

Cluster Optimization for Improved Web Usage Mining

Roshni Ali

Department of Computer Science & Engineering
RCERT

Chandrapur, India
roshniali27@gmail.com

Abstract— Now days, World Wide Web (WWW) has become rich and most powerful source of information. Conversely, it has become tricky and critical task to retrieve actual information due to its continuous expansion in dimensions. Web Usage Mining is a step-wise technique of extracting useful access patterns of the user from web. Web personalization makes use of web usage mining techniques, for knowledge acquisition process done by analyzing the user navigational patterns. The web page personalization involves clustering of different web pages having similar navigation patterns for an individual. Since cluster size expands due to the frequent access, optimization or shrinking the size of clusters becomes a chief consideration. This paper proposes a tactic of cluster optimization based on concept of swarm intelligence techniques. Later on based on the recognition of user access patterns, clustering is implemented using neural fuzzy approach i.e. NEF Class algorithm and cluster optimization is implemented using Ant Nest Mate Approach.

Keywords-Web Usage Mining, Web personalization, NEF Class, Cluster optimization, Swarm intelligence, Pheromone Values, Scouts, Searchers

I. INTRODUCTION

With about 30 million new web pages posted daily, the WWW is the prime, most used, knowledge source and the most perspective marketplace. In order to profitably hold on to users in this hastily rising surroundings, a web site must be build in such a way that it supports user personalization. To achieve this, an organization focuses on keeping track of user activities while browsing their web sites. Although there are many tools that helps to analyze this data using some of the web statistics methods, but the data provided is a sufficient information only for the web site and not for the designer. One of the ways to overcome this shortcoming is by applying data mining techniques on the Web. Data mining is a powerful new technology with great potential to help companies focus on the most important information in the data they have collected about the behavior of their potential clients. It identifies valued information within the data that queries and reports can't successfully divulge. Data mining, or knowledge discovery, is the computer-assisted process of digging through and analyzing enormous sets of data and then extracting the meaning of the data[1]. Data mining tools foresee behaviors and upcoming trends, allowing businesses to build proactive, knowledge-driven decisions.

Web mining - is the application of data mining techniques to discover patterns from the World Wide Web. Web mining is the integration of information gathered by traditional data mining methodologies and techniques with information gathered over the World Wide Web. Web mining is the use of data mining techniques to automatically discover and extract information from Web documents and services. Web mining can be divided into three different types – Web usage mining, Web content mining and Web structure mining [2].

The rapid e-commerce growth has made both business community and customers face a critical condition. Powerful & influential competition and the customer's option to choose from several different alternatives, business community has realized the necessity of intelligent marketing strategies and relationship management. Web servers record and accumulate data about user interactions whenever requests for resources

are obtained. Analysis of Web access logs can helps to understand the user behavior and the web structure. From the point of view of business and applications, knowledge acquired by the Web usage patterns can be directly applied to efficiently manage activities related to e-business, e-services, e-education and so on . Accuracy Web usage analysis helps to attract new customers, preserve existing customers, improvise cross marketing/sales, efficiency in promotional campaigns, tracking parting customers and find the most effective logical structure for their Web space. User profiles could be built by combining users' navigation paths with other data features, such as page viewing time, hyperlink structure, and page content [4]. As different brands morph into content providers, clients have become more and more habituated of seeing content that's customized, hence laying a basis for the need importance of personalization.

In the coming sections, in the paper we will come across detailed process of Web Usage Mining in Section II, Section III will illustrate basics of Clustering in web & Neuro-Fuzzy approach for the same. Section IV will focus on area of Swarm Intelligence & Ant Nest Mate method that has been used for cluster optimization. Section V will give detailed experimental results of the work done for this paper. Performance Analysis for same is illustrated in next Section VI.

