Weighted Fuzzy Possibilistic C-Means Algorithm for Clustering on Web Usage Mining to Discover the User Behavior

J. Vellingiri[1], Dr. S. Chenthur Pandian[2]

[1]Assistant Professor, Department of CSE, Kongunadu College of Engineering and Technology, Thottiam, Trichy (Dt) - 621 215, Tamil Nadu, India.
E-mail: vellingiri.research@gmail.com
[2]Principal, Dr. Mahalingam College of Engineering and Technology, Pollachi-642 003, Tamil Nadu, India

Abstract: Websites on the internet are the significant sources of information in our day-to-day activity. Hence, this made an enormous growth of World Wide Web in to amount of traffic, its size and complexity of websites. 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 behaviour and obtain their interest by analyzing the patterns. Weblog is one of the major sources which contain all the information regarding the users visited links, browsing patterns and time spent on a page or link. This information can be used in several applications like adaptive web sites, personalized services, customer profiling, pre-fetching and creating attractive web sites. 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. Pattern Discovery step is used in finding the user access patterns from web access log. Clustering is one of the typical algorithms in the field of data mining. The main objective of this paper is to organize a website into a set of clusters, which consists of "similar" data items based on the user behavior and navigation patterns. There are numerous clustering algorithms like K-Means, Possibilistic C-Means, etc., developed by many researchers. In recent times, Fuzzy C-Means is found to be superior as its embedded fuzzy logic. In noisy atmosphere, the memberships of FCM constantly do not correspond well to the degree of belonging of the data, and might be inexact. This paper uses a novel clustering algorithm called Weighted Fuzzy-Possibilistic C-Means (WFPCM) algorithm, which integrates extended partition entropy and inter class resemblance which is computed from the fuzzy set point of view and it needs only the membership matrix and possibilistic matrix, and is free from heavy distance computing. The performance of the proposed research work is evaluated using the parametric standards like accuracy, execution time and convergence behaviour for three data sets namely Anonymous Microsoft Web Dataset, MSNBC.com Anonymous Web Dataset and Real dataset.
[J. Vellingiri, Dr. S. Chenthur Pandian. **Weighted Fuzzy Possibilistic C-Means Algorithm for Clustering on Web Usage Mining to Discover the User Behavior.** *Academ Arena* 2018;10(10):70-76]. ISSN 1553-992X (print); ISSN 2158-771X (online). http://www.sciencepub.net/academia. 7. doi:10.7537/marsaaj101018.07.

Keywords: Web access log, User access patterns, Knowledge Extraction, Clustering, Possibilistic C-Means, Fuzzy C-Means (FCM), and Weighted Fuzzy Possibilistic C-Means (WFPCM).

1. Introduction

The rapid development of Internet, and the gradual increase the amount of information, it is estimated that there has 232 million web sites in 2009, and is increasing the speed of one millionper day. Google has recently confirmed that it has indexed over 4000 million web pages. The World Wide Web is the biggest database at the moment, and it is really a demanding task how to access efficiently these data.

An efficient technique in reducing these difficulties is web mining [20]. Web mining is that data mining technique is applied to web data to the discovery of the interesting usage patterns and implicit information [14].

Web usage mining includes three mostimportant steps:

- Data Preprocessing
- Knowledge Extraction
- Analysis of Extracted Results

After the preprocessing step, knowledge extraction is done by using clustering algorithm to extract the interesting patterns based on the user behavior. Clustering is one of the most common techniques used for data analysis and classification [17]. It is described as the method of grouping N item sets into individual clusters based on comparison or distance function [16]. A high-quality clustering technique possibly will yield clusters thus have more inter cluster and less intra cluster distance. The main purpose of clustering is to increase the resemblance of the data points within each cluster and also increase the dissimilarity across clusters.

Thus with the help of clustering technique, it is easy to extract and analyze the log data from the web sites by using the technology of web data mining and behavior analysis. Through this way, it is very simple to find out the information interrelated to the user's behavior, which is important for personalizing the web

sites. User's behavior on web sites specifies the character of the user, possibly will help in the personalization of web sites. The process of web site user's behavior is essentially composed of the data pre-processing, the discovery of website user's behavior model and the analysis of user's behavior. The user's behavior analysis is a series of activities from identifying users' objective to evaluating the final result.

