Original Research Paper

Recommendation Engine Formation Using Depth First Search and Genetic Approach

[1]J.S. Kanchana and [2]S. Sujatha

[1]*Department of IT, K.L.N College of Engineering, Sivagangai District, Tamil Nadu, India*
[2]*Department of MCA, Anna University of Technology, Trichy, Tamil Nadu, Pincode- 620 024, India*

Abstract: The requirement of online users in the website varies dynamically. The recommendation of web pages consisting of user expected information and data is performed by the online recommendation system. The recommendation engine must be self-adaptive and accurate. The existing algorithm uses Depth First Search (DFS) and bee's foraging approach to create navigation profiles by categorizing the current user activity. The prediction of navigations that are most expected to be visited by online users is also performed. In this study, the recommendation engine formation with optimized resource such as memory, CPU usage and minimum time consumption is proposed using DFS and Genetic Approach (GA). Here, initially the cluster formation is achieved using DFS approach. The method creates an eminent browsing pattern for each user using live session window. The performance of the approach is compared with the existing forager agent. The experimental results show that the proposed approach outperforms the existing methods in accomplishing accurate classification and anticipation of future navigation for the current online user.

Keywords: DFS Approach, Forager Agent, Genetic Approach, Imminent Browsing Pattern, Recommendation Engine

Introduction

Web Usage Mining (WUM) is the process of determining what the users are looking for on the internet. Some users might be interested in only textual data and some others might be looking at multimedia data. WUM is the application of data mining techniques to find the interesting usage patterns from web data to serve the needs of web based applications effectively. Usage data discover the identity of web users along with their browsing behavior at a web site. The web usage mining can be categorized depending on the type of usage data considered. In web server data, the user logs are gathered by the web server. Typical data consists of an IP address, access time and requested page.

The recommendation system is a subclass of the information filtering system that attempts to predict the rating or preference that user would provide to an item. For when a product is viewed on amazon.com, the store will recommend additional items depending on a matrix of what other shoppers purchased along with the currently selected item. Many recommender systems avoid unpopular and newly introduced items. It is a perfect alternative to search algorithms. It focuses on the items with enough ratings to be utilized in recommendation algorithms. Recommender systems generate a list of recommendations in one of the two ways:

- Collaborative filtering
- Content-based filtering

The collaborative filtering approach creates a model from the user's past behavior and also similar decisions made by other users. The items that the user can have an interestare predicted using that model. The content-based filtering approach uses a series of discrete characteristics of an item to recommend further items with similar properties. The hybrid recommender system combines collaborative and content based filtering.

The online recommendation system is used to recommend the navigations that are most likely to be searched or visited by the online users in the future. This is performed by the categorization of the current user activity into navigation profiles. The existing algorithm uses Depth First Search (DFS) for the formation of clusters along with the bee's foraging approach. It chooses the more profitable, efficient navigation profile for the current user activity. This Recommendation Engine (RE) system is self adaptive by capturing the changing needs of online user. Self adaptive systems have the ability to self adapt and reorganize itself. In this study, the recommendation engine formation with optimized resources such as memory, CPU usage and minimum time consumption is proposed. It uses DFS and genetic approach for the creation of eminent browsing pattern for each user. It includes the creation of a live session window. The proposed approach outperforms the existing methods by the effective and accurate prediction of current online user's future navigations.

The remaining part of the paper is organized as follows. Section II involves the works related to the recommendation engine formation. Section III involves a brief description of the existing method-bee's foraging approach in combination with DFS. Section IV involves the description of the proposed method-DFS with a Genetic Approach (GA). Section V involves the performance evaluation and comparison of GA and existing techniques in terms of memory and time consumption. The paper is concluded in section VI.

