Journal of Computer Science 11 (2): 372-382, 2015
ISSN: 1549-3636

doi:10.3844/jcssp.2015.372.382 Published Online 11 (2) 2015 (http://www.thescipub.com/jcs.toc)

A NOVEL APPROACH FOR USER NAVIGATION PATTERN DISCOVERY AND ANALYSIS FOR WEB USAGE MINING

Vellingiri, J., S. Kaliraj, S. Satheeshkumar and T. Parthiban

Department of Computer Science and Engineering,
Kongunadu College of Engineering and Technology, Trichy, India

Received 2014-06-09; Revised 2014-08-28; Accepted 2014-09-10

ABSTRACT

Websites on the internet are useful source of information in our day-to-day activity. Web Usage Mining (WUM) is one of the major applications of data mining, artificial intelligence and so on to the web data to predict the user's visiting behaviours and obtains their interests by analyzing the patterns.WUM has turned out to be one of the considerable areas of research in the field of computer and information science. Weblog is one of the major sources which contain all the information regarding the users visited links, browsing patterns, time spent on a page or link and this information can be used in several applications like adaptive web sites, personalized services, customer profiling, pre-fetching, creating attractive web sites etc. WUM consists of preprocessing, pattern discovery and pattern analysis. Log data is typically noisy and unclear, so preprocessing is an essential process for effective mining process. In the preprocessing phase, the data cleaning process includes removal of records of graphics, videos, format information, records with the failed HTTP status code and robots cleaning. In the second phase, the user behaviour is organized into a set of clusters using Weighted Fuzzy-Possibilistic C-Means (WFPCM), which consists of "similar" data items based on the user behaviour and navigation patterns for the use of pattern discovery. In the third phase, classification of the user behaviour is carried out for the purpose of analyzing the user behaviour using Adaptive Neuro-Fuzzy Inference System with Subtractive Algorithm (ANFIS-SA). The performance of the proposed work is evaluated based on accuracy, execution time and convergence behaviour using anonymous microsoft web dataset.

Keywords: Web Usage Mining (WUM), Robots Cleaning, Weighted Fuzzy-Possibilistic C-Means (WFPCM), Adaptive Neuro-Fuzzy Inf erence System with Subtractive Algorithm (ANFIS-SA), Modified Levenberg-Marquardt Algorithm

1. INTRODUCTION

The enormous amount of data stored in files, databases and other repositories, it is progressively more important to develop powerful means for analysis and perhaps interpretation of such data for the extraction of interesting knowledge that could help in decision-making.

Huge development of the information accessible by means of the Internet induces its complexity in manageability. The beginning of the World Wide Web (WWW) has overwhelmed home computer users with vast amount of information (Berners-Lee *et al*., 1994). By using internet, almost any kind of topic one is in need, can find certain pieces of information that are made obtainable by other internet citizens, ranging from individual users that upload an inventory of their record gatherings, to major companies that do business through the web.

Internet activity resulting from the user interaction produces a huge quantity of data accumulated in web access log files. The owners of the websites, in particular commercial websites, are concerned in obtaining more information regarding their customers with the purpose

Corresponding Author: Vellingiri, J., Department of Computer Science and Engineering, Kongunadu College of Engineering and Technology, Trichy, India

of understanding the better target of cross-marketing campaigns, to show the relevant ads to their users and to restructure the website for a smoother navigation. WUM exploits data mining techniques to analyze the user access to websites.

In WUM, data can be collected in server logs, browser logs, proxy logs or obtained from an organization's database. These data collections differ in terms of the location of the data source, the kinds of data available, the segment of population from which the data was collected and methods of implementation.

The data sources used in WUM may include web data repositories like:

- Web server logs
- Proxy server logs and
- Browser logs

The following information can be obtained with the help of WUM:

- Number of hits
- Number of visitors
- Visitor referring website
- Visitor referral website
- Time and duration
- Path analysis
- Visitor IP address
- Browser type
- Cookies and
- Platform

Several actions can be carried out after analyzing the web log files. The following are the common actions executed from the analyzed results:

- Shortening paths of high visit pages
- Eliminating or combining low visit pages
- Redesigning pages to help user navigation and
- Redesigning pages for search engine optimization

The major phases in WUM are:

- Preprocessing
- Pattern discovery and
- Pattern analysis

Several WUM applications are growing at faster rate, especially because of the business interest in e-commerce websites and the associated web-marketing applications. Furthermore, the growing interest in the web semantic field and the recent field of web semantic mining will bring new perspectives for the WUM-related applications.

The main aim of this study is to develop an improved web usage mining technique. The other goals of this study include the following:

- Developing a novel technique for preprocessing the web log to obtain the required logs and eliminating the unnecessary logs (Huiying and Wei, 2004)
- Reducing the navigation behaviour which suits to predict the future behaviour of user (Velasquez *et al*., 2004)
- Improving the accuracy of determining the user navigation pattern by using Fuzzy based clustering technique (Suresh *et al*., 2011)
- Usage of neural network in predicting the future user behaviour (Raju and Satyanarayana, 2007)
- This research focuses on using Adaptive Neuro-Fuzzy Inference System with Subtractive Algorithm (ANFIS-SA) for prediction which includes the advantage of fuzzy and neural network (Khatibinia *et al*., 2011)

This study improves the web usage mining by enhancing the techniques for preprocessing the web logs, user navigation determining and improving the prediction result.

