

Intelligent Systems of Machine Learning Approaches for Developing E-Services Portals

Waleed M. Ead^{1,*}, Mohamed M. Abbassy¹

¹Faculty of Computers and Artificial Intelligence, Beni-Suef University, Egypt.

E-Mail: Waleeadead@bsu.edu.eg, Mmabbassy@fcis.bsu.edu.eg

Abstract

This paper provides a framework for intelligent servers that provide e-services on portals and applications for e-commerce purposes, such as an intelligent server that can dynamically plan and structure to respond to future users' needs and provide the appropriate e-services at the right time. To figure out how web users learn the log files collected from web users' connections and the internet, we use data mining techniques. In our experimental study, we use real data sets. Different data mining techniques are also used for Developing Intelligent E-Services Portals in future work, and it is a very difficult task in big data to identify regular trends. As the duration of the trends has to be discovered to expand the analytical space that grows exponentially.

Keywords: Intelligent Servers, Web Logs, Data Mining, E-Services, E-Commerce, energy-saving applications, Machine Learning Approaches.

Received on 30 October 2020, accepted on 25 November 2020, published on 02 December 2020

Copyright © 2020 Waleed M. Ead *et al.*, licensed to EAI. This is an open access article distributed under the terms of the [Creative Commons Attribution license](#), which permits unlimited use, distribution, and reproduction in any medium so long as the original work is properly cited.

doi: 10.4108/eai.2-12-2020.167292

*Corresponding author. E-mail: Waleeadead@bsu.edu.eg

1. Introduction

Intelligent web server growth, particularly in e-commerce applications, has been more popular in intelligent business services. Intelligent servers can change their architectures automatically to satisfy potential product requirements. The layout of the smart cloud server is seen in Figure 1. We've got a customer insight at the edge of the infrastructure to have correct knowledge at the right moment for the right users. Web usage mining is the task of detecting the behavior of web users while they browse the web. The aim is to identify site users' navigation preferences. The online portals can be customized, the network layout is enhanced, and web server efficiency is optimized. Cloud mining activities is displayed in Figure 2.

1.1 Web User Identification

According to [1], it shows the web user identity and identifies the most difficult measures. Such a dilemma relies on the

purpose of web mining operations to be carried out. For brief periods, the IP user address may have an appropriate response. There are two types of specific detection strategies[2]; proactive strategies and reactive methods. The proactive approaches are structured to distinguish users before or after accessing a website. Reactive methods aim to link individuals to log entries after the log has been written [3].

The rest of the paper is organized as the detailed related work in section 2. The detailed experimental design is discussed in section 3. The proposed algorithm is stated in sections 4, 5 and 6, followed by the conclusion stated in section 7.

2. Related Work

Web usage mining is a hot research era in the field of web mining. The World Wide Web (Web for short) expansion has generated a large amount of data stored on the logs of the web servers of the web sites. Different hidden patterns can be discovered, such as user navigational patterns and the frequently visited web pages or page sets. etc... The

information obtained from weblog data can be used for various purposes.