Related Work

Gupta and Shrivastava (2012) designed an enhanced genetic Fuzzy C-Means (FCM) algorithm for web usage data clustering. In this paper, they suggested a new framework to enhance the web session's cluster quality from FCM using improved Genetic Algorithm (GA). Ali and Sheikh (2014) suggested a review for web usage mining. In this paper, they focused on the process of web mining, which, involved basically three imperative tasks includes, preprocessing, pattern discovery and pattern analysis. Paramasivam and Srinivasan (2013) used web usage mining and clustering sessions for perfection and phychiatring user profile in websits. In this paper, a complete framework and findings were presented in mining, web usage patterns from web log files on a real web site. Chaudhary and Gupta (2013) recommended a tools and methods for web usage mining. This paper has attempted to provide a review for the rapidly growing area of web usage mining. This paper was concentrated on web usage mining area and its application in various fields. Yin and Guo (2013) suggested a new formulation for the Website Structure Optimization (WSO) problem based on a comprehensive survey of existing works and practice considerations. Mohanraj *et al*. (2012) implied an Ontology Driven Bee's Foraging Approach (ODBFA) based self adaptive online recommendation system.

In this study, they contemplated the ODBFA that exactly classified the current user activity to any of the navigation profiles. Hung *et al*. (2013) suggested a paper about web usage mining for scrutinzing elder self care behavior patterns. The main intent of this study was to understand the self care service. Karaboga and Ozturk (2011) implied a novel clustering approach based Aritifical Bee Colony (ABC) algorithm. In this analysis, ABC was used for data clustering on benchmark problems and the performance of the ABC algorithm was compared with Particle Swarm Optimization (PSO) algorithm. Bojnord (2014) suggested a Swarm intelligence approach for clustering based on fuzzy honey bees foraging optimization. In this scrutiny, fuzzy operator was employed to enhance the performance of the proposed approach and prevent premature convergence.

Kumar and Hemalatha (2014) suggested an enhanced algorithm based on Artificial Bee Colony (ABC) optimization. Based on classical association rule mining, a new approach has been developed expanding it by using fuzzy sets. Upadhyay and Purswani (2013) recommended a paper for pattern discovery of web usage mining. Web usage mining describes the discovery of useful from the web contents. Gong (2011) presented an enhanced person correlation similarity measure method in the personalized collaborative filtering recommendation algorithm. Omkar *et al*. (2011) implied an Artificial Bee Colony (ABC) algorithm for multi obejctive design optimization of composite strutures. In this paper, they presented a generic method for multi objective design optimization of laminated composite components, based on Vector Evaluated ABC.

Mohanraj and Chandrasekaran (2011) contemplated an ontology based approach to implement the online recommendation system. In this study, they presented an ontology based recommendation framework for identifying the user intent and predict future browsing pattern of online user using novel web usage mining method. Rathipriya *et al*. (2011) retrieved the global optimal bicluster from the web usage data. In this work, the swarm intelligent technique was combined with biclustering approach.

Salehi *et al*. (2012) suggested a new recommendation approach based on implicit attributes of learning material. In this analysis, Genetic Algorithm (GA) was used for attribute weight optimization to solve sparsity problem. Jarukasemratana and Murata (2013) designed a web caching replacement algorithm based on web usage data. They developed a system that record user's browsing behavior at the resource level. Suresh *et al*. (2011) implied an enhanced FCM algorithm for clusteirng on web usage mining. They presented a clustering web usage data, which is useful in finding the user access patterns and the order of the visits of the hyperlinks of the each user. Senkul and Salin (2012) investigated the effect of semantic information on the patterns generated for web usage mining in the form of frequent sequences. Bhushan and Nath (2012) suggested a paper for an automatic recommendatin of web pages for online users using web usage mining. In this paper, they implied a web recommendation approach which recommends user a list of pages based upon user's historic pattern.

Bee's Foraging Approach

In this section, the existing bee's foraging approach for the online recommendation engine formation is described. The approach utilizes navigation pattern mining to organize the similar navigation pattern of user into clusters or navigation profiles. Each navigation profile is a set of identical web page access of users across various user sessions. This cluster formation is performed by DFS approach. The initialization of Foragers corresponding to the number of clusters in the navigation profile is performed by the Onlooker agent.

Each Forager Agent (FA) executes greatest common subsequence detection in the navigation profiles to detect the imminent browsing patterns or the subsequence of web user. The Onlooker Agent (OA), after receiving the profitability score and discovered subsequence from each FA, determines the closely competing navigation profiles. It is submitted as input to the recommendation engine. The selection of the best profitable source of navigation profile is performed by the recommendation engine after receiving profitable sore and subsequence. Score based similarity comparison is used in the selection of right set of pages from clusters. The output of the recommendation engine is the imminent browsing patterns of the user.