II. WEB USAGE MINING

Web usage mining refers to the automatic discovery and analysis of patterns in click streams, user transactions and other associated data collected or generated as a result of user interactions with web resources on one or more Web sites [7]. With the continual growth and propagation of e-commerce, Web services, and Web-based information systems, the volumes of click stream, transaction data, and user profile data collected by Web-based organizations in their daily operations has reached astronomical proportions [8]. Proper Analysis of such web data helps the organizations to determine the life-time value of clients, also helps in designing cross-marketing strategies for products and services, evaluating effectiveness

of promotional campaigns, optimizing the functionality of Web-based applications, provides more tailored content to viewers, and helps in finding the most effective logical structure for Web space. This kind of analysis helps in automatic discovery of significant patterns and associations among huge collection of chiefly semi-structured data stored in Web servers and applications server access logs and related operational data sources. The goal is to confine, model, and examine the behavioral patterns and profiles of users interacting with a Web site. The discovered patterns are usually represented as collections of pages, objects, or resources that are frequently accessed or used by groups of users with common needs or interests [10].

Web Usage mining Consists of three phases, mainly pre-processing, pattern discovery, and pattern analysis [11]. Fig 1. Below shows the sequence of Web Usage Mining process.

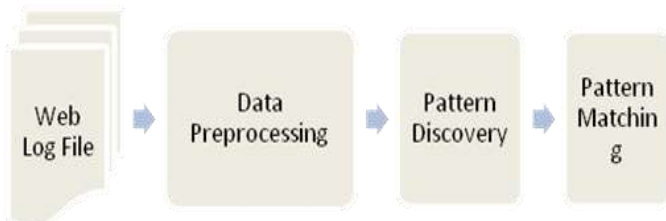


Fig 1: Web Usage Mining Process

1. Pre-Processing.

It is the process of converting the unstructured data into useful information by applying some algorithm. Web usage data sources must be integrated, filtered, cleaned, and transformed, such that gaps will be possibly filled, irrelevant information will be thrown away, and user sessions and transactions will be identified. These sources of data are mainly Web server log files, agent logs and other interfaces. The data present in the log file cannot be used as it is. Therefore, contents of the web log file should be cleaned in this preprocessing step. The unwanted data are bumped of and a minimized log file is obtained.

2. Pattern Discovery

After converting the data in log file into a formatted data the pattern discovery process is done. With the existing data of the log files many useful patterns are identified either with user id's, session details, time outs etc. It is the key component for analyzing the pre-processed data. In this phase the process is done through various algorithm and knowledge discovery techniques. It can be done using various techniques such as association rules, classification, clustering, sequential pattern and statistical analysis.

3. Pattern analysis

This process eliminates the irrelevant rules or patterns that were generated. They extract the interesting rules or patterns from the output of the pattern discovery. The most familiar form of pattern analysis comprises of a knowledge query mechanism such as SQL (Structured Query Language) or loads the usage data into a data cube toper form OLAP (Online analytical processing) operations.

III. CLUSTERING

Clustering is a method of data mining that collectively groups set of items having similarities in there characteristics revealed. In usage domain, we have observed that there are broadly two basic clusters i.e. user clusters and page clusters [12]. Clustering on user records (sessions or transactions) is referred as most common analysis done in Web usage mining and Web analytics. Clustering of users together tends to create groups exhibiting similar kind of browsing or access patterns. This knowledge is especially helpful for inferring user demographics in order to perform market segmentation in e-commerce applications or provide personalized Web content to the users with similar interests [8]. Further on, analysis of user groups based on their demographic attributes can lead to the discovery of valuable business intelligence. Moreover, Usage-based clustering has also been used to create Web-based "user communities" reflecting similar interests of users, and to study user patterns that can be used to provide dynamic recommendations in Web personalization applications.

One straightforward approach in creating an aggregate view of each cluster is to compute the centroid (or the mean vector) of each cluster. The dimension value for each page view in the mean vector is computed by finding the ratio of the sum of the page view weights across transactions to the total number of transactions in the cluster. If page view weights in the original transactions are binary, then the dimension value of a page view p in a cluster centroid represents the percentage of transactions in the cluster in which p occurs. Thus, the centroid dimension value of p provides a measure of its significance in the cluster. Page views in the centroid can be sorted according to these weights and lower weight page views can be filtered out [8]. The resulting set of page view-weight pairs can be viewed as an "aggregate usage profile" representing the interests or behavior of a significant group of users.