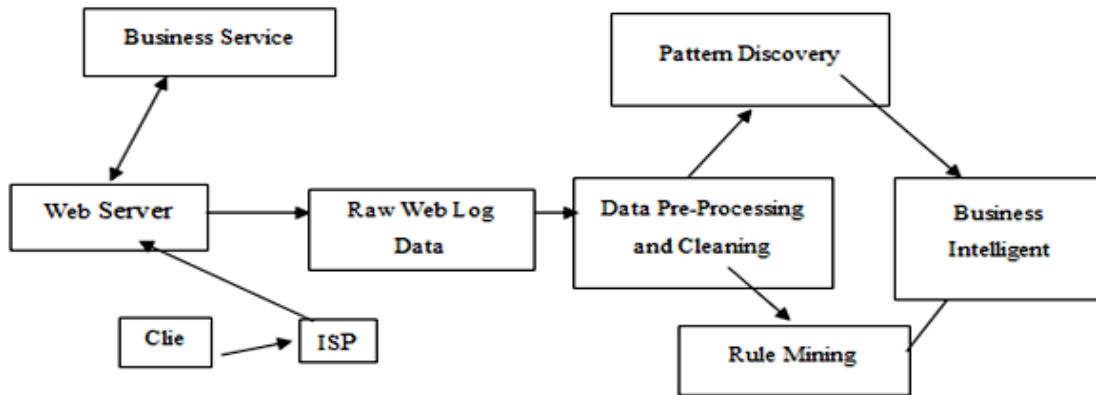


Figure 1. The proposed architecture of intelligent server

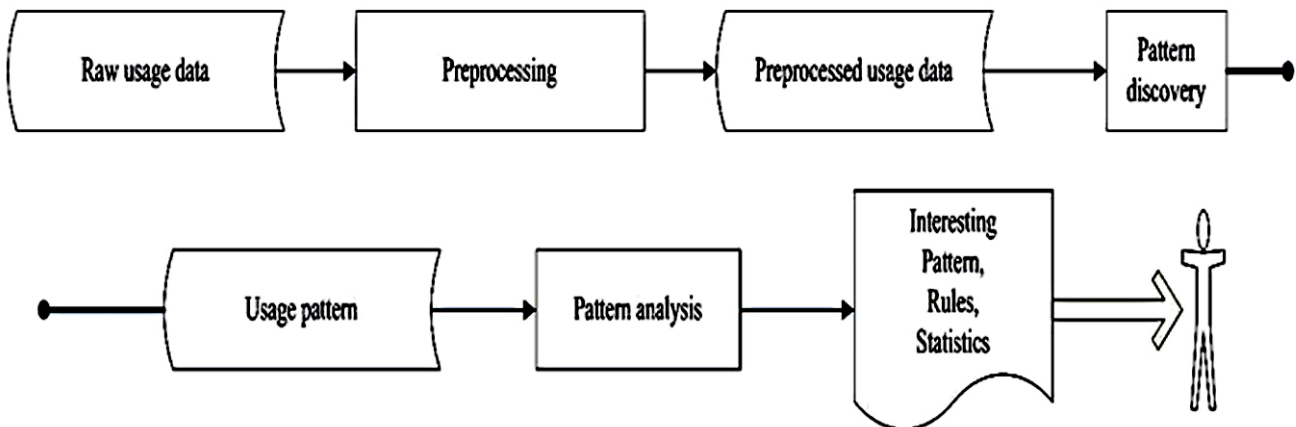


Figure 2. web usage mining tasks

[1] implements the association rule mining techniques that can help find the pages that are visited together even if they are not directly connected, revealing the correlations between groups of users sharing a specific event/interest [4-5]. The derived hidden patterns can restructure web sites by adding links between those pages that are visited together. Another issue of the web mining technique is web user clustering to identify the categories to discover user profiles. Several studies have been suggested to analyze the nature of user navigation. [6] A clustering partitioning approach is introduced that visualizes the user's navigation routes in-cluster. [7] An expectation-maximization (EM) algorithm has been suggested to find the parameters' highest likelihood in a probabilistic model. The function relies on unobserved latent variables to evaluate user navigation patterns. [8-9] introduced a guideline framework for navigation pattern mining by website usage mining to forecast future user movements.

3. Experimental Design

We are mainly interested in users' navigation to websites. The Web servers also record information about web access in a log file when users visit a website. Each logfile record is a website request performed by a web user on the server. Each logfile record/raw usually includes the following information: IP (Internet Protocol) address of the user, connection date and time, page URL (Uniform Resource Locator) requested, request protocol, request status code, and page size (if the request is successful). It also contains the different Web user requests, as shown in Figure 3.

```
190.144.44.36- 2009-08-26:00:06:34 "GET
/logs/administrator/components/com_virtuemart/export.php?
mosConfig_absolute_path=http://filtry.in.ua/components/co
m_virtuemart HTTP/1.1" 404 255 "1.109.250.29" 2009-08-
```

```

26                               10:22:06,"GET
/show_news.php?cutepath=http://filtry.in.ua//components/co
m_virtuemart HTTP/1.1",404,211,,"Mozilla/5.0"
"1.109.250.29",2009-08-26       10:22:06,"GET
/logs/show_news.php?cutepath=http://filtry.in.ua//componen
ts/com_virtuemart HTTP/1.1",404,216,,"Mozilla/5.0"
    
```

Figure 3. Examples of web user’s request

In table 1 below, this indicates a web user request representation in a log file. The different status code types that may appear in a log file are in table 2.

Table 1. Web User’s Request Interprets

Attribute	Value
IP	190.144.44.36
Date and time	2009-08-26:00:06:34
URL(referrer)	http://filtry.in.ua
Requested page	/logs/administrator/components/com_virtuemart/export.php?mosConfig//components/com_virtuemart/
Request protocol	GET
Status code	404
Size of page	255

Table 2. Status Code Categories

Status Code	Description
1xx	Informational
2xx	Success
3xx	Redirection
4xx	Client Error
5xx	Server Error

One of the five answer groups is specified in the first digit of the status code. Microsoft IIS may use added decimal sub-codes to provide more specific information, but they are not included here. For example, 2xx each has a group of subclasses from 200 to 207, in which each represents a valid HTTP (Hypertext Transfer Protocol) request, for example, 200(OK). The pre-processed log file that is ready for mining work is shown in Table 3.

3.1 Navigation Patterns Mining

A user navigation pattern is typical for a group of users’ navigations. The web mining clustering techniques will be used to find the hidden pattern of the users’ navigations. Furthermore, the data obtained from the log file cannot be used directly. Therefore, it must be preprocessed in an appropriate format for mining tasks, as previously mentioned. Table 4 displays log file contents obtained and used for the analysis by the webserver www.aceonlinestore.com.

Table 3. Preprocessed Log File

Trans ID	Client IP	Datetime	Method	Server IP	Port	URL Request
1	99.141.48.225	26/Aug/2009 : 00:00:20	GET	64.221.88.169	80	/images
2	99.141.48.225	26/Aug/2009 : 00:00:22	POST	64.221.88.169	80	/cgi-bin
3	61.250.92.97	26/Aug/2009 : 00:01:49	GET	64.221.88.169	80	/index.php
4	72.11.145.8	26/Aug/2009 : 05:14:18	GET	64.221.88.169	80	/administrator/components
...

Table 4. Contents of The Log File Used

# of records	# of users	# of sessions
21412	730	2465

In Table 5, the dataset contains about 184 URLs. Each URL address is first assigned to sequential numerical values.

Table 5. URLs Address Assigned to A Numeric Value

Index	URLs Address
0	/administrator/components/com_jjgallery/admin_jjgallery_p
1	/catalog/
2	/components/
3	/images/items/
....

3.1.1 Session identification

A user session is a finite collection of websites that the same user accesses during a particular visit. According to [7], [10] defines a user session as a collection of user accesses from the same IP address during a predefined period. When the period between requests approaches a certain amount, it may assume that the user is starting a new session. So, we use a 30-minute default timeout.

A unique number $j=1$ if the Web is made up of n pages. Every URL shall contain n . A user sits then defined by an n -dimensional vector that displays the user's preferences in the j -th web page with the j -th term. In various forms, a web page may be described to interest levels. In this analysis, the degree to which a URL meaning is mainly linked to the extent of website use (number of pages visited / total number of pages during the sitting) and the time spent on the user's link is considered. Formal, a vector represents the i device session $s^{(i)}=(s_1^{(i)}, s_2^{(i)}, s_3^{(i)}, \dots, s_n^{(i)})$ with the property:

$$s^{(i)} = \begin{cases} f_j^{(i)} t_j^{(i)} & \text{if the user visits the } j\text{-th URL} \\ 0 & \text{otherwise} \end{cases}$$

For $j=1 \dots n$, while the connection duration and time spent for the consumer on the j -th page in i th session are displayed by $f_j(i)$ and $t_j(i)$, respectively. The set of N sessions is defined from the log info following the pre-processing step. Table 6 represents the URL requests for each session that are useful for the clustering phase are set to a range of frequencies.