This hierarchy of operations by the FA, OA and the recommendation engine for the selection of profitable score and imminent browsing pattern enhance the time and resource consumption of the approach. The limitations of the bee's foraging approach are overcome by using DFS along with the Genetic Approach (GA).

Genetic Approach

Figure 1 Shows the creation of recommendation engine using the genetic approach. Initially, when the user accesses information on the web pages, the details such as hostname, session id, page request, no of visits and session in and out time are stored in the log file.

Web Page Linking

All the web pages accessed details during various sessions obtained from the log system is used for the creation of undirected acyclic graph G = (V, E). Figure 2 shows the undirected acyclic graph formed from the log system details. It consists of 17 vertices. Each vertex V denotes the web page present in the web server model. In the graph, the node 1 is directly connected to the vertices 2, 3 and 13. The vertex 2 is directly connected to 5 vertices.

The edge weight is then determined for all edges in the graph using Equation 1:

$$EW = \frac{S_{ij}}{max\left\{S_{ij}\right\}} \tag{1}$$

Here, EW is the edge weight of the edge (i, j) in the graph. Denotes the number of sessions containing both pages i and (i, j) and are the number of sessions containing page i and page j respectively. The determined edge weights are stored in the Weighted Adjacency matrix (WA). Each entry EW in the matrix contains the value calculated using Equation 1. Table 1 depicts a sample WA for 5 vertices in the graph. A threshold value is set by the observation of web page usage. The elements of EW whose value is less than the specified threshold (for e.g., 0.3), is removed to reduce the number of edges in the graph.

Population Generation

The identical navigation patterns of users are grouped into clusters by applying Depth First Search (DFS) for each of the nodes in the undirected graph. The vertices directly reachable from each and every vertex are determined by the DFS approach (Table 2). The number of clusters created is minimized by applying minimum cluster size threshold (for e.g., 2).
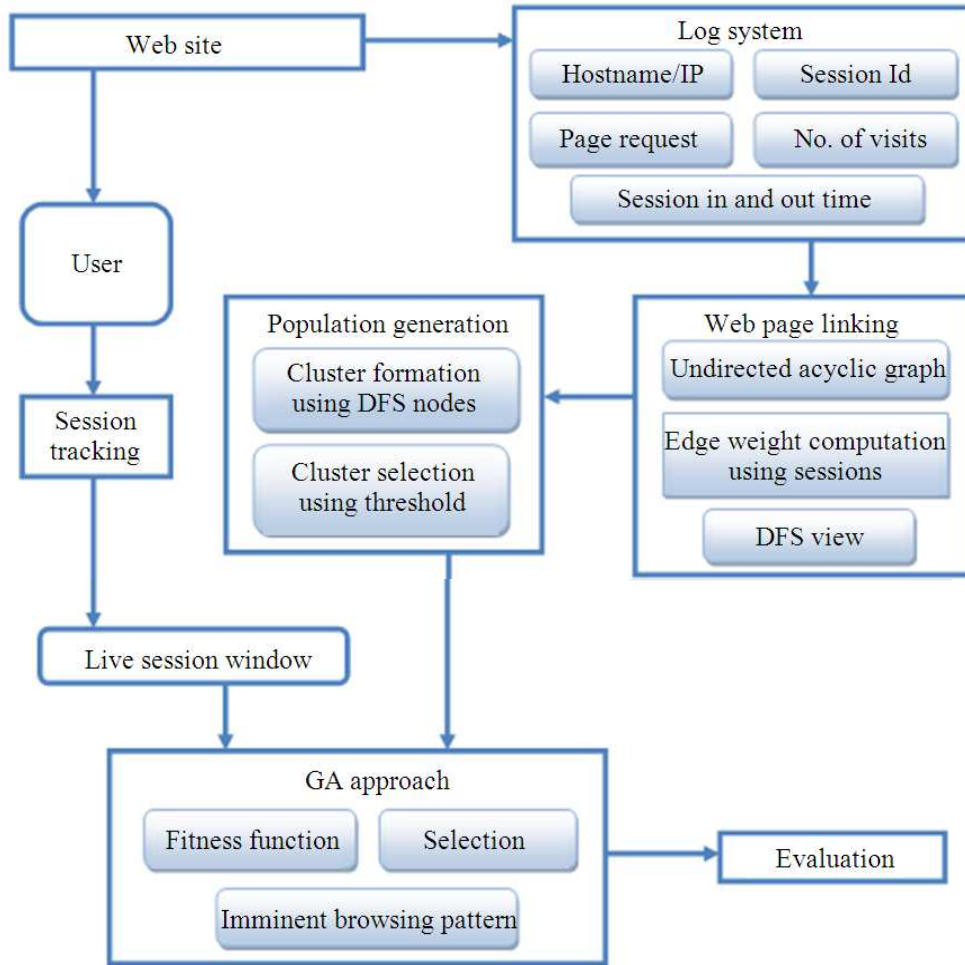
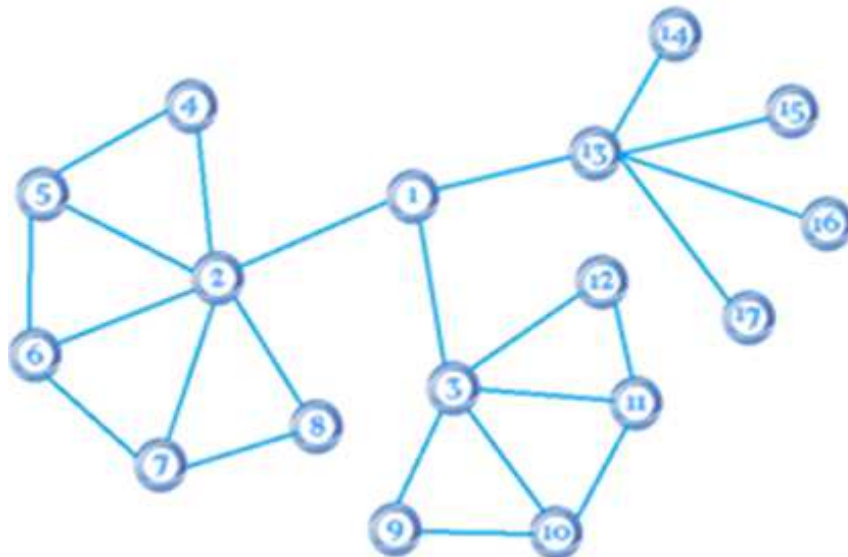Fig. 1. Genetic Approach (GA) for recommendation engine formation