Since a fuzzy systems describes the knowledge it encodes better but it can't learn or adapt its knowledge from training examples. On contrary neural network learns from training examples but cannot elucidate what it has learnt hence it's not viable to infer the end result in natural language. Neural networks and fuzzy systems both have their own strengths and weaknesses. The unification of neural networks and fuzzy logic in neuro-fuzzy models provides learning & readability. Hence, researchers have made many attempts to assimilate these specific methods to craft hybrid models which will unite the advantages/merits of both methods. In the conventional approach to fuzzy clustering the model designer based on a prior knowledge fixes the membership functions and the consequent models. However, in cases where this set is unavailable and instead a set of input-output data is observed from the process, the components of fuzzy system i.e. membership and consequent models can be represented in a parametric form and the parameters are tuned with the help of neural networks. In such situations the fuzzy methods turn into neurofuzzy methods. Neuro-fuzzy methods combine the uncertainty handling capability of fuzzy systems and the learning ability of neural networks. Thus, neurofuzzy (NF) computing has become a popular framework for solving complex problems in general and clustering problems in particular. In case the knowledge about clustering or any

general problem can be expressed in linguistic rules, then a fuzzy inference system (FIS) can be built, and if it is in data, or can be learned from a simulation or training then artificial neural networks (ANNs) can be applied[16][17].

IV. OPTIMIZATION THROUGH SWARM INTELLIGENCE

Particle Swarm Optimization (PSO) was originally designed and introduced by Eberhart and Kennedy. The PSO algorithm is a population base seek algorithm based on social behavior of birds, bees or a school of fishes [20]. Originally swarm intelligence focuses on graphically simulating the graceful and unpredictable choreography of bird folk. Every single individual is represented as vector in multidimensional search area. Thus same vector have one assigned vector that can determine the subsequent progress of the particle called as velocity vector. The PSO then determines methods to revise the velocity of a particle. Each particle then updates its velocity based on present velocity and the finest arrangement explored so far [20].The PSO practice is then iterated for some fixed number of times till minimum error based on preferred performance index is attained. It has been shown that this simple model can deal with difficult optimization problems efficiently. The PSO, in the beginning, was developed for real valued spaces but many troubles are, however, defined for discrete valued spaces where the domain of the variables is finite.

Recently a family of nature have inspired lots of technical algorithms, known as *Swarm Intelligence* (SI).It has fascinated number of researchers from the areas of pattern recognition and clustering [21]. Various clustering techniques that are based on this have allegedly presented many classical methods of partitioning a complex real world dataset. This area of Swarm Intelligence is a relatively new interdisciplinary field of research that has gained huge popularity now a day. Different algorithms resembling to the domain portray inspiration from the collective intelligence emerging from the behavior of a group of social insects (like bees, termites and wasps). When acting as a community together, these insects with very limited individual capability cooperatively perform many complex tasks necessary for their continued existence. Troubles of finding and storing foods, selecting and picking up materials for future usage need a thorough planning, and are solved by insect colonies without any kind of supervisor or controller. Particle Swarm Optimization (PSO) is another very popular SI algorithm for global optimization over continuous search spaces.

The complex social behavior of ants and other social insects requires multiple levels of recognition. Thus, Ant Nest mate approach suggests that ants can distinguish nest mates from non-nest mates, which allow them to limit altruism and cooperation to members of their own colony and protect their colony from exploitation by outsiders. Ants that have the same odor will be in the same nest. The clusters obtained are feed into an ant based clustering approach that checks for the similarity of the pheromone values of the artificial ants. This is done on the fact that ants belonging to the same nest will have similar odor. In this algorithm clusters are considered as the ants nest and the url combinations in each cluster is considered as the artificial ants.