Previous approaches have been used Fuzzy C-Means (FCM) algorithm to extract the user behaviour which is a typical clustering algorithm which has been extensively used in medicine imaging, bioinformatics, pattern recognition and data mining. Since the standard FCM clustering technique utilizes the squared-norm to evaluate the relationship among prototypes and data points, it can be useful in clustering only the 'spherical' clusters and many algorithms are obtained from the FCM to cluster more common dataset [8]. Since FCM technique is very susceptible to noise and to avoid the drawback of FCM the constraint of memberships in FCM are removed and the Possibilistic C-Means (PCM) algorithm is developed. To categorize a data point, an approach is deducted that the data point must strongly have their cluster centroid and it is the responsibility of membership. Also for the evaluation of the centroid, the typicality is used for improving the unnecessary effect of outliers. Previous approaches have been used FPCM [7] to find the user behaviour or navigation patternwhich combines the characteristics of both fuzzy and Possibilistic C-Means [9] to identify the user behavior. In this work proposed a clustering algorithm called Weighted Fuzzy Possibilistic C-Means (WFPCM) includes weighting function with FPCM to find appropriate clusters.

2. Related Works

Krol et al., [1] proposed an approach for the study of information Web system user's activity using clustering technique. On the basis of a Web server log, unidentified sessions are found out in the form of a 65 dimensional vector, where dimensions denote individual Web system pages. Each dimension consists of the value of a measure of interest of the user in a page during a given session. The ratio of time user spent visiting a given page to the total time of a session provides the value. Then the entire set of sessions is clustered using HCM (Hard C-Means) algorithm. The resulting clusters are considered as the user activity patterns and among them clusters dominated by a page are chosen as those where the value of user interest is higher than the given threshold value e.g.50 %. The sessions of named users, registered in the system, are obtained through an application log of user activity. The frequencies of named user sessions, consisting of individual clusters, are computed for a given period of time e.g. one month. The user activity can be evaluated by analyzing frequencies obtained. For instance, the user behavior can be considered as deviated from standard pattern if the frequency of a session in a cluster dominated by a page is below a determined threshold value e.g. 10 %. The approach was evaluated using data from a cadastral Web system exploited in an extranet.

Based on the combination of five user minimum browsing behaviors, Ling Zheng et al., [2] focused on a complete analysis of the quantitative relations between the user browsing time and the user interest rate, and proposed a novel technique of computing degree of thematic interest; this approach is based on association rule [15]. User's interest is attained by fuzzy clustering approach, and creates a user interest model. From the experimental results, it is observed that the approach has effective user interest rate, the user interest model responses user interest exactly, and is significantly used in personalized service.

A fuzzy rough approximation technique is proposed by Cuifang Chen [3] to cluster user access patterns from Web logs. In the process of clustering user access patterns, the Web page visited discloses a Web user's interest during a surfing. Time duration on a Web page shows the level of a Web user's interest. And it is featured as a fuzzy linguistic variable [13]. Each user access pattern is represented by a fuzzy vector representing visited Web pages and time durations during a surfing. Ultimately, rough approximation technique is used to cluster user access patterns represented by fuzzy vectors. User access patterns with similar surfing behaviors are clustered into one class based on this technique.

Analyzing and predicting navigational behavior of Web users play a vital role in finding user friendly and effective websites. Web personalization is a general technique for adapting the content of a website to the requirements of each particular user. An approach for dynamic recommendation based on fuzzy clustering techniques, applicable to currently on-line users is proposed by Nadi et al., [4]. The approach focuses on the aspects of both web content mining and web usage mining. By using fuzzy web mining approaches, the model concludes the user's preferences from IIS web server's access logs. The fuzzy clustering technique used in this work provides the possibility of obtaining the uncertainty among Web user's behaviors. This proposed model is implemented and tested as a recommender system for personalizing website of "Information and Communication Technology Center" of Isfahan municipality in Iran. From the results, it is observed that integrating fuzzy approach provide more interesting and useful patterns which accordingly

makes the recommender system more functional and robust.