The methodologies used in this study are:

- A novel technique for weblog mining with better data cleaning and transaction identification
- Weighted fuzzy possibilistic c-means algorithm for pattern discovery on web usage mining
- User navigation pattern analysis using adaptive neuro-fuzzy inference system with subtractive algorithm for web page prediction

2. RELATED WORKS

The growth of the web is tremendous both in the amount of web sites available and in the amount of their users. This development generated enormous quantities of data related to the user interaction with the web sites, accumulated in web log files. In addition, the web sites owners expressed the need to better understand their visitors in order to better serve them.

Web usage mining is one of the active research areas and extensive research work has been carried out in the recent years. A number of effective techniques have been proposed by various authors to understand the user needs for better service. This section discusses few works done by various researchers for pre-processing, pattern

discovery and pattern analysis (Velmurugan *et al*., 2012) of usage patterns from web data.

2.1. Pre-Processing

The data in the log file must be preprocessed to enhance the effectiveness and ease of the mining process. The major task of data preprocessing phase is to remove noisy and unrelated data and to lessen data volume for the pattern discovery phase. Aye (2011) largely concentrated on the data preprocessing phase of WUM with actions, such as, field extraction and data cleaning algorithms. Field extraction algorithm carries out the process of extracting fields from the single line of the log file. Data cleaning approach removes unrelated or unnecessary items from the web log data (Vellingiri *et al*., 2011).

Shin and Jo (2008) developed a novel automatic web information extractor called dasiacatch crawlerpsila which uses style sheet to obtain necessary data on an objective site.

2.2. Pattern Discovery

Liang *et al*. (2006) discussed the concept of web service usage patterns and pattern discovery through service mining. The author described three different phases of service usage data: (a) User request phase, (b) template phase and (c) instance phase. At each phase, the author analyzed patterns of service usage data and the discovery of these patterns. An approach for service pattern discovery at the template phase is presented.

The discovery of the users' navigational patterns using Self Organizing Map (SOM) is proposed by (Etminani *et al*., 2009). The author used the Kohonen's SOM to pre-processed web logs for extracting the common patterns.

A Pro-patterned Extendable Neural Network (PENN) is established by (Xu *et al*., 2006) to determine and identify user's objective based on two schemes: Template matching and attention focus changing mechanisms.

Alam *et al*. (2008) illustrated a new web session clustering algorithm that exploits PSO. The authors analyzed the existing web usage clustering approaches and developed swarm intelligence dependent PSO-clustering algorithm for the clustering of web user sessions.

2.3. Pattern Analysis

Wang *et al*. (2004) developed a technique that can discover users' frequent access patterns based on users browsing web behaviours. Initially, the author introduced the concept of access pattern based on a user's access path and then puts forward a revised algorithm (FAP-Mining) according to the FP-tree algorithm to mine frequent access patterns. The new algorithm initially constructs a frequent access pattern tree and then mines users' frequent access patterns on the tree.

Kudelka *et al*. (2008) have created own Pattrio method of pattern detection on web pages. The detected patterns on web pages illustrate the web page structural design from the external viewpoint of the user. The knowledge of this structural design can be used in various connections.

3. METHODOLOGY

3.1. Preprocessing

Preprocessing is an important step because of the complex nature of the web architecture which takes 80% in mining process (Chitraa and Davamani, 2010). The raw data is pretreated to get reliable sessions for efficient mining. It includes the domain dependent tasks of data cleaning, user identification, session identification, path completion and construction of transactions. Preprocessing of a web log file simply reformats the entries of a log file into a form that can be used directly by the subsequent steps of the log analyzer. In the preprocessing phase of this research, the following steps are carried out in sequence as illustrated in **Fig. 1**.

Data cleaning is the task of removing irrelevant records that are not necessary for mining. Data cleaning includes the following sub-phases as illustrated in **Fig. 2**.

The process of data cleaning is removal of outliers or irrelevant data. Analyzing the huge amounts of records in server logs is a cumbersome activity. So initial cleaning is necessary. If a user requests a specific page from server entries like gif, JPEG, etc., are also downloaded which are not useful for further analysis are eliminated. The records with failed status code are also eliminated from logs. Automated programs like web robots, spiders and crawlers are also to be removed from log files. Thus removal process in the data cleaning phase includes.

3.1.1. Elimination of Local and Global Noise

Web noise can be normally categorized into two groups depending on their granularities.

Global Noise: It corresponds to the unnecessary objects with huge granularities, which are no smaller than individual pages. This noise includes mirror sites, duplicated web pages and previous versioned web pages.