As the size of the cluster goes on increasing due to increase in users or growth of interest of users it has become inevitable need to optimize the clusters. Here we introduces a cluster optimizing methodology based on ants nest mate recognition ability and is used for eliminating the data redundancies that may occur after the clustering done by the web usage mining methods. "Ant Nest Mate approach for cluster optimization" is presented to personalize web page clusters of target users. Hierarchy relationship exists within groups. These complex behaviors can be instantiated with a fact that ants can distinguish between nest mates and non-nest mates. The level of interaction and cooperation among ants of different colony is nearly nil as to protect the exploitation of the colony from outsiders. Ants can distinguish nest mates from non-nest mates, which allow them to limit altruism and cooperation to members of their own colony and protect their colony from exploitation by outsiders.

V. EXPERIMENTAL RESULTS

1) *Creating Web log File.*

The Windows Firewall log allows advanced users to collect and identify inbound traffic. You can log dropped packets and successful connections. Once logging is turned on all of the information is written to a file called, pfirewall.log. The log file is stored in the %system root%\Windows directory. This log file contains fields like date, time, action, protocol src-ip ,dst-ip ,src-port ,dst-port ,size ,tcpflags, tcpsyn ,tcpack ,tcpwin, icmptype ,icmrcode & info path.This log file is filtered & pre-processed initially.

Step 1: Take pfirewall.log file as an input.

Step 2: Parsing the pfirewall.log.

(Selecting the only required attributes.i.e.Src ip,dst ip & size(no. of . Packets)

Step3: Store in array.

Step 4: Check which entry have same destination & source ip and apply the accumulation filter / Discretion filter.

Step 5: Display output in console.

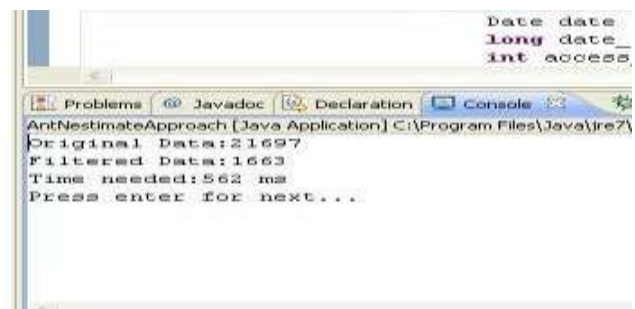


Fig 2: Filtered data from Web log File

2) *Implementing NEF CLASS.*

Out of this filtered data,Src & Dest IP are separated & its total access of time, total accessed data packets & total hits on link is calculated & displayed in console. As NEF CLASS suggests, using Src & Dest IP & Size of data, total accessed data packets & total hits on link is calculated. We had group of inputs that had 3 data values. After implementation of NEF Class we had obtained more 2 attributes i.e. Total accessed

time & Total accessed counts. Now using these inputs values, decision will be taken to add data to clustering.



Fig 3: Implementation of NEF-CLASS

Once we have calculated mean access time and mean access count, entries are then added for clustering by applying condition.

Condition Applied.

- If Access Time & Access Count \geq Mean Access Time & Access Count.....
- Then, ADD ENTRY TO CLUSTER.
- Else..Reject Or Discard.

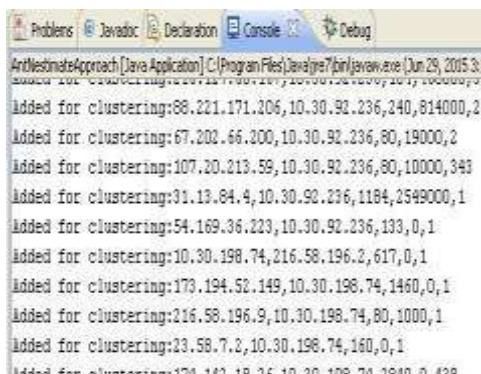


Fig 4:Centroid Cluster.

Once the cluster have been created, then entries of these clustered are parsed again to identify the similarities and difference of data accessed. This parsing is done by applying Swarm intelligence optimization technique on clusters & Ant Nest mate approach is used for same.

- Centroid = Cluster of data from web log file obtained by pattern matching & Analysis
- Pheromone Values = Navigation Links that user follows by continuous browsing.
- Scouts = Similarities of Links that has all relevant matching contents of user's need.
- Searchers = Dissimilarities in Links has all irrelevant items that don't matches up to data that was searched.