Clustering method is very sensitive to the initial center values, requirements on the data set too high, and cannot handle noisy data. The proposal method by Suresh et al., [5] uses information entropy to initialize the cluster centers and establish weighting parameters to adjust the location of cluster centers and noise problems. The navigation datasets are sequential in nature. Clustering web data is obtaining the groups which share common interests and behavior by examining the data collected in the web servers, this improves clustering on web data effectively using improved Fuzzy C-Means (FCM) clustering [19]. Web usage mining is the application of data mining approaches to web log data repositories. It is used in obtaining the user access patterns from web access log. Web data Clusters are formed using MSNBC web navigation dataset.

3. Methodology

Data cleaning is one of the preprocessing techniques which are used to eliminate inappropriate records that are not essential for mining. The unwanted and unrelated data are removed using the preprocessing steps. Knowledge extraction is the next step after the preprocessing phase. 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 work, Weighted Fuzzy Possibilistic C-Means Algorithm is used to find out the user behaviour.

3.1.1 Fuzzy C-Means

Data cleaning is one of the preprocessing techniques which are used to eliminate inappropriate records that are not essential for mining.

Knowledge extraction is the next step after the preprocessing phase. The main objective of this step is to find out the user behavior and the navigation patterns. For this purpose clustering algorithm is used. Clustering plays a significant role in data analysis and understanding the behavior 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 paper, Fuzzy Possibilistic C-Means Algorithm is used to find out the user behavior.

The Fuzzy C-Means (FCM) can be regarded as the fuzzified form of the k-means algorithm. It is an approach of clustering which permits one piece of data to fit in to two or more clusters. This method is commonly used in pattern recognition. The algorithm is an iterative clustering approach that generates an optimal c partition by reducing the weighted within group sum of squared error objective function J_{FCM}:

$$J_{FCM}(V,U,X) = \sum_{i=1}^{c}\sum_{j=1}^{n} u_{ij}^{m}\, d^2(X_j, v_i), 1 < m < +\infty \quad (1)$$

Where $X = \{x_1, x_2, \ldots, x_n\} \subseteq Rp$ is the data set in the p -dimensional vector space, p denotes the number of data items, c represents the number of clusters with $2 \le c \le n-1$. $V = \{v_1, v_2, \ldots, v_c\}$is the c centers or prototypes of the clusters, v_i is the p-dimension center of the cluster i, and $d^2(X_j, v_i)$ is a distance [16] between object $x_j$and cluster centre v_i. $U = \{\mu_{ij}\}$symbolizes a fuzzy partition matrix with $u_{ij} = u_i(x_j)$ is the degree of membership of x_j in the ith cluster;x_j is the jth of p-dimensional measured data. The fuzzy partition matrix satisfies:

$$0 \prec \sum_{j=1}^{n} \mu_{ij} \prec n, \forall i \in \{1, \ldots, c\} \quad (2)$$

$$\sum_{i=1}^{c} \mu_{ij} = 1, \forall j \in \{1, \ldots, n\} \quad (3)$$

The parameter m is a weighting exponent on every fuzzy membership and finds out the amount of fuzziness of the resultant classification [12]; it is a predetermined number larger than one. The objective function J_{FCM} can be reduced under the parameter of U. Particularly, taking of J_{FCM}in accordance with the u_{ij} and v_i and zeroing them correspondingly, it is essential but not adequate conditions for $J_{FCM}$to be at its local extrema will be as the following:

$$\mu_{ij} = \left[\sum_{k=1}^{c}\left(\frac{d(X_j v_i)}{d(X_j v_k)}\right)^{2/(m-1)}\right]^{-1}, 1 \le i \le c, 1 \le j \le n. \quad (4)$$

$$v_i = \frac{\sum_{k=1}^{n} \mu_{ik}^{m} x_k}{\sum_{k=1}^{n} \mu_{ik}^{m}}, 1 \le i \le c. \quad (5)$$

FCMAlgorithm

Step 1: Choose the number of clusters.

Step 2: Assign randomly to each point coefficients for being in the clusters.

