85-percent-of-americans-use-mobile-devices-to-access-news-and-seniors-are-driving-that-number-up/. [Accessed: 01-Jan-2018].

[2] D. Z. Morris, "85% of U.S. Adults Read the News on a Mobile Device," Fortune. [Online]. Available: http://fortune.com/2017/06/18/mobile-news-reading-pew/. [Accessed: 01-Jan-2019].

[3] H. Shim, K. H. You, J. K. Lee, and E. Go, "Why do people access news with mobile devices? Exploring the role of suitability perception and motives on mobile news use," Telemat. Informatics, vol. 32, no. 1, pp. 108–117, 2014.

[4] C. Castillo, M. El-haddad, and M. Stempeck, "Characterizing the Life Cycle of Online News Stories Using Social Media Reactions," 2012.

[5] R. Omar, A. O. M. Tap, and Z. S. Abdullah, "Web usage mining: A review of recent works," in In Information and Communication Technology for The Muslim World (ICT4M), 2014 The 5th International Conference, 2014, pp. 1–5.

[6] B. Mobasher, H. Dai, T. Luo, and M. Nakagawa, "Discovery and Evaluation of Aggregate Usage Profiles for Web Personalization."

[7] V. N. Padmanabhan and J. C. Mogul, "Using Predictive Prefetching to Improve World Wide Web Latency."

[8] P. L. T. Pirolli and J. E. Pitkow, "Distributions of surfers' paths through the world wide web," WWW, vol. 2, no. 1, pp. 29–45, 1999.

[9] F. Khalil, J. Li, and H. Wang, "A Framework of Combining Markov Model with Association Rules for Predicting Web Page Accesses."

[10] K. Olmstead, A. Mitchell, and T. Rosenstiel, "Navigating News Online: Where People Go, How They Get There and What Lures Them Away," Journalism, vol. 9, 2011.

[11] O. Westlund, "News consumption in an age of mobile media : Patterns , people , place , and participation," 2016.

[12] P. Bari and P. Chawan, "Web usage mining," J. Eng. Comput. Appl. Sci., vol. 2, no. 6, pp. 34–38, 2017.

[13] J. Srivastava, R. Cooley, M. Deshpande, and P.-N. Tan, "Web usage mining: Discovery and applications of usage patterns from web data," ACM SIGKDD Explor. Newsl., vol. 1, no. 2, pp. 12–23, 2000.

[14] R. Cooley, B. Mobasher, and J. Srivastava, "Data preparation for mining world wide web browsing patterns," Knowl. Inf. Syst., vol. 1, no. 1, pp. 5–32, 1999.

[15] B. Liu, B. Mobasher, and O. Nasraoui, "Web Usage Mining," in Web Data Mining, Springer, 2011, pp. 527–603.

[16] G. Piatetsky-Shapiro, "Discovery, analysis and presentation of strong rules," in Knowledge discovery in databases, 1991, pp. 229–248.

[17] R. Agrawal, T. Imielinski, and A. Swami, "Mining association rules between sets of items in large databases," in ACM SIGMOD Record, 1993, p. 207–216.

[18] S. G. Langhnoja, M. P. Barot, and D. B. Mehta, "Web Usage Mining Using Association Rule Mining on Clustered Data for," vol. 2, no. 1, 2013.

[19] O. Raphaeli, A. Goldstein, and L. Fink, "Analyzing online consumer behavior in mobile and PC devices: A novel web usage mining approach," Electron. Commer. Res. Appl., vol. 26, pp. 1–12, 2017.

[20] D. Mistry, K. J. Sharma, and S. A. Patel, "Recommend websites through weblog files using association rule," Int. J. Comput. Appl., vol. 126, no. 2, pp. 1–8, 2015.

[21] M. D. G. García-hernández, J. Ruiz-pinales, and A. Reyes-ballesteros, "Association Rule-Based Markov Decision Processes."

[22] K. M. Jeon, D. Y. Lee, H. K. Kim, and M. J. Lee, "Acoustic surveillance of hazardous situations using non-negative matrix factorization and hidden Markov model," in Audio Engineering Society Convention, 2014, pp. 1–5.

[23] V. Narasimhan, P. Danecek, A. Scally, Y. Xue, C. Tyler-Smith, and R. Durbin, "Bcftools/roh: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data," Bioinformatics, vol. 32, no. 11, pp. 1749–1751, 2016.

[24] M. Z. Uddin, J. Torresen, and T. Jabid, "Human activity recognition using depth body part histograms and hidden Markov models," in Innovations in Science, Engineering and Technology (ICISET), 2016, pp. 1–4.

[25] M. Awad and I. Khalil, "Prediction of User â€TM s Web-Browsing Behavior : Application of Markov Model Prediction of User ‟ s web - browsing behavior : Application of Markov Models," no. August 2015, 2012.

[26] M. Deshpande and G. Karypis, "TR 00-056 Selective Markov Models for Predicting Web-Page Accesses," 2000.

[27] X. Dongshan and S. Junyi, "A New Markov Model For Web Access Prediction," Comput. Sci. Eng., vol. 4, no. 34, 2002.

[28] J. Zhu, J. Hong, and J. G. Hughes, "Using Markov models for web site link prediction," in roceedings of the thirteenth ACM conference on Hypertext and Hypermedia, 202AD, pp. 169–172.

[29] R. R. Sarukkai, "Link prediction and path analysis using Markov chains," Comput. Networks, vol. 33, no. 1, pp. 377–386, 2000.

[30] S. Sorout, "Web Page Anticipation System Using Markov Model," Int. J. Adv. Res. Comput. Sci., vol. 8, no. 7, pp. 867–872, 2017.

[31] M. Trevisiol and L. M. Aiello, "Cold-start News Recommendation with Domain-dependent Browse Graph Categories and Subject Descriptors," 2012.

[32] P. Huntington, D. Nicholas, and H. R. Jamali, "Website usage metrics : A re-assessment of session data," vol. 44, pp. 358–372, 2008.