Table 7 demonstrates the assigning of binary representation to the association between various URL requests for each URL request that occurs in each session. Such pre-processing in table 7 will be useful to discover the frequent correlated patterns.

Table 6. URLs assigning for clustering

URL Req.	1	2	3	4	5	6	...
Sessions	URLs Number appear in the sessions						
1	0	4	8	0	58	0	...
2	0	0	0	0	0	0	...
3	0	0	2	0	32	0	...
4	0	1	2	5	53	1	...
...

Table 7. URLs assigning for binary code for finding an association between requests

Sessi on	URLs Number appear in the sessions						
1	0	1	1	0	1	0	...
2	0	0	0	0	0	0	...
3	0	0	1	0	1	0	...
4	0	1	1	1	1	1	...
...

4. Clustering Algorithms

Data clustering is an efficient means of identifying data similarities and merging related data [11]. Any clustering techniques are adequate for the number of clusters. The algorithm tries to divide the data into the number of clusters defined in K-means in this situation. For certain instances, the number of clusters found from the beginning is not enough. Rather, the algorithm starts with searching the first big cluster and then seeks the second, etc. The changed k-means are displayed and explained.

4.1 K-means clustering

Clustering costs are dependent on the distance between the Euclidean vector x_k in group j and the related cluster / group-center c_i , identified by [12]:

$$J = \sum_{i=1}^c J_i = \sum_{i=1}^c \left(\sum_{k, x_k \in G_i} \|X_k - c_i\|^2 \right)$$

Where $J_i = \sum_{k, x_k \in G_i} \|X_k - c_i\|^2$ is the cost function in group i . In the $c \times n$ binary membership matrix, U determines the clustered groups where the variable u_{ij} is 1 if the j th data point x_j is the group i and 0. The decreasing u_{ij} is as follows after the cluster centers are set [12]:

$$u_{ij} = \begin{cases} 1 & \text{if } \|X_j - c_i\|^2 \leq \|X_j - c_k\|^2, \text{ for each } k \neq i \\ 0 & \text{otherwise.} \end{cases}$$

$$u_{ij} = \begin{bmatrix} 1 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & 1 \end{bmatrix}$$

Then x_j belongs to group i because c_i is the nearest, closest to all centers. On the other side, the optimum core c_i that minimizes the mean of all the vectors in cluster i [12] when the membership matrix is set, i.e., when u_{ij} is defined:

$$c_i = \frac{1}{|G_i|} \sum_{k, x_k \in G_i} X_k$$

Where $|G_i|$ is the size of G_i , or $|G_i| = \sum_{i=1}^n u_{ij}$.

The algorithm will be provided in data set x_i , $i = 1, \dots, n$; it will then evaluate the cluster centers c_i and its membership matrix U using the stages as follows:

Stage 1: Dynamically select c points from data dots to initialize the cluster center c_i .

Stage 2: Identify the matrix U of the membership.

Stage 3: Measure the cost feature to avoid if a certain resistance value is either underneath.

Stage 4: Cluster Core Change. Move to stage 2.

It must run many times because the K-means output relies on the cluster centers' initial position. Clustering involves an attempt to reduce each cluster's expense function. The measurement of the algorithm is performed with the test set's precision. The measurement data vectors are then allocated to their respective clusters according to each vector's distance and each cluster core after determining the cluster centers. An error calculation is measured, and for that reason, the root implies square error (Root Mean Square Error-RMSE). The algorithm is consistently evaluated for better results.

4.2 Modified K-means clustering

The enhanced K-means clustering for very broad datasets is a scalable analytics cluster algorithm. This is based on the K-mean algorithm; it may be categorical as well as constant. This needs only one data transfer. Two phases [12] have to be taken:

- For certain small sub-clusters, the Pre-cluster records are seen in Figure 4.

- The Cluster of the sub-clusters occurring from the pre-cluster phase through the number of clusters required. The number of clusters may also be identified automatically.

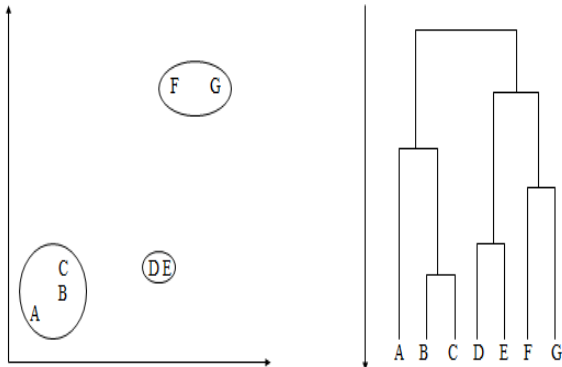


Figure 4. Shows the pre-cluster step of MK-means

For conceptual and range areas, it uses a log-like distance calculation. This is a probability-based distance. The difference between the two clusters is correlated with the log structure decrease as it is merged into one cluster. For the computation of log-likely, regular distributions for field ranges and multiple distributions for symbolic fields are presumed. The areas must always be autonomous, and the archives will be separate. The i to j class has [13]:

$$d(i, j) = \xi_v = -N_v \left(\sum_{k=1}^{K^A} \frac{1}{2} \log(\sigma_k^2 + \sigma_{vk}^2) \right)$$

Where,

K^A is the number of range type input fields,

N_v is the number of records in cluster v ,

σ_k^2 is the estimated variance of the k th continuous variable

for all records,

σ_{vk}^2 is the estimated variance of the k th continuous variable

for records in the v th cluster, and

$\langle i, j \rangle$ is an index representing the cluster formed by combining clusters i and j .