Step 3: Initialize the fuzzy partition matrix with $u_{ij} = u_i(x_j)$ is the degree of membership of x_j in the ith cluster;x_j is the jth of p-dimensional measured data.

Step 4: Determine the amount of fuzziness of the resultant classification.

Step 5: The objective function J_{FCM} can be reduced under the parameter of U.

3.1.2 Possibilistic C-Means

Even though FCM is a very helpful clustering approach, its memberships will not constantly communicate well to the degree of belonging of the data, and possibly will be inexact in a noisy surroundings, because the real data inevitably engages with some noises. To recover from this disadvantage of FCM, and to generate memberships that have a better explanation for the degree of belonging for the data, Krishnapuram and Keller [10] relaxed the constrained condition (3) of the fuzzy c-partition to achieve a possibilistic type of membership function and they proposed PCM for unsupervised clustering. The component produced by the PCM communicates to a dense region in the data set; all clusters are independent of the other clusters in the PCM approach. The objective function of the PCM can be represented as follows:

$$J_{PCM}(V,U,X) = \sum_{i=1}^{c}\sum_{j=1}^{n} \mu_{ik}^{m} d^2(X_j, v_i) + \sum_{i=1}^{c} \eta_i \sum_{j=1}^{n} (1-u_{ij})^m \qquad ((6)$$

Where

$$\eta_i = \frac{\sum_{j=1}^{n} \mu_{ik}^{m} ||x_j - v_i||^2}{\sum_{j=1}^{n} \mu_{ij}^{m}} \qquad (7)$$

$\eta_i$is the scale parameter at the ith cluster,

$$u_{ij} = \frac{1}{1+\left[\frac{d^2(x_j,v_i)}{\eta_i}\right]^{\frac{1}{m-1}}} \qquad ((8)$$

is the possibilistic typicality value of training sample x_j belonging to the cluster i. $m \in [1,\infty)$is a weighting factor called the possibilistic constraint. Similar to other cluster techniques, PCM technique is also based on the initialization. In PCM approaches, the clusters do not have a group of mobility, because every data point is categorized as only one cluster at a time rather than all the clusters at the same time. Consequently, an appropriate initialization is necessary for the algorithms to converge to nearly global minimum.

PCMAlgorithm

Step 1: Choose the number of clusters.

Step 2: Initialize the scale parameter η_i.

Step 3: Then initialize the possibilistic typicality value (u_{ij}) of training sample x_j belonging to the cluster$i$.

Step 4: $m \in [1,\infty)$ is a weighting factor and it is the possibilistic constraint.

Step 5: The objective function of the PCM generated.

3.1.3 Weighted Fuzzy Possibilistic C-Means

Pal defined a clustering technique that integrates the features of both fuzzy and possibilisticC-Means: Memberships and typicalities are very significant for the accurate characteristic of data substructure in clustering difficulty [11]. In PCM approaches, the clusters do not have a group of mobility, because every data point is categorized as only one cluster at a time rather than all the clusters at the same time. Consequently, an appropriate initialization is necessary for the algorithms to converge to nearly global minimum.

In order to preserve the neighborhood structure information in the weight, the weighting function is integrated into the objective function of FPCM which is defined as follows

$$S_{ij} = e^{-(||x_j - v_i||^2)/t_i} \qquad (9)$$

wheret_i is a scaling parameter. When $t_i \to 0$, the weight matrix becomes the most significant component of the clustering result while the weights are very similar with each other. In this scenario, the weighted clustering will produce poor clustering result. Alternatively, when $t_i \to \infty$, the weight matrix has entries all equal to 1 and therefore the weighted clustering is degraded into non-weighted clustering.

In order to choose appropriate values for the weights, a local scale for t_i is used as follows:

$$t_{i(Scaling\ Parameter)} = \begin{cases} \sigma_i^2, x_j \epsilon N_{ik} \\ \left(\frac{1}{c}\sum_{i=1}^{c}\sigma_i\right)^2 otherwise \end{cases} \qquad (10)$$

where $\sigma_i = (1/k)\sum_{j=1}^{k} ||x_j - v_i||^2$, k is the number of the neighbors of the ith center. $N_{ik}$is the k Nearest Neighbor (k-NN) neighborhoods of the i^{th} cluster.