Fig. 2. Undirected acyclic graph formation

Table 1. Weighted Adjacency matrix (WA)

Vertex	1	2	3	4	5
1	0.000	1.296	1.0	0.000	0.000
2	1.296	0.000	0.0	1.059	0.296
3	1.000	0.000	0.0	0.000	0.000
4	0.000	1.059	0.0	0.000	0.462
5	0.000	0.296	0.0	0.462	0.000

Table 2. Cluster formation by DFS after applying threshold

Cluster	nodes
C1	{1, 2, 3, 13}
C2	{2, 4, 5, 6, 7, 8}
C3	{3, 9, 10, 11, 12}
C4	{13, 14, 15, 16, 17}

The live session window ({14, 13, 15, 13, 1, 3, 10, 11, 9, 3, 1, 13, 17}) created from the session tracking on the website and the cluster formed by the DFS is given as input for the fitness value computation of the genetic approach.

Fitness Value Computation

The number of live session window matches with the GA data is performed. Table 3 describes the URL matches in LSW and the created clusters. The number of live session window matches with each and every cluster is divided by the total number of URL count in the live session window to get the fitness value. Here, the total number of URL count in the LSW is 13. For the number of LSW match in cluster 1 is 3. This value is divided by 13 to retrieve the fitness value 0.23 for C1.

A threshold value is calculated to filter the minimum fitness value. The computation of threshold value is performed using Equation 2:

$$threshold\ value = \frac{minimum\ fitness\ value + maximum\ fitness\ value}{2} \quad (2)$$

Normalized fitness value for every path is computed using Equation 3. Table 4 shows the normalization of fitness value for three clusters:

$$normalized\ fitness\ value = \frac{fitness\ value}{sum\ of\ all\ fitness\ value} \quad (3)$$

Table 5 depicts the computation of an accumulated fitness value.

The difference between accumulated fitness value and the normalized fitness value of every path is determined to retrieve the Browsing Pattern (BP) value. Table 6 shows the computation of BP value.