Clusters as a whole that are referred as Centroid are divided in Scouts (Cluster 1) showing similar access patterns of user & Searchers (Cluster 2) showing dissimilar ones.

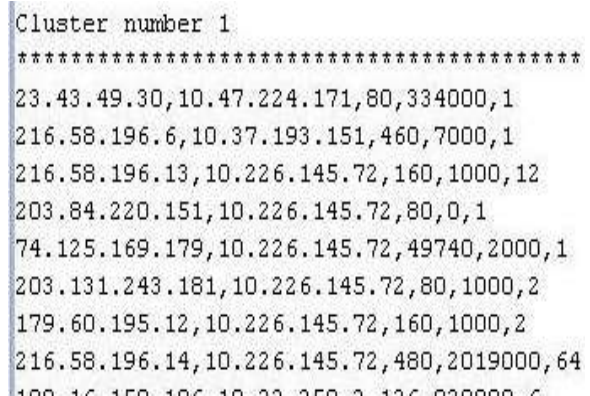


Fig 5:Cluster as Scouts.

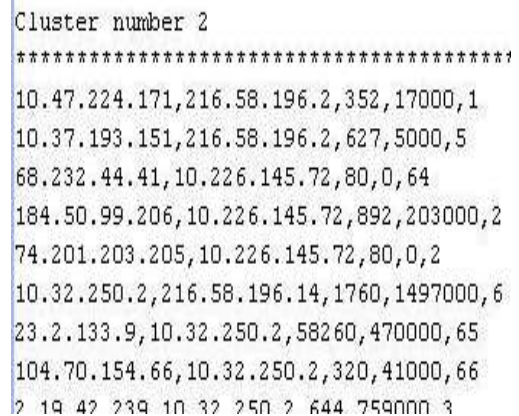


Fig 6.Cluster as Searchers

Once the Clusters are optimized and divided into Scouts & Searchers, then on basis of this optimized data, User IP's are tracked ,access to which other user has been done is identified & mean access, mean time & mean size are calculated & displayed.



Fig 7: Tracked User Profiles

VI. PERFORMANCE ANALYSIS

Once the user profiles have been tracked, Pie-charts are generated for the same analyzing the behavior of each user on the various ip address. Fig 8, below represent the user access times showing that which ip destination is accessed for how much time span. In the result below green color represent the maximum time of access & ip address related for same was 179.60.192.7. and remaining other ip's were not accessed for much longer duration as compared to mentioned one.

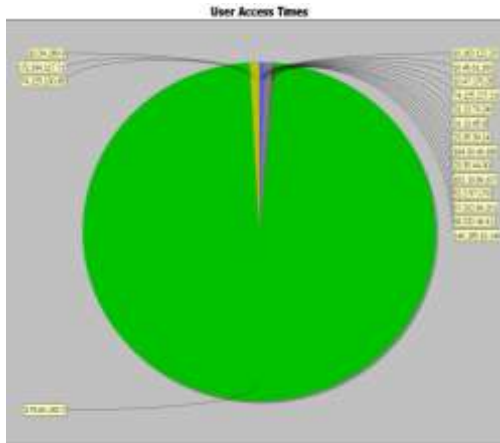


Fig 8: Pie chart for User Access Times

User access counts Pie chart represented below in Fig 9, mentions the number of times particular ip address is visited on an on by the user. As from the below Pie chart we can say ip address 31.13.95.8 have got maximum number of hit counts by the user, in the time span observed. Next to it was 10.47.176.28 represented by bit broad area in green color. Other ip addresses like 10.163.122.233, 30.45.61.160, 23.35.39.14 & others as can be seen have fair number of access counts.