Pal defined a clustering technique that integrates the features of both fuzzy and PossibilisticC-Means:

Memberships and typicalities are very significant for the accurate characteristic of data substructure in clustering difficulty. Thus, an objective function in the WFPCM depending on both memberships and typicalities is represented as:

$$J_{WFPCM}(U,T,V) = \sum_{i=1}^{c}\sum_{j=1}^{n} S_{ij}(\mu_{ij}^{m} + t^{n})\, d^{2}(X_j, v_i) \quad (11)$$

with the following constraints:

$$\sum_{i=1}^{c} \mu_{ij} = 1, \forall j \in \{1, \dots, n\} \quad (12)$$

$$\sum_{j=1}^{n} t_{ij} = 1, \forall i \in \{1, \dots, c\} \quad (13)$$

A clarification of the objective function can be acquired by means of an iterative procedure where the degrees of membership, typicality and the cluster centers are revise through:

$$\mu_{ij} = \left[\sum_{k=1}^{c} S_{ij}\left(\frac{d(X_j, v_i)}{d(X_j, v_k)}\right)^{2/(m-1)}\right]^{-1}, 1 \le i \le c, 1 \le j \le n \quad (14)$$

$$t_{ij} = \left[\sum_{k=1}^{n} S_{ij}\left(\frac{d(X_j, v_i)}{d(X_j, v_k)}\right)^{2/(\eta-1)}\right]^{-1}, 1 \le i \le c, 1 \le j \le n \quad (15)$$

$$v_i = \frac{\sum_{k=1}^{n} S_{ij}(\mu_{ik}^{m} + t_{ik}^{\eta})X_K}{\sum_{k=1}^{n} S_{ij}(\mu_{ik}^{m} + t_{ik}^{\eta})}, 1 \le i \le c \quad (16)$$

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.

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.

WFPCM 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 (2) and (3).

Step 3: The objective function of J_{FCM} is determined using 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) and used PCM.

Step 5: The objective function of J_{PCM} 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_{WFPCM} as in 11.

4. Experimental results

The performance of the proposed WFPCM Clustering technique to predict the user behaviour is evaluated on the basis of the following parameters,

- Clustering Accuracy and
- Execution Time

This dataset collects data from www.microsoft.com by randomly-selecting 37711 anonymous users. The performance of the proposed WFPCM technique is evaluated against the PCM and FCM based on the clustering accuracy and execution time.

4.1.1Clustering Accuracy

The clustering accuracy can be calculated using the following formula,

$$\text{Clustering Accuracy} = \frac{Numberofrecordscorrectly Clustered}{TotalNumberofrecords} \times 100$$

Clustering accuracy is calculated for anonymous Microsoft web dataset for PCM, FCM and the proposed WFPCM using the above formula.

Table 2 shows the comparison of the accuracy of clustering results in predicting the user behaviour for the proposed WFPCM method with the PCM and FCM. The number of clusters [6] was varied from 2 to 10 for a predetermined 10 number of iterations and the most excellent results were obtained at the end of the iterations.

Figure 3 shows accuracy comparison of PCM, FCM and proposed WFPCM to predict the user behaviour in Anonymous Microsoft Web Dataset.

Table 2. Clustering accuracies of PCM, FCM and proposed WFPCM to predict the user behaviour in anonymous Microsoft web dataset

Number of Clusters	Accuracy (%)		
	PCM	FCM	WFPCM
2	78.1	82.6	92.5
3	71.2	73.5	89.6
4	76.2	81.9	92.2
5	86.5	93.5	96.1
6	69.2	73.3	89.9
7	72.5	79.8	93.4
8	63.9	70.9	88.2
9	66.2	73.3	91.5
10	71.9	78.9	93.5