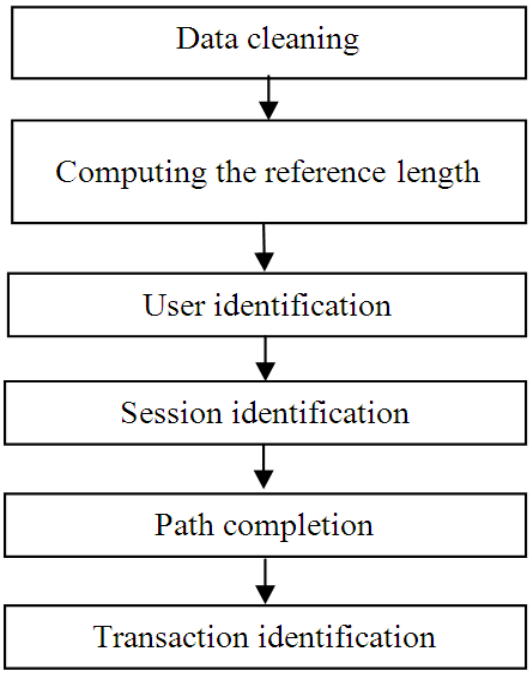

Fig. 1. Steps in preprocessing phase

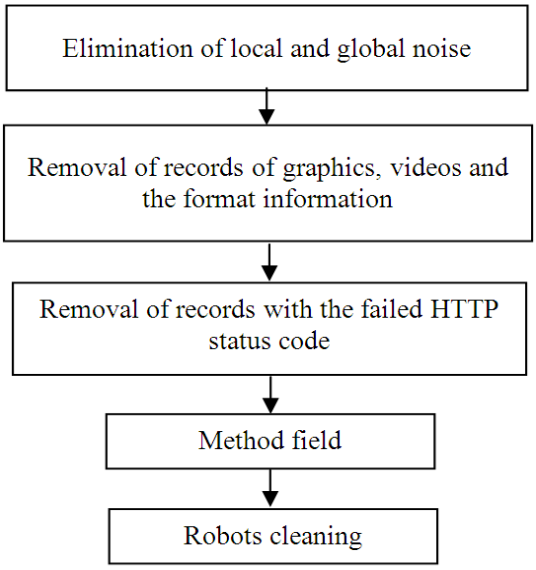

Fig. 2. Sub-phases in data cleaning

Local (Intra-Page) Noise: It corresponds to the irrelevant items inside a web page. Local noise is typically incoherent with the major content of the page. This noise includes banner ads, navigational guides, decoration pictures. These noises have to removed for better results.

3.1.2. The Records of Graphics, Videos and the Format Information

The records have filename extension of GIF, JPEG, CSS and so on, which can be found in the URI field of the every record, can be removed. This extension files are not actually the user interested web page, rather it is just the documents embedded in the web page. So it is not necessary to include in identifying the user interested web pages. This cleaning process helps in discarding unnecessary evaluation and also helps in fast identification of user interested patterns.

3.1.3. The Records with the Failed HTTP Status Code

The HTTP status code is then considered in the next process for cleaning. By examining the status field of every record in the web access log, the records with status codes more than 299 or lesser than 200 are removed. This cleaning process will further reduce the evaluation time for determining the used interested patterns.

3.1.4. Method Field

Records having value of POST or HEAD in Method field are reserved in present study for acquiring more accurate referrer information which is different from most other researches.

3.1.5. Robots Cleaning

Web Robot (WR) (also called spider or bot) is a software tool that periodically scans a web site to extract its content. Web robots automatically follow all the hyperlinks from a web page. Search engines, such as Google, periodically use WRs to gather all the pages from a web site in order to update their search indexes. The number of requests from one WR may be equal to the number of the web site's URIs. If the web site does not attract many visitors, the number of requests coming from all the WRs that have visited the site might exceed that of human-generated requests.

Eliminating WR-generated log entries not only simplifies the mining task that will follow, but it also removes uninteresting sessions from the log file. Usually, a WR has a breadth (or depth) first search strategy and follows all the links from a web page. Therefore, a WR will generate a huge number of requests on a web site. Moreover, the requests of a WR are out of the analysis scope, as the analyst is interested in discovering knowledge about users' behaviour.

Most of the web robots identify themselves by using the user agent field from the log file. Several databases referencing the known robots are maintained. However, these databases are not exhaustive as each day new WRs appear or are being renamed, making the WR identification task more difficult.

To identify web robots' requests, the data cleaning module implements two different techniques.

In the first technique, all records containing the name "robots.txt" in the requested resource name (URL) are identified and straightly removed.

The next technique is based on the fact that the crawlers retrieve pages in an automatic and exhaustive manner, so they are distinguished by a very high browsing speed. Therefore, for each different IP address, the browsing speed is calculated and all requests with this value more than a threshold are regarded as made by robots and are consequently removed. The value of the threshold is set up by analyzing the browser behaviour arising from the considered log files.

This helps in accurate detection of user interested patterns by providing only the relevant web logs. Only the patterns that are much interested by the user will be resulted in the final phase of identification if this cleaning process is performed before start identifying the user interested patterns.