However, the improved k-mean uses a two-stage mechanism that deals with a clustering algorithm system to evaluate the clusters' numbers automatically. It is considered that k-means need more time than modified k-means due to the number of iterations to find the optimal clusters.

4.3 Kohonen Clustering Algorithm

The model Kohonen [13] consists of the input and output layers, two layers of neurons or groups. The input layer is fully related to the output layer and the weight of each link.

Another way of approaching is the network structure. The Kohonen model parameters are represented as weights between input units and output units or as cluster core connected to each output unit. Output records are inserted into the network and cluster centers: The clusters are spatially organized in a two-dimensional matrix. Each record impacts the entity (cluster) to which it is allocated and the winning team's neighborhood units, as shown in Figure 5.

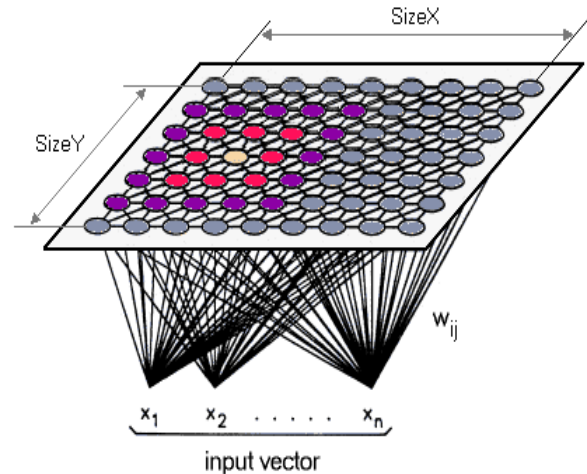


Figure 5. Kohonen network model [13]

Distance in a Kohonen network is calculated through a Euclidean between the encoded variable and the cluster core for the output unit. The reference k th for the reference database is set, and the output unit j th is set to the input field k th. The Euclidean distance between the output unit variable (the unit's center) and the input variable is the activation of the output function. Recollect that the winning machine for Kohonen networks is the lowest driven production unit. In addition to other forms of neural networks, enhanced activation is a stronger response.

$$d_{ij} = \sqrt{\sum_k (x_{ik} - w_{jk})^2}$$

X_{ik} is the value of the k th input field for the i th record, and w_{jk} is the weight for the k th input field on the j th output unit.

5. Clustering algorithms experimental results

Finding a relative navigation pattern can improve the content on clustered websites/pages, particularly in a website for a customer interested in such pages.

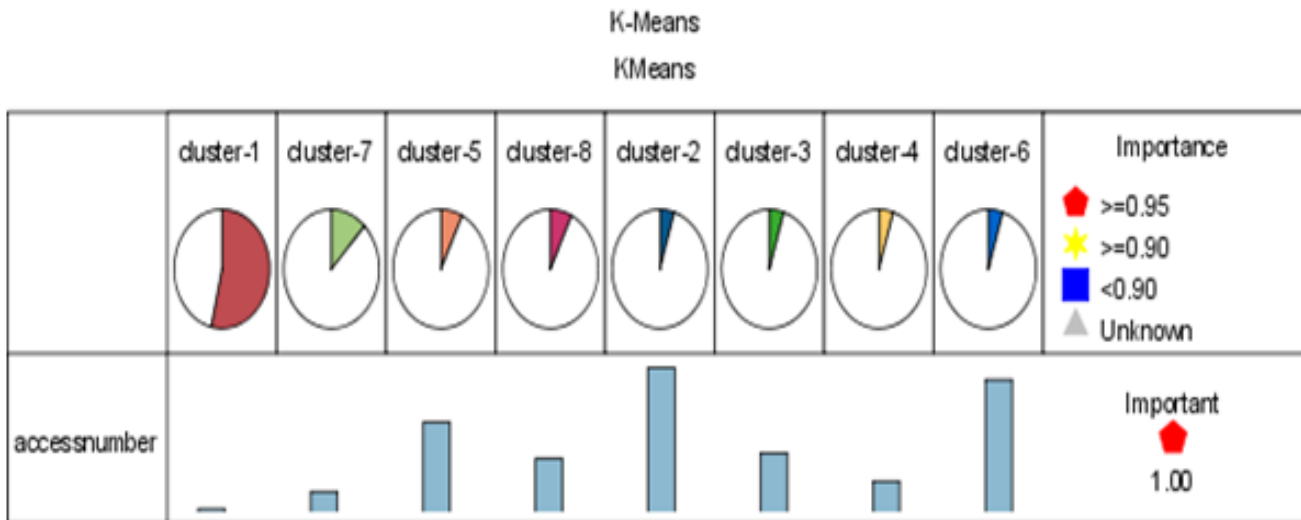


Figure 6. Cluster numbers

This is one of the targets to make www adaptable to its users and leads us on its own to the smart server. By the number of accesses to the URL requests, we cluster our data into 8 classes using the K-means algorithm, and figure 6 shows the clusters. The Importance column indicates the field's overall importance to the model. This result will appear as 1 minus the p-value (probability value was used to measure significance from the t-test or chi-square test). The number of accesses per cluster is shown in Table 8. Two clusters have the highest access numbers, Cluster 2 and Cluster 6, so that the webmaster takes both clusters into account to update it continuously. To make the sites in such clusters more appealing to consumers. Table 9 displays how many URLs the users have used in each cluster.