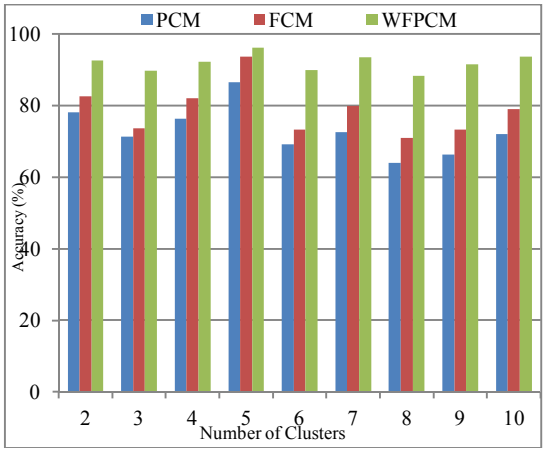

Figure 3: Accuracy comparison of PCM, FCM and proposed WFPCM to predict the user behaviour in anonymous Microsoft web dataset

Figure 3 reveals that the proposed WFPCM technique for predicting user behaviour provides better prediction results than the other two with more accuracy.

4.1.2 Execution Time

Figure 4shows the execution time taken by the PCM, FCM and the proposed WFPCM during the process of predicting user behaviour in anonymous Microsoft web dataset.

From the Figure 4, it is observed that this proposed WFPCM technique in predicting the user behaviour in Anonymous Microsoft Web Dataset takes very low execution time when compared with the PCM and FCM which takes 9.6 and 11.9 seconds respectively in Anonymous Microsoft Web Dataset. It can be observed that the time required for predicting the user behaviour in the proposed WFPCM technique for anonymous Microsoft web dataset is 4.5 seconds, whereas, more time is needed by other clustering techniques for predicting the user behaviour.

Experiment result clearly shows that accuracy and execution time of the WFPCM is comparatively better than PCM and FCM in anonymous Microsoft Web dataset.

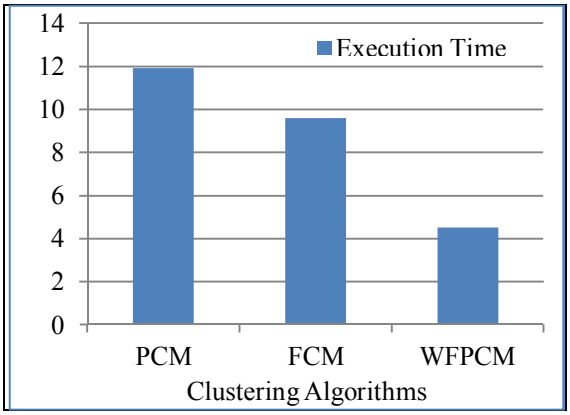

Figure 4: Comparison of time taken for predicting the user behaviour in anonymous Microsoft web dataset

5. Conclusion

The proposed method focuses on efficient web usage mining techniques. The three important phases of web usage mining are preprocessing phase, user behaviour analysis and navigation pattern discovery is concentrated in this research work. The user behaviour is predicted with the help of the Weighted Fuzzy Possibilistic C-Means (WFPCM) approach. Based on the WFPCM, an algorithm to mine user's preferred paths is predicted. To predict the user behavior existing approaches have been 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 behavior. The algorithm was implemented and the experiment result proves that this method is very effective in predicting user behavior. This approach is suitable for applications in business, such as to design personalized web service. It is observed from the experimental results that the proposed approaches outperform the other existing approaches. From the results it is confirmed that the proposed technique for extracting web user behaviour results in better prediction of user behaviours when compared to the existing web usage mining techniques.

This work can be extended to develop a path-mining algorithm for constrained navigational paths with gaps that would not make use of pre-defined thresholds.

Corresponding Author:
J. Vellingiri, Assistant Professor, Department of CSE, Kongunadu College of Engineering and Technology, Thottiam, Trichy (Dt) - 621 215, Tamil Nadu, India.
E-mail: vellingiri.research@gmail.com