3.2. Pattern Discovery

The main objective of this step is to cluster the user behaviour and the navigation patterns. For this purpose clustering algorithm is used. Clustering plays a significant role in data analysis and understanding the behaviour of users in the websites. It combines the data into classes or clusters with the intention that the data objects inside a cluster have huge similarity in relationship to one another, but are very dissimilar to those data objects in other clusters. In this study, Weighted Fuzzy Possibilistic C-Means Algorithm is used to find out the user behaviour.

FPCM generates memberships and possibilities at the same time, together with the usual point prototypes or cluster centers for each cluster. WFPCM is an integration of both Possibilistic C-Means (PCM) and Fuzzy C-Means (FCM) with weighting function that is supposed to circumvent a variety of difficulties of PCM and FCM. WFPCM completely ignores the noise sensitivity deficiency of FCM, overcomes the coincident clusters problem of PCM.

Algorithm

Step 1: Initialize FCM which is an iterative clustering technique and produce an optimal c partition by minimizing the weighted within group sum of squared error objective function.

Step 2: The fuzzy partition matrix should satisfy the condition Equation 1 and 2:

$$0 \prec \sum_{j=1}^{n} \mu_{ij} \prec n, \forall i \in \{1,\ldots,c\} \quad (1)$$

$$\sum_{i=1}^{c} \mu_{ij} = 1, \forall j \in \{1,\ldots,n\} \quad (2)$$

Step 3: The objective function of J_{FCM} is determined using the the algorithm of FCM.

Step 4: To improve the weakness of FCM and to produce memberships that have a good explanation for the degree of belongingness for the data relaxed the constrained condition 3.2 and used PCM.

Step 5: The objective function of J_{FCM} is found using the algorithm of PCM.

Step 6: Characteristics of both Fuzzy and Possibilistic C-Means are combined in this step by using the weight factor S_{ij} to obtain the J_{WFPM} Equation 3:

$$J_{WFPCM}(U,T,V) = \sum_{i=1}^{c}\sum_{j=1}^{n} S_{ij}(\mu_{ij}^{m} + t^{n}) d^{2}(X_{j}, v_{i}) \quad (3)$$

To predict the user behaviour existing approaches used FCM and PCM. But the existing approaches are inadequate because of its sensitivity towards noise. Thus with the help of WFPCM, noise is reduced, provides more accuracy and thus provides better result in predicting the user behaviour.

3.3. Pattern Analysis

The third important phase is classifying the user behaviour and ANFIS-SA is used for this purpose in this research. In the online phase, when a new request appears at the server, the wanted URL and the session to which the user belongs are determined, the primary knowledge base is restructured and a list of implication is suggested to the demanded page. After the clustering is carried out, the output will be a set of clusters $np' = \langle nP_1, nP_2,\ldots,nP_n\rangle$ where $np_i = \langle P_1, P_2,\ldots,P_k\rangle$ where k represents the set of web pages identified as user navigation patterns and $1 \le i \le n$. Sequence $W' = \langle P_1, P_2,\ldots,P_m\rangle$ represents a current active session and m indicates size of active session window. The web pages present in the active session are sorted and after this the prediction list is determined by means of classification.

After this process, for building the prediction list, the technique used is adaptive neuro-fuzzy inference system with subtractive algorithm. The ANFIS is a framework of adaptive technique to assist learning and adaptation (Kumar and Punithavalli, 2011). This kind of framework formulates the ANFIS modeling highly organized and not as much of dependent on specialist involvement. To illustrate the ANFIS architecture, two fuzzy if-then rules according to first order Sugeno model are considered Equation 4 and 5:

$$Rule1: If\ (xisA_1)\ and\ (yisB_1)\ then\ (f_1 = p_1x + q_1y + r_1) \quad (4)$$

$$Rule2: If\ (xisA_2)\ and\ (yisB_2)\ then\ (f_2 = p_2x + q_2y + r_2) \quad (5)$$

where, x and y are represents the inputs, A_i and B_i indicating the fuzzy sets, f_i indicates the outputs within the fuzzy region indicated by the fuzzy rule, pi, qi and ri shows the design parameters that are determined while performing training procedure. The ANFIS architecture to execute these two rules is represented in **Fig. 3**, in which a circle shows a fixed node and a square shows an adaptive node.

All the nodes in the initial layer are adaptive. The outcomes from these adaptive nodes are fuzzy membership grade of the inputs that are indicated by Equation 6 and 7:

$$O_i^1 = \mu_{A_i}(x) i = 1,2 \quad (6)$$

$$O_i^1 = \mu_{B_{i-2}}(y) i = 3,4 \quad (7)$$

where, $\mu_{Ai}(x)$, $\mu_{Bi-2}(y)$ can allow any fuzzy membership function. For instance, if the bell shaped membership function is utilized, $\mu_{Ai}(x)$ is given by Equation 8:

$$\mu_{A_i}(x) = \frac{1}{1 + \left\{\left(\frac{x - c_i}{a_i}\right)\right\}^{b_i}} \quad (8)$$

where, a_i, b_i and c_i is nothing but the parameters of the membership function which controls the bell shaped functions accordingly.