Table 3. Number of matches in Live Session Window (LSW) and clusters

Path	Number of matches
LSW and C1	3
LSW and C2	0
LSW and C3	4
LSW and C4	4

Table 4. Normalization of fitness value

Path	Fitness	Normalized fitness value
LSW and C1	0.23	0.27
LSW and C3	0.31	0.36
LSW and C4	0.31	0.36

Table 5. Calculation of accumulated fitness value

Path	Normalized fitness value	Accumulated fitness value
LSW and C1	0.27	0.27
LSW and C3	0.36	0.63
LSW and C4	0.36	0.99

Table 6. Browsing pattern value

Path	Computation	BP value	Diff (%)
LSW and C1	0.27-0.27	0.00	0.00
LSW and C3	0.63-0.36	0.27	42.86
LSW and C4	0.99-0.36	0.63	63.64

The path with has zero BP value is discarded. The pattern with the minimum BP value is the imminent browsing pattern. Here, the path with the imminent browsing pattern value (42.86%) is LSW and C3. The imminent browsing pattern retrieved is {3, 9, 10, 11, 12}.

Generation of Imminent Browsing Pattern

The accumulated fitness value is calculated using the normalized fitness value. i.e., the accumulated fitness value of LSW and C1 is equal to its normalized fitness value. The accumulated fitness value of LSW and C3 is obtained by adding its normalized fitness value to the normalized fitness value of LSW and C1 and so on.

Performance Analysis

The proposed genetic approach for the imminent browsing pattern generation is compared with the bee's foraging approach. The performance analysis shows that the proposed approach achieves reduced time and memory consumption, minimum CPU usage, greater accuracy in less time than the existing methods.
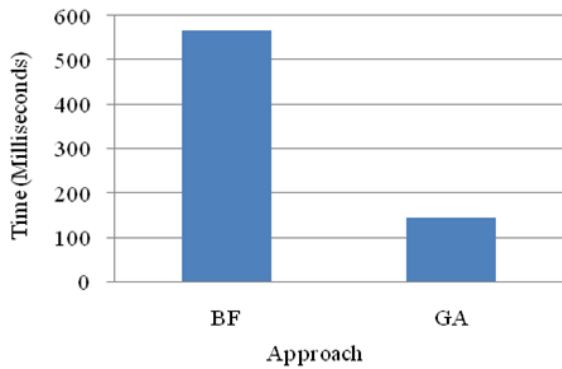
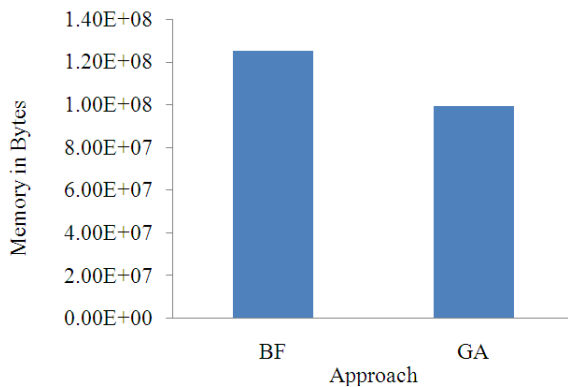Fig. 3. Time consumption of bee's foraging and genetic approach

Fig. 4. Memory consumption of bee's foraging and genetic approach

Time Consumption

Figure 3 Presents the comparison of time consumption of Bee's Foraging approach (BF) and the Genetic Approach (GA). The analysis shows that GA achieves less time consumption than BF.

Memory Consumption

The memory consumption comparison of the Bee's Foraging (BF) and the Genetic Approach (GA) is shown in Fig. 4. The analysis describes that the proposed GA requires less memory usage when compared to BF.

The CPU usage of GA is compared with the BF. The result is shown in Fig. 5. The analysis shows that the proposed approach achieves minimum CPU usage in percent than the BF.

Accuracy

The accuracy of GA is compared with the BF. The result is shown in Fig. 6. The analysis shows that the proposed approach achieves greater accuracy than the BF.