References

1. D. Krol, M. Scigajlo and B. Trawinski, Investigation of internet system user behaviour using cluster analysis. International Conference on Machine Learning and Cybernetics, 2008; Vol. 6, Pp. 3408 – 3412.
2. Ling Zheng, Shuo Cui, Dong Yue and Xinyu Zhao. User interest modeling based on browsing behavior.3rd International Conference on Advanced Computer Theory and Engineering (ICACTE), 2010; Vol. 5, Pp. 455 – 458.
3. Cuifang Chen. A Fuzzy Rough Approximation Approach for Clustering User Access Patterns. WRI World Congress on Software Engineering (WCSE '09), 2009; Vol. 1, Pp. 276 – 280.
4. S. Nadi, M. Saraee and M. Davarpanah-Jazi. A fuzzy recommender system for dynamic prediction of user's behavior. International Conference for Internet Technology and Secured Transactions (ICITST), 2010; Pp. 1 – 5.
5. K. Suresh, R. Madana Mohana, A. Rama Mohan Reddy and A. Subramanyam. Improved FCM Algorithm for Clustering on Web Usage Mining. International Conference on Computer and Management (CAMAN), 2001; Pp. 1 – 4.
6. Yanfeng Zhang, Xiaofei Xu and Yunming Ye. An Agglomerative Fuzzy K-means clustering method with automatic selection of cluster number. 2nd International Conference on Advanced Computer Control (ICACC), Vol. 2,2010; Pp. 32-38.
7. Xiao-Hong Wu and Jian-Jiang Zhou. Possibilistic Fuzzy c-Means Clustering Model Using Kernel Methods. International Conference on Intelligent Agents, Web Technologies and Internet Commerce, 2005; Vol. 2, Pp. 465 – 470.
8. Yang Yan and Lihui Chen. Hyperspherical possibilistic fuzzy c-means for high-dimensional data clustering. ICICS 2009. 7th International Conference on Information, Communications and Signal Processing, 2009; Pp. 1-5.
9. Vuda. Sreenivasarao and Dr. S. Vidyavathi. Comparative Analysis of Fuzzy C- Mean and Modified Fuzzy Possibilistic C -Mean Algorithms in Data Mining. IJCST, 2010; Vol. 1, No. 1, Pp. 104-8106.
10. R. Krishnapuram and J.M. Keller. The possibilistic C-means algorithm: insights and recommendations. IEEE Transactions on Fuzzy Systems, Vol. 4, No. 3, pp. 385 – 393.
11. N. R. Pal, K. Pal and J. C. Bezdek. A mixed c-means clustering model. Proceedings of the Sixth IEEE International Conference on Fuzzy Systems, 1997; Vol. 1, pp. 11-21.
12. W.L. Hung, M Yang, D. Chen. Parameter selection for suppressed fuzzy c-means with an application to MRI segmentation, 2005; Pattern Recognition Letters.
13. Y. Zhonghang, T. Yangang, S. Funchun and S. Zengqi. Fuzzy Clustering with Novel Serable Criterion. Tsinghua Science and Technology, 2006; Vol. 11, pp. 50-53.
14. J. Srivastava, R. Cooley, M. Deshpande and PN. Tan. O. Web usage mining: discovery and applications of usage patterns from web data. SIGKDD Explorations, 2000; Vol. 1, No. 2, pp.12–23.
15. Jian Chen, Jian Yin, Tung, A.K.H. and Bin Liu. Discovering Web usage patterns by mining cross-transaction association rules. International Conference on Machine Learning and Cybernetics, 2004; Vol. 5, Pp. 2655-2660.
16. P. Zhang, X. Wang, and P. X. Song. Clustering categorical data based on distance vectors. The Journal of the American Statistical Association, 2006; Vol. 101, No. 473, pp. 355–367.
17. A. Vakali, J. Pokorný and T. Dalamagas. An Overview of Web Data Clustering Practices. EDBT Workshops, 2004; pp. 597-606.
18. Lin Zhu, Fu-Lai Chung and Shitong Wang. Generalized Fuzzy C-Means Clustering Algorithm with Improved Fuzzy Partitions, 2009; IEEE Transactions on Systems.
19. Cheul Hwang, Frank Chung-Hoon Rhee. Uncertain Fuzzy Clustering: Interval Type-2 Fuzzy Approach to C-Means, 2007; IEEE Transactions on Fuzzy Systems.
20. Chu-Hui Lee and Yu-Hsiang Fu. Web Usage Mining Based on Clustering of Browsing Features. Eighth International Conference on Intelligent Systems Design and Applications, 2008; Vol. 1, Pp. 281-286.

10/25/2018