In the second layer, the nodes are fixed. These nodes are named with M, indicating that they perform as a simple multiplier. The outcome of this layer can be given by Equation 9:

$$O_i^2 = w_i = \mu_{A_i}(x)\mu_{B_i}(y) i = 1,2 \quad (9)$$

Which represents the firing strengths of the rules.

Also, the nodes in third layer are fixed. They are named with N, indicating that they are occupied in a normalization function to the firing strengths from the earlier layer.

The outputs of this layer can be represented as Equation 10:

$$O_i^3 = \bar{w}_i = \frac{w_i}{w_1 + w_2} i = 1,2 \quad (10)$$

Which represents the normalized firing strengths.

All nodes that are in fourth layer are adaptive. The outcome of the entire node in this layer is just the multiplication of the normalized firing strength with the first order polynomial. As a result, the outcome of this layer is represented by Equation 11:

$$O_i^4 = \bar{w}_i f_i = \bar{w}_i (p_i x + q_i y + r_i) i = 1,2 \quad (11)$$

There is only one node named 'S' in the layer 5. This node performs the addition of all the incoming signals. Thus, the overall output of the model is given by Equation 12:

$$O_i^5 = \sum_{i=1}^{2} \bar{w}_i f_i = \frac{\sum_{i=1}^{2} w_i f_i}{w_1 + w_2} \quad (12)$$

It can be distinguished that layer 1 and the layer 4 are adaptive layers. Layer 1 composes of three adjustable parameters like a_i, b_i and c_i that is related to the input membership functions.

These parameters are represented as premise parameters. In layer 4, there exists three adjustable parameters namely p_i, q_i and r_i, related to the first order polynomial. These parameters are called consequent parameters.

3.3.1. ANFIS with Subtractive Algorithm

The optimum number of ANFIS fuzzy rules is determined by Subtractive Algorithm (SA) (Khatibinia *et al*., 2011) in ANFIS-SA. In this study for the purpose of user navigation pattern analysis, an efficient method is developed to train ANFIS with high performance.

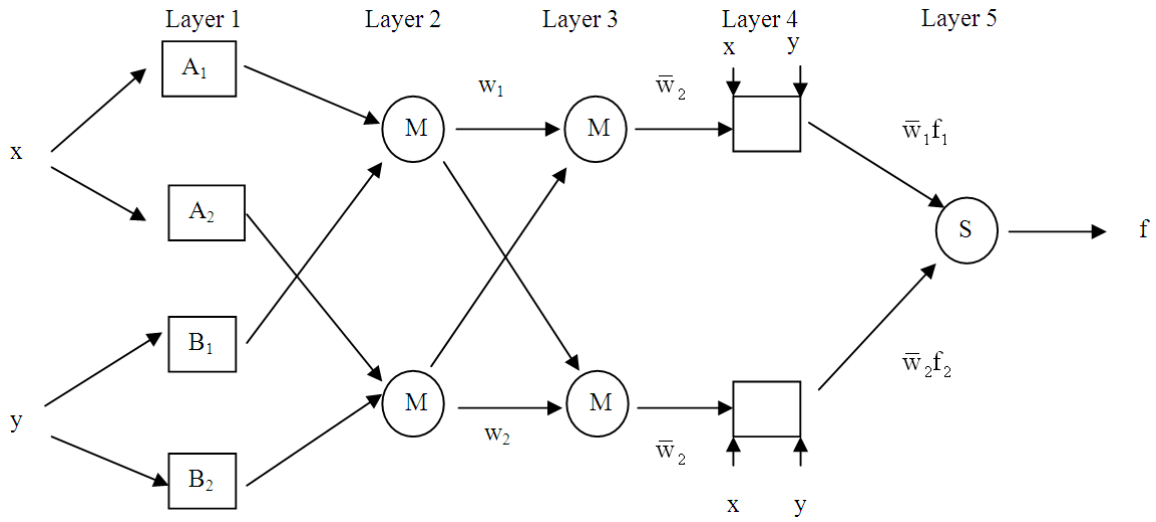

Fig. 3. ANFIS architecture

In this study, the chief purpose is to optimize the shape of membership functions of fuzzy variables with the least number of fuzzy rules. These parameters are determined by subtractive algorithm.

Subtractive algorithm is a well-organized approach for discovering the optimal number of data clusters. In SA the center candidates are the data samples themselves.

3.2. Subtractive Algorithm

The procedure of finding optimum number of ANFIS fuzzy rules and potential reduction is repeated until the stopping criterion is met. Consider, Z = {z_1, z_2,…,z_n} be a set of n records and P is the potential associated to z_i. The algorithm of SA is as follows:

Step 1: If $P_k > \epsilon^{up}$ P_1: Accept z_k as the next fuzzy rules and continue.
Step 2: Otherwise, if $P_k < \epsilon^{down}$ P_1: Reject z_k and finish the algorithm.
Step 3: Otherwise, let f_{min} be the optimum number of fuzzy rules.
Step 4: If the rule doesn't satisfy the condition, reject z_k and assign it the potential 0.0.
Step 5: Select the point with higher potential as new z_k and repeat the process.