Table 8. Numbers Of Accesses For Each Cluster

Cluster	Access number
1	18
2	677
3	281
4	149
5	423
6	622
7	100
8	252

Table 9. Allocating URL request to each cluster

URL Request	Access no	Cluster no
/administrator/components/com_jigallery/admin_jigallery.p	11	Cluster-1
/catalog/	418	Cluster-8
/components/	695	Cluster-6

/images/	621	Cluster-6
...

The selection of a certain number of clusters depends on the domain expertise as we do not know how many clusters we want.

We begin with the number of the initial clusters to see the grouping of results. The comparison of the standard deviation of the two techniques is shown in Table 10. Table 10 shows that cluster members' similarities in the MK-means are more related to each other than K-means in most clusters. We can refer to this less variation in most clusters to the computation of log-likelihood and the hierarchal cluster (pre-cluster) step.

Table 10. Comparison between modified k-mean and k-mean

Cluster no	Standard deviation	
	K-mean	Modified k-mean
Cluster -1	8.992	1.604
Cluster -2	25.456	1.258
Cluster -3	18.385	10.878
Cluster -4	5.657	8.4
Cluster -5	5.568	71.963
Cluster -6	2.121	1.291
Cluster -7	21.615	82.995
Cluster -8	10.017	34.75

Table 11 shows each cluster's standard deviation in the Kohonen Clustering algorithm and the number of accesses in each. When determining the correct clustering technique, the nature of the training knowledge collection will be essential. The following figures 7 and 8 shows the deviations between the three techniques from the standard deviation

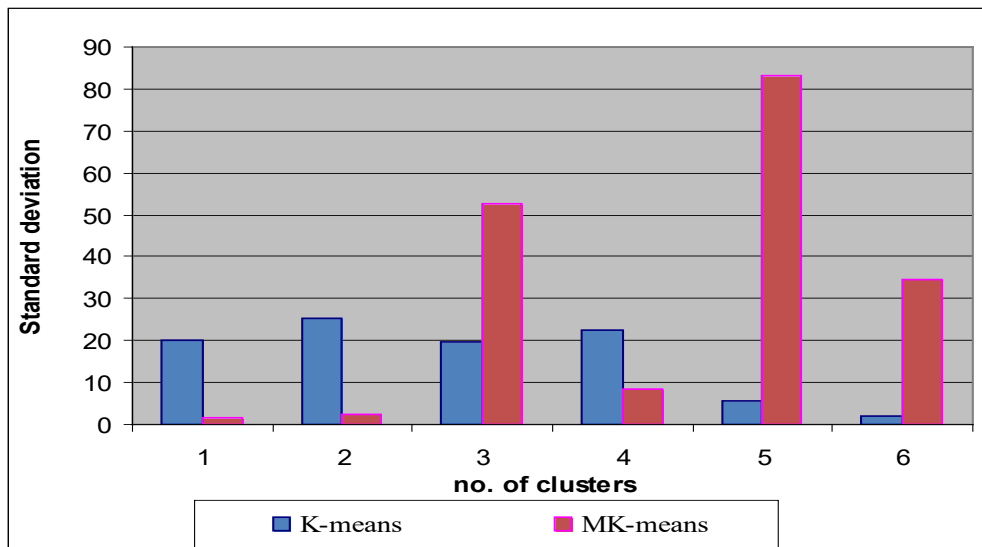


Figure 7. Difference between K-means and MK-means in 6 clusters

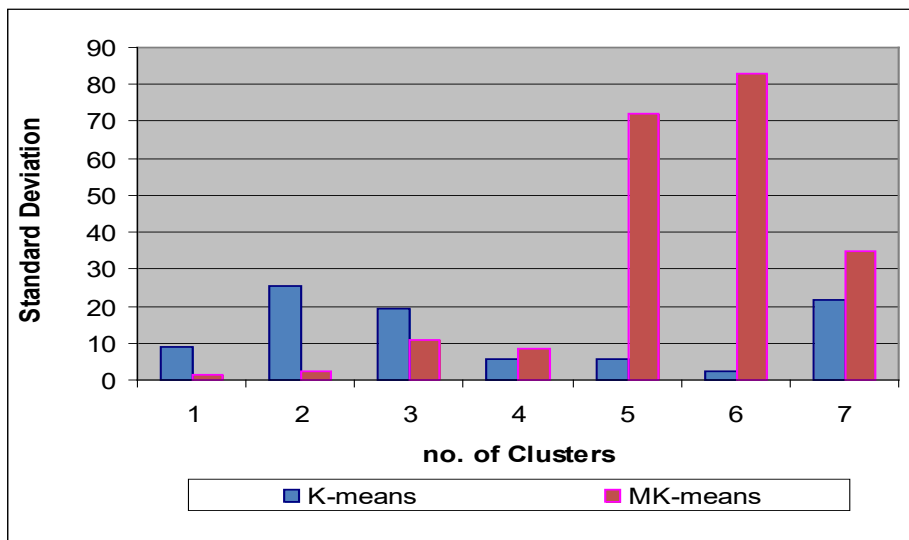


Figure 8. Difference between K-means and MK-means in 7 clusters

Table 11. Kohonen standard deviation

Cluster	Standard deviation	Access numbers
1	34.75	16
2	5.568	40
3	18.385	94
4	10.02	65
5	2.309	127
6	5.657	263
7	3.32	423
8	10	649

As Figure 9 shows, MK means have a reduced difference for most of the clusters in our data set and appear with a lesser difference in more than one cluster. Still, they are expected to have a low overall difference in the all largest datasets.