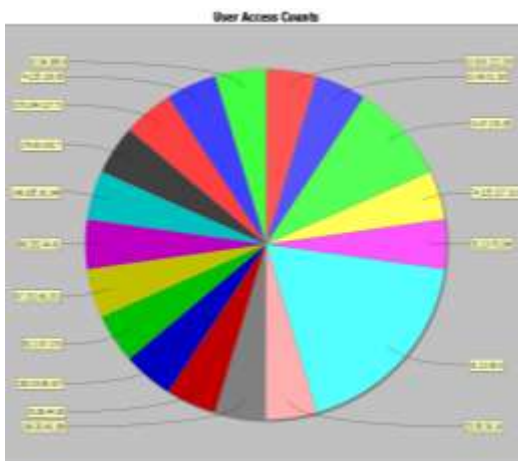


Fig 9: Pie chart for User Access Counts

User access size pie chart in Fig 10, below mentions amount of data accessed on particular ip address. Here ip address 10.34.189 can be observed to be have largest access size, next to this comes ip,s like 10.163.122..203, 10.45.61.160, 10.47.176.28 etc, and other that can be seen has average access size

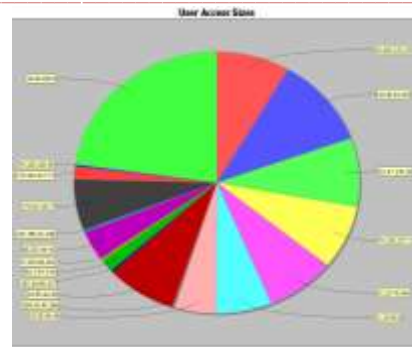


Fig 10: Pie Chart for User Access Sizes

CONCLUSION

Internet is the broad media for the users today, where ample amount of data is available to them on a single click, this data is updated day by day and growing larger in size. This paper deals with the phenomena of providing user a better access opportunity with less complexity in minimum amount of time. This paper proposes optimization technique of clusters based on swarm intelligence technique of Ant Nest Mate. As user access data on web, entries of its detailed activity gets logged into web log file. Web log file tracks the record of each and every event of customer. The web log is accessed and performs data cleaning to remove crawlers request and request to graphics. Accumulation filtering is done for log file filtering. This filtered data is then used for clustering. Clustering is done using NEF Class algorithm which is a Neuro-Fuzzy approach. Thus filtered input pattern is used for clustering & clusters thus formed are feed to Ant Nest Mate Algorithm, where cluster optimization is performed on ip's accessed. Optimization is done on the basis of swarm intelligence properties, where user following the similar navigational links will be grouped together in one cluster referred as Scouts and the other one with dis-similar behavior will be placed in Searchers. Thus the cluster or centroid gets optimized in more better way. Then we perform tracking user profiles by extracting user profiles from each cluster as a set of relevant url's and the user profiles discovered in certain period of time is compared with the user profiles discovered in later period. Based on the user profiles the web page is personalized. As a future enhancement scalability can be taken into consideration. Concept of Big data or thin data may come across as for the storage of data in clusters. Hence improved algorithm for same could be implemented for better storage.

REFERENCES

- [1] <http://www.laits.utexas.edu/~anorman/BUS.FOR/course.mat/Alex/>
- [2] Lin, C.-W. and Hong, T.-P. (2013), A survey of fuzzy web mining. WIREs Data Mining Knowl Discov, 3: 190–199. doi: 10.1002/widm.1091
- [3] Srivastava J, Desikan P and V Kumar, "Web Mining-Concepts, Applications & Research Direction" in 2002 Conference.
- [4] Abraham, Ajith, He Guo, and Hongbo Liu. Swarm intelligence: foundations, perspectives and applications. Springer Berlin Heidelberg, 2006.
- [5] Srivastava J, Desika & n P and V Kumar, "Web Mining-Accomplishment Future Directions" in 2004 Conference.
- [6] R. Kosala, and H. Blockeel, "Web Mining Research: A Survey, SIGKDD Explorations, Newsletter of the ACM Special Interest