In the above algorithm ϵ^{up} specifies a threshold above which the point is selected as an optimum number of fuzzy rules without any doubts and ϵ^{down} specifies the threshold below which the fuzzy rules is definitely rejected. The learning algorithm used by ANFIS-SA is modified levenberg-marquardt algorithm.

4. EXPERIMENTALS RESULTS

In order to evaluate the proposed preprocessing phase with robots cleaning, experiments were carried out using UCI machine learning repository (University of California, Irvine). Anonymous microsoft web dataset (http://archive.ics.uci.edu/ml/datasets/Anonymous+Microsoft+Web+Data) is selected from the UCI Machine Learning Repository datasets for the evaluation purpose. The proposed approach is evaluated against existing approaches like LCS, ANFIS and ANFIS-SA. The performance of the proposed approaches are evaluated using the following parameters like:

- Prediction accuracy
- Convergence behaviour
- Execution time

4.1. Prediction Accuracy

The prediction accuracy can be calculated using the following formula:

$$\text{Prediction accuracy} = \frac{\text{Number of records correctly predicted}}{\text{Total number of records}} \times 100$$

The prediction accuracy of the proposed approaches is compared in **Fig. 4**.

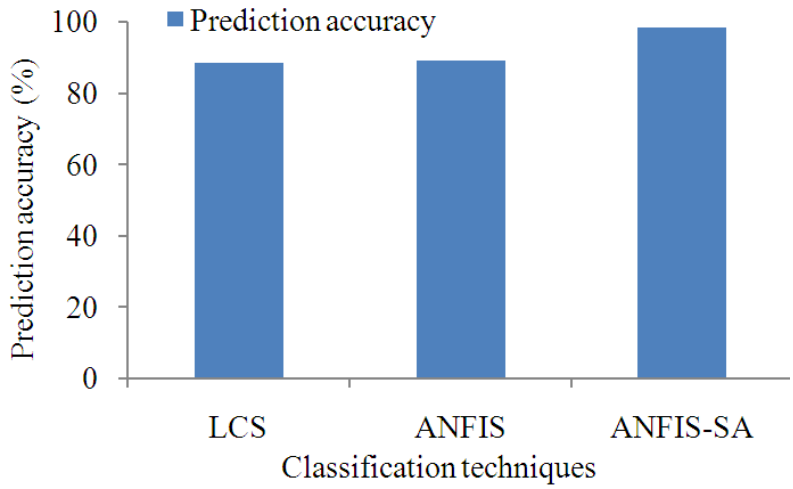

Fig. 4. Comparison of prediction accuracy

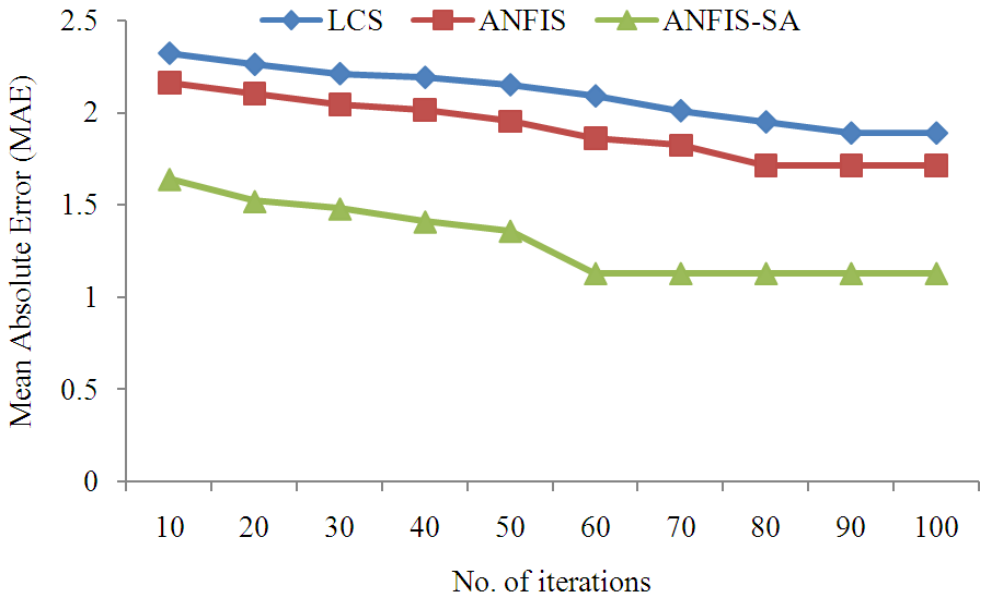

Fig. 5. Comparison of convergence behaviour

The prediction accuracy of the proposed approaches is obtained and tabulated. From the table, it is observed that the proposed ANFIS-SA approach has higher prediction accuracy since it predicts the user navigation pattern accurately when compared with the other proposed approaches. The accuracy obtained by the proposed ANFIS-SA is 98.52% whereas the accuracy obtained by LCS and ANFIS approaches are 88.36 and 89.29% respectively.