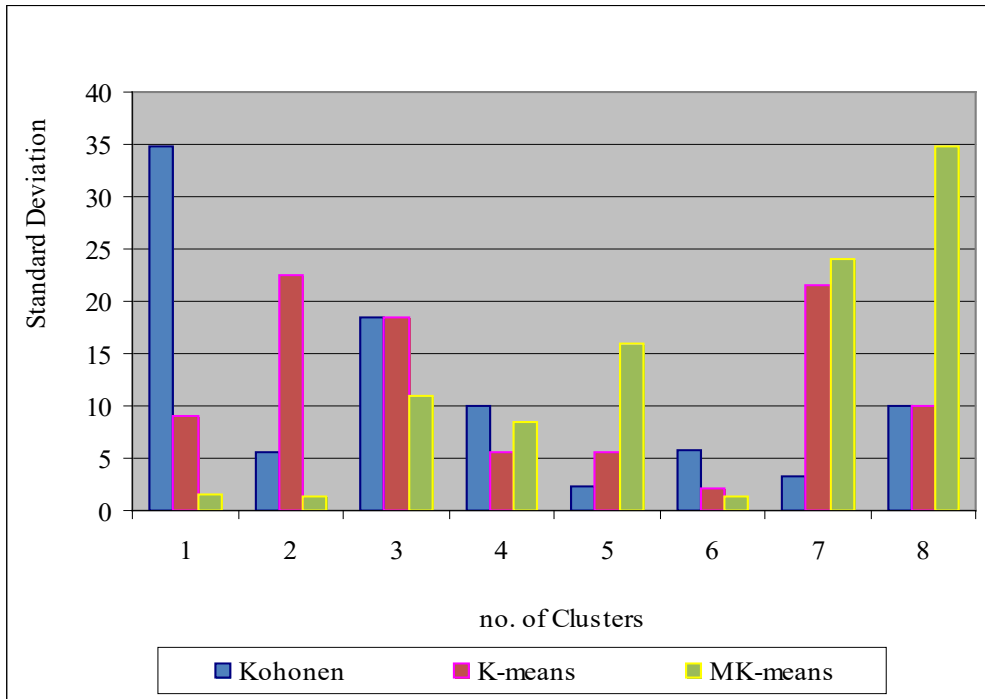


Figure 9. Difference between K-means, MK-means, and Kohonen

6. Frequent patterns algorithms

As noted above, every E-service’s web site’s customization will offer useful insights into consumer behavior and allow end-users to recognize web users’ surfing behavior through E-service web sites. For illustration, user navigation’s association may also be improved by suggesting specific products/services for website users based on their surfing behavior. Therefore, we need to use common design strategies to accomplish such a significant connection for the user’s navigation patterns.

6.1 CARMA Algorithm

The Continuous Association Rule Mining Algorithm (CARMA) is an alternative to Apriori, reducing I / O costs, time, and space requirements. This requires only two passes and offers results with much lower rates of assistance than Apriori [14].

This also permits improvement to the level of assistance during execution. CARMA covers transactions of items and documents. Items are conditions like a flag that displays whether a particular item is present or not in a particular transaction. A collection of items is a group of items that may or may not coexist in transactions.

Within two phases, CARMA continues. Initially, frequent itemsets are defined in the data, and then rules are created from frequent itemsets in [14]. The following performs:

1. Find all similar items:

- Find Frequently items:
 - items that are more than or equal to the level for minimum support in the database.
- Get Frequent itemset:
 - Generate frequent item candidates.
 - Prune results to identify common itemset.

2. Establish strict rules for the interaction of frequent items

- Rules that reach the level of minimum. Support and minimum confidence.

The following flow charts show the generation of correlated rules, as Figure 11 shows more details about generating strong rules.

Rule support is the ratio of data transactions that are important for tasks, and the trend is true. The measurement of trust or belief in each discovered pattern is confidence, as shown in figure 10.

The following example illustrates how to create a rule as seen:

Assume our file includes five transactions and the element reference indicates a URL as follows:

- A=/components/,
- B=/skin/,
- C=/includes/,
- D=/skins/ and E= /catalog/

Let the min.support=50% and min.confidence=50%.

Note the support and confidence calculated as follows:

$$\text{Support (URL1} \rightarrow \text{URL2)} = \frac{\text{No.of transactions containing both URL1 and URL2}}{\text{No.of total transactions}}$$

$$\text{Confidence (URL1} \rightarrow \text{URL2)} = \frac{\text{No.of transactions containing both URL1 and URL2}}{\text{No.of total transactions containing URL1}}$$

TransID	items
1	A, C, D
2	B, C, E
3	A, B, C, E
4	B, E

Step 1: find all frequent itemsets

1.itemsets	Support
{A}	2
{B}	3
{C}	3
{D}	3

L1

Generate C2 candidates from L1

itemset
{A,B}
{A,C}
{A,E}
{B,C}
{B,E}
{C,E}

Scan for a count of each candidate

itemset	support
{A,B}	1
{A,C}	2
{A,E}	1
{B,C}	2
{B,E}	3
{C,E}	2

Compare with min.support, L2

itemset	support
{A,C}	2
{B,C}	2
{B,E}	3
{C,E}	2

Generate C3 from L2

itemset
{B,C,E}

Scan for a count each candidate

itemset	support
{B,C,E}	2

Compare with min. support, L3

itemset	support
{B,C,E}	2

Step 2:

Generate strong rule from step 1

Rule	Support(X,Y)	Support (X)	Confidence
{A}->{C}	2	2	100
{B}->{C}	2	3	66.6
{B}->{E}	3	3	100
...
{B}->{C,E}	2	3	66.66
{C}->{B,E}	2	3	66.66
{E}->{B,C}	2	3	66.66
...
{C,E}->{B}	2	2	100
{B,E}->{C}	2	3	66.66
{B,C}->{E}	2	2	100
...

Figure 10. The demonstration shows how to generate the association rules

6.1.1 Experimental Results

As user research on web site, he/she wants to reach its target more in a quicker time. Finding a correlation between user navigation paths can improve the navigation plan of requests for www. This will also become a powerful way in case we want the user to see it possible as a large number of available products/services. Furthermore, it will lead to an increase in the income of the E-services portals. It may also make a part in an advertisement and the rest of it on the correlated page where such an issue can give more view about more advertisements to users.