- Group on Knowledge Discovery and Data Mining”, Vol. 2, No. 1 pp 1-15, 2000.
- [7] Puzis, Yury, et al. "Predictive web automation assistant for people with vision impairments." Proceedings of the 22nd international conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2013.
- [8] Mobasher, Bamshad. "Data mining for web personalization." The adaptive web. Springer Berlin Heidelberg, 2007. 90-135.
- [9] Qingtian Han, Xiaoyan Gao, Wenguo "Study on Web Mining Algorithm based on usage Mining", Computer Aided Industrial design and Conceptual design, 2008 CAID/CD 2008.
- [10] Facca, Federico Michele, and Pier Luca Lanzi. "Mining interesting knowledge from weblogs: a survey." Data & Knowledge Engineering 53.3 (2005): 225-241.
- [11] Jaideep shrivastav, Robert Colley, Mukund Deshpande, Pang-Ning Tan, "Web Usage Mining: discovery and Application of usage pattern from web data",ACM SIGKDD,jan2000.
- [12] www.springer.com/cda/content/.../cda.../9783642539640-c1.pdf
- [13] WANG Tong HE Pi-lian, "Web Log Mining by an Improved AprioriAll Algorithm", proceedings of world academy of science, engineering and technology volume 4 February 2005 ISSN 1307-6884, © 2005 WASET.ORG.
- [14] Mohd Helmy Abd Wahab, Mohd Norzali Haji Mohd, Hafizul Fahri Hanafi, and Mohamad Farhan Mohamad Mohsin, " Data Pre-processing on Web Server Logs for Generalized Association Rules Mining Algorithm", World Academy of Science, Engineering and Technology ,2008.
- [15] Kobra Etminani, Mohammad-R. Akbarzadeh-T, and Noorali Raaeeji Yanehsari, "Web Usage Mining: users' navigational patterns extraction from web logs using Ant-based Clustering Method," in Proc. IFSA-EUSFLAT ,2009.
- [16] Roohi, Farhat. "NEURO FUZZY APPROACH TO DATA CLUSTERING: A FRAMEWORK FOR ANALYSIS." European Scientific Journal 9.9 (2013).
- [17] Cordón, Oscar. Genetic fuzzy systems: evolutionary tuning and learning of fuzzy knowledge bases. Vol. 19. World Scientific, 2001.
- [18] R. Cooley, B. Mobasher, and J. Srivastava,"Web Mining: Information and Pattern Discovery on the World Wide Web," IEEE Computer Society,2009, pp. 558
- [19] R. Cooley, B. Mobasher, and J. Srivastava, "Data Preparation for Mining World Wide Web Browsing Patterns," KNOWLEDGE AND INFORMATION SYSTEMS, vol. 1,1999.
- [20] He, Jie, and Hui Guo. "A modified particle swarm optimization algorithm." TELKOMNIKA Indonesian Journal of Electrical Engineering 11.10 (2013): 6209-6215.
- [21] Ali, Yasir Hassan, Roslan Abd Rahman, and Raja Ishak Raja Hamzah. "Acoustic emission signal analysis and artificial intelligence techniques in machine condition monitoring and fault diagnosis: a review." Jurnal Teknologi 69.2 (2014).
- [22] D.Vasumathi, and A.Govardan,"BC-WASPT : Web Access Sequential Pattern Tree Mining," IJCSNS International Journal of Computer Science and Network Security., Vol.9,June-2009, pp. 569–571.
- [23] S.Vijayalakshmi V.Mohan, S.Suresh Raja,"Mining Constraint-based Multidimensional Frequent Sequential Pattern in Web Logs," European Journal of Scientific Research., Vol.36, pp 480-490,2009.
- [24] Ming-Syan Chen, Jong Soo Park, Philip S. Yu, "Efficient Data Mining for Path Traversal Patterns", Ieee Transactions On Knowledge And Data Engineering, Vol. 10, No. 2, March/April 1998.
- [25] F.M. Facca, P.L. Lanzi "Mining interesting knowledge from Weblogs: a survey", Data and Knowledge Engineering Vol. 53, No. 3,June 2005, pp 225-241.
- [26] C. W. Cleverdon "The Cranfield Tests on Index Languages Devices". In Readings in Information Retrieval, Morgan Kaufmann Publisher. Inc, San Francisco, California, pp.47-60.1997.