It is observed from the graph that the proposed ANFIS-SA shows better prediction accuracy than LCS and ANFIS approaches.

4.2. Convergence Behaviour

Convergence behaviour is the fixed iteration number or the result which does not change after a certain number of iterations. It is also the property of classification or clustering technique to attain the minima it was intended to identify.

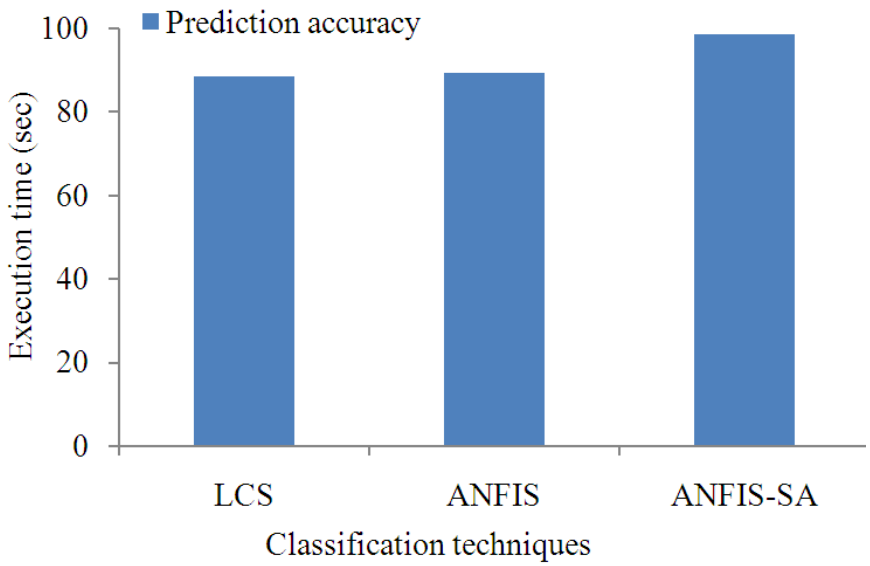

Fig. 6. Comparison of execution time

The convergence behaviour is measured against the Mean Absolute Error (MAE). MAE is a quantity used to measure how close forecasts or predictions are to the eventual outcomes. The MAE is given by:

$$MAE = \frac{1}{n}\sum_{i=1}^{n} | f_i - y_i |$$

As the name suggests, the MAE is an average of the absolute errors $e_i = f_i - y_i$, where f_i is the prediction, y_i the true value and n is the total number of records. It is obtained for different number of iterations.

Figure 5 represents the convergence behaviour of the proposed approaches for the anonymous microsoft web dataset.

It is observed from the **Fig. 6** that the existing LCS approach converges in 90 iterations and ANFIS converges in 80 iterations. But, the proposed ANFIS-SA converges in just 60 iterations. Thus, the proposed ANFIS-SA has better convergence behaviour than the other proposed approaches.

4.3. Execution Time

The time taken for the predicting the user navigation behaviour from the anonymous microsoft web dataset is considered as one of the important performance measure. The time taken for predicting the user navigation behaviour by LCS, ANFIS and ANFIS-SA is evaluated and is tabulated.

From the **Fig. 6**, it is observed that the proposed ANFIS-SA takes very less time for predicting the user navigation behaviour when compared to the other approaches like LCS and ANFIS. Time taken by the proposed ANFIS-SA is very low when compared to the time taken by LCS and ANFIS approach which takes 2.6 and 2.3 sec respectively.

It is observed from the **Fig. 6** that the ANFIS-SA approach takes only 0.92 sec.

5. CONCLUSION

Web mining is the extraction of interesting and useful knowledge and implicit information from artifacts or activity related to the WWW. Web servers record and accumulate data about user interactions whenever requests for resources are received. Analyzing the web access logs can help understand the user behavior and the web structure. An important knowledge that can be obtained from web log files is the user's navigation pattern. The navigation pattern knowledge can be used to help users from getting loss in the cyberspace by predicting their future request. In the preprocessing phase, the data cleaning process includes removal of records of graphics, videos, format information, records with the failed HTTP status code and robots cleaning. In the second phase, the user behaviour is organized into a set of clusters using Weighted Fuzzy-Possibilistic C-Means (WFPCM), which consists of "similar" data items

based on the user behaviour and navigation patterns for the use of pattern discovery. In the third phase, classification of the user behaviour is carried out for the purpose of analyzing the user behaviour using Adaptive Neuro-Fuzzy Inference System with Subtractive Algorithm (ANFIS-SA). The experimental section reveals that the proposed approach performs better in predicting the web user behavior.

6. ACKNOWLEDGEMENT

We thank our colleagues in Computer Science Department of Kongunadu College of Engineering and Technology, Trichy, India who provided insight and expertise that greatly assisted the research.