Finally, implementing the mining of frequent patterns on the server-side will help us make our server intelligence to predict future customer needs. Moreover, we can also provide some programming facilities on the server to help adapt the website by improving the navigational plan, design, and content related to navigational patterns. We can capture the correlation of requested pages into a prescribed framework. With some facilities, we inherent this generated framework into a specified site to automatically reorganize the navigational plan, design, and change web pages' content. As previously mentioned, it will lead to what is called the concept of object-oriented navigation.

Figure 11 illustrates the correlated pages, which help us in finding association rules by CARMA.

There are three types of linkage between URL requests: strong, medium, and weak linkage. The following tables in tables 12 to 14 show this linkage of URL requests. In our experiment, we start to support and confidence in a low percent approximately 30% increase, and until the satisfaction of results occurs. This may depend on domain expertise. Table 15 shows the correlated request with support and confidence degree sorted by confidence. The following can lead us to display more ads in the strongly correlated pages and improve web catalogs for E-service sites. We mention that CARMA works at low support and confidence than Apriori. The following tables show the difference between the two techniques. In Table16, we use initial support and confidence 20% to see how far both techniques construct rules. It has been seen that CARMA has more generating rules than Apriori. In table 17, we use

initial support and confidence of 30% for Apriori. It is noticed that Apriori does not operate in low support and confidence. Increasing or decreasing this ratio to see how related rules are generated.

7. Conclusion

This paper has addressed how to make the e-services portals adaptive and intelligent based on its user's navigations patterns. There are three aspects of achieving the development of intelligent portals. Firstly, the aspect is capturing user navigation patterns by using algorithms such as CARMA and Apriori. Secondly, grouping similar user's requests by implementing clustering techniques such as K-means, MK-means, and SOM (A system on a module). Thirdly, the classification of users' URLs being either malicious or benign will be a future work by proposed an adaptive classification technique.

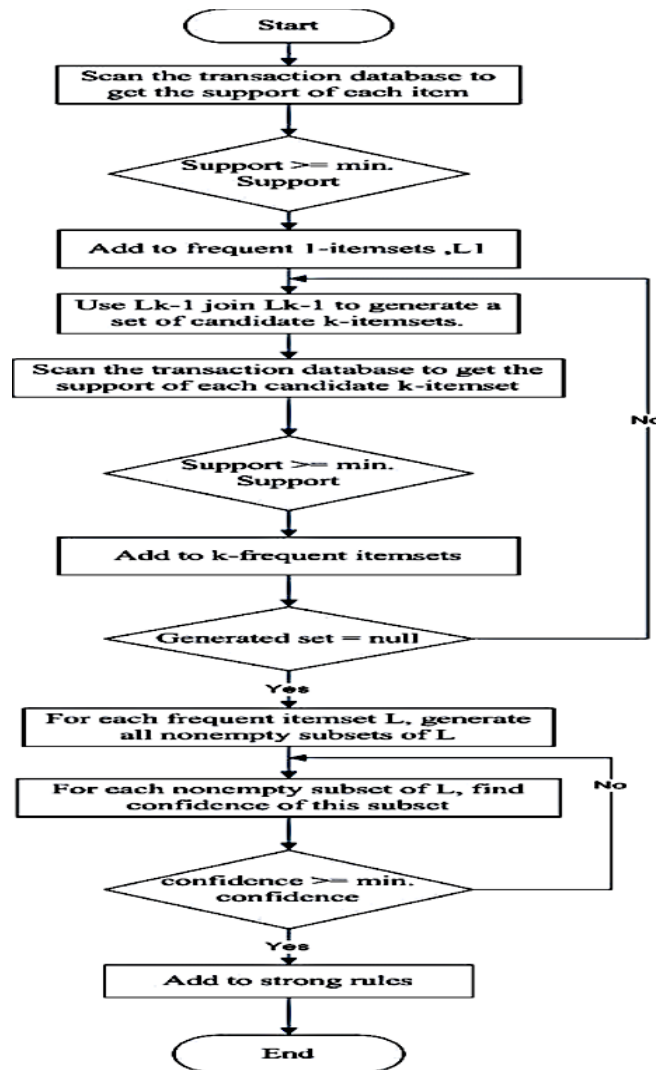


Figure 11. Generate an association rule flow chart

Table 12. Strong links between URL requests

Link	Field 1	Field 2
229	/components/ = "T'	/skin/ = "7'
190	/include= "7'	/skin/ = "7'
189	/components/ = "T'	flogslcomponenteadmin jjgallery_php =
176	/catalog/ = 'T'	/skin/ = 'T'
158	/logs/error/ = "1'	/skin/ = 'T'
150	flogecomponenteadmin jjgallery_php =	/skin/ = 'T'
145	/components/ =	/include= "7'
142	/catalog/ = "7'	/components/ = "7'
138	/components/ =	flogs/error/ = "T'
129	/catalog/ = 'T'	flogs/effort = "T'
120	/catalog/ = 'T'	/include = 'T'
120	/include = 'T'	Ilogsferrorl = "T'

Table 13. Medium links between requests

Link	Field 1	Field 2
35	/images/ = "T"	/images/items/ =
35	/scripts/ = "T"	/skin/ = "T"
34	/catalog/ = "T"	/logs/index_php =
33	/logs/modules/ = "T"	/search/ = "7"
33	/logs/componentsIadmin jjgallery_php = "T"	/logs/skin/ = "T"
30	/catalog/ = "T"	/scripts/ = "7"
29	/components/ = "T"	/logsImodules/ =
29	/logs/componentsIadmin jjgallery_php = "T"	/scripts/ = "7"
29	/logsIerror/ = "T"	/logs/skin/ = "7"
28	/include/ = "T"	/scripts/ = "7"
28	/logs/index_php = "T"	/modules/= "T"
28	/logs/index_php = "T"	/search/ = "7"
27	/logs/ = "T"	/search/ = "7"
26	/logs/components/com_virtuemart/show_image_in_imgtag_php = "T"	/skin/ = 'T'
26	flogs/componentsIadmin jjgallery_php = "T"	/logs/modules/ =
25	/include/ = "T"	/logs/modules/ =