- [27] Qingtian Han, XiaoyanGao, Wenguo "Study on Web Mining Algorithm based on usage Mining", Computer Aided Industrial design and Conceptual design, 2008 CAID/CD 2008.
- [28] Jaideepshrivastav, Robert Colley, MukundDeshpande, Pang-Ning Tan, "Web Usage Mining: discovery and Application of usage pattern from web data",ACM SIGKDD,jan2000.
- [29] Bhaskaran, V.S.a.V.M., "Data Preparation Techniques for Web Usage Mining in World Wide Web-An Approach,"International Journal of Recent Trends in Engineering, Vol 2, No.4, 2009.
- [30] C, G., M. M, and D. K, "Preprocessing of Web Log Files in Web Usage Mining,"The Icfai Journal of Information Technology, 35J-2008-03-06-01(35J-2008-03-06-01): pp. 55-66, 2008.
- [31] Mohd Helmy Abd Wahab, M.N.H.M., Hafizul Fahri Hanafi,Mohamad Farhan and Mohamad Mohsin."Data Pre-processing on Web Server Logs for Generalized Association Rules Mining Algorithm,"World Academy of Science, Engineering and Technology 48, 2008.
- [32] L.K. Joshila Grace, V.M.a.D.N., "ANALYSIS OF WEBLOGS AND WEB USER IN WEB MINING," International Journal of Network Security & Its Applications(IJNSA), Vol.3, No.1, 2011.
- [33] K, D.a.V.v.M., "Session Reconstruction in Web Usage Analysis," The Icfai Journal of Information Technology, 2008.
- [34] R.Krishnamoorthi, K.R.S.a., "Identifying User Behavior by Analyzing Web Server Access Log File," IJCSNS International Journal of Computer Science and Network Security, Vol.9, No. 4, 2009.
- [35] V.Chitraa, A.S.D., "A Survey on Preprocessing Methods for Web Usage Data," (IJCSIS) International Journal of Computer Science and Information Security, Vol. 7, No. 3, 2010.
- [36] Arshi Shamsi, R.N., Pankj Pratap Singh and Mahesh Kumar Tiwar,"Web Usage Mining by Data Preprocessing," International Journal of Computer Science And Technology, Vol.3, No. 1, 2012.
- [37] R, C.R.S.J.a.D., "Web Usage Mining: Discovery and Application of Interesting Patterns from Web Data," Ph.D Thesis, University of Minnesota, 2000.
- [38] M.-S. Chen, J. S. Park, and P. S. Yu. Data mining for path traversal patterns in a web environment. In Proc. of the 16th International Conference on Distributed Computing Systems,pages 385–392, May 1996.
- [39] T. Nakayama, H. Kato, and Y. Yamane. Discovering the gap between web site designers' expectations and users' behavior. In Proc. of the Ninth Int'l World Wide Web Conference,Amsterdam, May 2000.
- [40] J. Pei, J. Han, B. Mortazavi-asl, and H. Zhu. Mining access patterns efficiently from web logs. In Proc. of the 4th Pacific-Asia Conf. on Knowledge Discovery and Data Mining,pages 396–407, April 2000.
- [41] M. Perkowitz and O. Etzioni. Adaptive web sites:Automatically synthesizing web pages. In Proc. of the Fifteenth National Conf. on Artificial Intelligence (AAAI), pages 727–732, 1998.
- [42] M. Perkowitz and O. Etzioni. Towards adaptive sites: Conceptual framework and case study. In Proc. of the Eighth Int'l World Wide Web Conf, Toronto, Canada, May 1999.
- [43] C. Shahabi, A. M. Zarkesh, J. Abidi, and V. Shah. Knowledge discovery from users web-page navigation. In Proc. of the 7th IEEE Intl. Workshop on Research Issues in Data Engineering (RIDE), pages 20–29, 1997.
- [44] M. Spiliopoulou, L. C. Faulstich, and K. Wilkler. A data miner analyzing the navigational behaviour of web users. In *Proc. Of the Workshop on Machine Learning in User Modelling of the ACAI99*, Greece, July 1999.