7. REFERENCES

Alam, S., G. Dobbie and P. Riddle, 2008. Particle swarm optimization based clustering of web usage data. Proceedings of the IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Dec. IEEE Xplore Press, Sydney, NSW, pp: 451-454. DOI: 10.1109/WIIAT.2008.292

Aye, T.T., 2011. Web log cleaning for mining of web usage patterns. Proceedings of the 3rd International Conference on Computer Research and Development, Mar. 11-13, IEEE Xplore Press, Shanghai, pp: 490-494. DOI: 10.1109/ICCRD.2011.5764181

Vellingiri, J. and S.C. Pandian, 2011. A novel technique for web log mining with better data cleaning and transaction identification. J. Comput. Sci., 7: 683-689. DOI: 10.3844/jcssp.2011.683.689

Berners-Lee, T., R. Cailliau, A. Luotonen, H.F. Nielsen and A. Secret, 1994. The World-Wide Web. Commun. ACM, 37: 76-82. DOI: 10.1145/179606.179671

Chitraa, V. and A.S. Davamani, 2010. An efficient path completion technique for web log mining. Proceedings of the IEEE International Conference on Computational Intelligence and Computing Research.

Etminani, K., A.R. Delui, N.R. Yanehsari and M. Rouhani, 2009. Web usage mining: Discovery of the users' navigational patterns using SOM. Proceedings of the 1st International Conference on Networked Digital Technologies, Jul. 28-31, IEEE Xplore Press, Ostrava, pp: 224-249. DOI: 10.1109/NDT.2009.5272158

Huiying, Z. and L. Wei, 2004. An intelligent algorithm of data pre-processing in web usage mining. Proceedings of the 5th World Congress on Intelligent Control and Automation, Jun. 15-19, IEEE Xplore Press, pp: 3119-3123. DOI: 10.1109/WCICA.2004.1343095

Khatibinia, M., J. Salajegheh, M.J. Fadaee and E. Salajegheh, 2011. Prediction of failure probability for soil structure interaction system using modified ANFIS by hybrid of FCM-FPSO. Asian J. Civil Eng. (Building and Housing), 13: 1-27.

Kudelka, M., V. Snasel, O. Lehecka and E. El-Qawasmeh, 2008. Web content mining using web design patterns. Proceedings of the IEEE International Conference on Information Reuse and Integration, Jul. 13-15, IEEE Xplore Press, Las Vegas, NV, USA, pp: 232-237. DOI: 10.1109/IRI.2008.4583035

Kumar, A.K. and M. Punithavalli, 2011. Efficient cancer classification using Fast Adaptive Neuro-Fuzzy Inference System (FANFIS) based on statistical techniques. Int. J. Adv. Comput. Sci. Applic.

Liang, Q.A., J.Y. Chung, S. Miller and Y. Ouyang, 2006. Service pattern discovery of web service mining in web service registry-repository. Proceedings of the IEEE International Conference on e-Business Engineering, Oct. 24-26, IEEE Xplore Press, Shanghai, pp: 286-293. DOI: 10.1109/ICEBE.2006.90

Raju, G.T. and P.S. Satyanarayana, 2007. Knowledge discovery from web usage data: Extraction of sequential patterns through ART1 neural network based clustering algorithm. Proceedings of the International Conference on Conference on Computational Intelligence and Multimedia Applications, Dec. 13-15, IEEE Xplore Press, Sivakasi, Tamil Nadu, pp: 88-92. DOI: 10.1109/ICCIMA.2007.289

Shin, K. and G.S. Jo, 2008. Catch crawler: Automatic web information extractor using style sheet. Proceedings of the IEEE International Workshop on Semantic Computing and Applications, Jul. 10-11, IEEE Xplore Press, Incheon, pp: 99-102. DOI: 10.1109/IWSCA.2008.23

Suresh, K., R. MadanaMohana A.R. Reddy and A. Subramanyam, 2011. Improved FCM algorithm for clustering on web usage mining. Proceedings of the International Conference on Computer and Management, May 19-21, IEEE Xplore Press, Wuhan, pp: 1-4. DOI: 10.1109/CAMAN.2011.5778781

Velasquez, J., A. Bassi, H. Yasuda and T. Aoki, 2004. Mining web data to create online navigation recommendations. Proceedings of the 4th IEEE International Conference on Data Mining, Nov. 1-4, IEEE Xplore Press, pp: 551-554. DOI: 10.1109/ICDM.2004.10019

Velmurugan, K. and M.A.M. Mohamed, 2012. A study of network traffic pattern and its impact on performance in implementation of web services. Am. J. Eng. Applied Sci., 5: 63-69. DOI: 10.3844/ajeassp.2012.63.69

Wang, X., Y. Ouyang, X. Hu and Y. Zhang, 2004. Discovery of user frequent access patterns on web usage mining. Proceedings of the 8th International Conference on Computer Supported Cooperative Work in Design, May 26-28, IEEE Xplore Press, pp: 765-769. DOI: 10.1109/CACWD.2004.1349127

Xu, X., C. Zhou and G. Hu, 2006. Discovering and recognizing user's intention based on pro-patterned extendable network in web active service. Proceedings of the 4th International Conference on Natural Computation, Oct. 18-20, IEEE Xplore Press, Jinan, pp: 455-459. DOI: 10.1109/ICNC.2008.748