Table 14. Weak links between requests

Link	Field 1	Field 2
14	flogs/index_php = "T'	/logsImodules/ = "T"
14	/logsIskin/ = "T"	/scripts/ = "7"
14	/logs/ = "T"	/logs/modules/ = "T"
13	1e404_php = "T"	/skin/ = "7"
13	/logs/e404_php = "T"	/skin/ =
13	/componentsfr "T"	flogs/components/com_galleria/galleria_html_php =1'
13	flogs/componentsIadmin jjgalleiy_php = "T"	flogs/componentsIcom_galleria/galleria_html_php = "T"
13	flogs/componentsIcom_galleria/galleria_html_p	/skin/ = "1'
13	/catalog/ = "T"	/logs/mods/ = "T"
13	/logs/index_php = "7'	/logs/mods/ = "T"
13	!components! = "T'	/e404_php = "T"
13	/components/ = "T"	/logs/e404_php = "T"
12	flogsf "7'	flogs/componentsIcom_virtuemart/show_image_in_imgtag_

Table 15. Correlated requests

Consequent	Antecedent	Support %	Confidence %
/components/	/logs/components/admin_figallery_php	21.834	98.667
	/skin/		
/components/	/logs/components/admin_figallery_php	28.093	97.927
/skin/	/include/	31.587	87.558
/skin/	/components/		
	/logs/components/admin_figallery_php	27.511	78.307
/skin/	/logs/components/admin_figallery_php	28.093	77.72
/skin/	/components/	43.377	76.846
/components/	/logs/components/admin_figallery_php	28.093	76.684
/skin/			
/skin/	/logs/error/	31.587	72.811
/skin/	/catalog/	35.517	72.131
/components/	/skin/	48.617	68.563
/components/	/include/	31.587	66.82
/logs/component_s/admin_figallery_php	/components/	33.333	64.629
	/skin/		
/components/	/logs/error/	31.587	63.594
/logs/component_s/admin_figallery_php	/components/	43.377	63.423
/components/	/catalog/	35.517	58.197
/include/	/skin/	48.617	56.886
/catalog/	/skin/	48.617	52.695
/logs/component_s/admin_figallery_php	/components/	43.377	49.664
/skin/			
/include/	/components/	43.377	48.658

Table 16. Difference CARMA and Apriori

Analysis parameters	CARMA	Apriori
Number of valid transactions	687	731
Number of rules	24	11
Min. support	21.837%	20.246%
Min. confidence	44.311%	47.651%
Max. support	48.617%	45.691%
Max. confidence	98.667%	65.823%

Table 17. Analysis of Apriori

Analysis parameters	Apriori
Number of valid transactions	731
Number of rules	3
Min. support	31.327%

Min. confidence	47.651%
Max. support	45.691%
Max. confidence	55.022%

References

- [1] R. Forest, G. Heyner, K. Kanngiesser, and H. Renta I., and S?ndor J.?Analysis of Web User Identification Methods? World Academy of Science, Engineering, and Technology 34,2007.
- [2] Spiliopoulou, Myra, et al. "A framework for the evaluation of session reconstruction heuristics in the web-usage analysis." *Inform Journal on Computing* 15.2 (2003): 171-190.
- [3] Tan, Shun-Hua, Miao Chen, and Guo-Hai Yang. "User behavior mining on large scale web log data." *The 2010 International Conference on Apperceiving Computing and Intelligence Analysis Proceeding*. IEEE, 2010.
- [4] Eirinaki, Magdalini, and Michalis Vazirgiannis. "Web mining for web personalization." *ACM Transactions on Internet Technology (TOIT)* 3.1 (2003): 1-27.
- [5] Celik, Serra. "They Know What You Will Do Next Click." *Interdisciplinary Approaches to Digital Transformation and Innovation*. IGI Global, 2020. 100-122.
- [6] Diep, Nguyen Ngoc, et al. "An Unsupervised Method for Web User Interest Analysis." *2019 6th NAFOSTED Conference on Information and Computer Science (NICS)*. IEEE, 2019.
- [7] Mustapha, Norwati, Manijeh Jalali, and Mehrdad Jalali. "Expectation maximization clustering algorithm for user modelling in web usage mining systems." *European Journal of Scientific Research* 32.4 (2009): 467-476.
- [8] Sneha, Y. S., G. Mahadevan, and Madhura Prakash. "A personalized product-based recommendation system using web usage mining and the semantic web." *International Journal of Computer Theory and Engineering* 4.2 (2012): 202.
- [9] Jalali, Mehrdad, et al. "A web usage mining approach based on lcs algorithm in online predicting recommendation systems." *2008 12th International Conference Information Visualisation*. IEEE, 2008.
- [10] Kaur, Babaljeet, et al. "Recommender System: Towards Classification of Human Intentions in E-Shopping Using Machine Learning." *Journal of Computational and Theoretical Nanoscience* 16.10 (2019): 4280-4285.
- [11] Bezdek, James C. *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media, 2013.
- [12] Hammouda, Khaled, and Fakhreddine Karray. "A comparative study of data clustering techniques." *University of Waterloo, Ontario, Canada* 1 (2000).
- [13] Modeler, IBM SPSS. "Algorithms Guide." IBM Corporation (2012).
- [14] Hidber, Christian. "Online association rule mining." *ACM Sigmod Record* 28.2 (1999): 145-156.