Figure 7 shows the accuracy and time consumption of GA and BF. The analysis shows that the proposed approach achieves greater accuracy in less time when compared to BF.
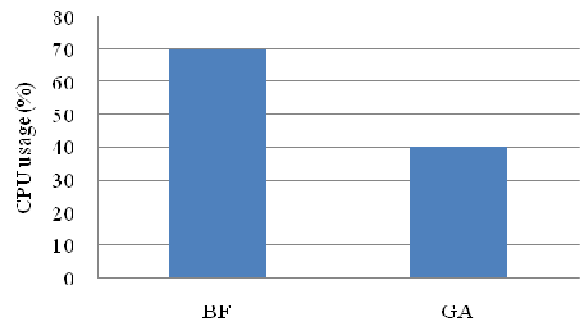

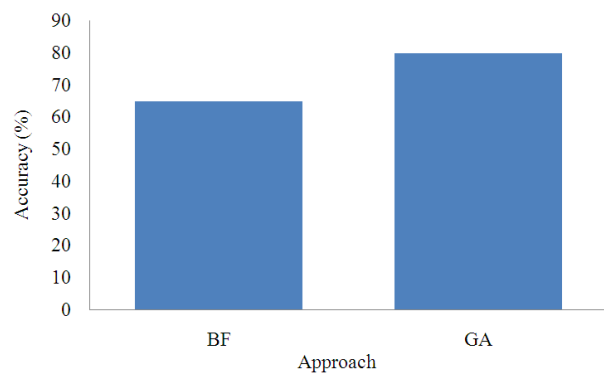Fig. 5. CPU usage of bee's foraging and genetic approach

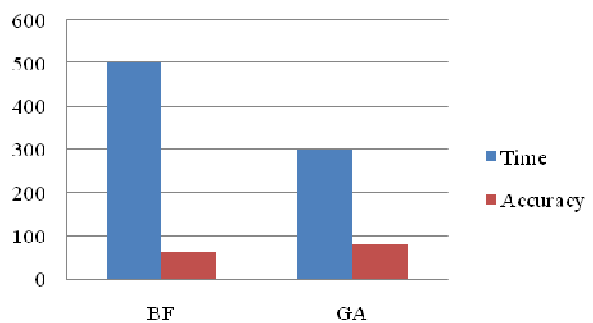Fig. 6. Accuracy of bee's foraging and genetic approach

Fig. 7. Accuracy and time consumption of bee's foraging and genetic approach

Conclusion

An effective recommendation engine creation using DFS and Genetic Approach (GA) is proposed. The method generates imminent browsing pattern for each user using live session window. Initially, DFS approach is applied to group the similar navigation patterns of users into clusters. The created cluster and Live Session Window (LSW) from session tracking is processed by GA. The proposed GA is analyzed against time, accuracy and resource usage. The performance of the approach is compared with the bee's foraging approach. The experimental results

show that the genetic approach achieves optimized time, memory and resource consumption, greater accuracy in less time when compared to the existing methods. As a future work, with more number of web pages and different categories, sequential category prediction can be performed.

Funding Information

The authors have no support or funding to report.

Author's Contributions

All authors equally contributed in this work.

Ethics

This article is original and contains unpublished material. The corresponding author confirms that all of the other authors have read and approved the manuscript and no ethical issues involved.

References

Ali, R.S. and R.H. Sheikh, 2014. Web usage mining-a review. CSI Digital Resource Center.

Bhushan, R. and R. Nath, 2012. Automatic recommendation of web pages for online users using web usage mining. Proceedings of the International Conference on Computing Sciences, Sept. 14-15, IEEE Xplore Press, Phagwara, pp: 371-374. DOI: 10.1109/ICCS.2012.17

Bojnord, I., 2014. Fuzzy honey bees foraging cptimization: Swarm intelligence approach for clustering. J. Artificial Intell., 7: 13-23. DOI: 10.3923/jai.2014.13.23

Chaudhary, K. and S.K. Gupta, 2013. Web usage mining tools and techniques: A survey. Int. J. Scientific Eng. Res., 4: 1762-1768.

Gong, S.J., 2011. An enhanced similarity measure used in personalized recommendation algorithms. Adv. Materials Res., 159: 671-675. DOI: 10.4028/www.scientific.net/AMR.159.671

Gupta, K. and M. Shrivastava, 2012. Web usage data clustering using improved genetic fuzzy c-means algorithm. Int. J. Adv. Comput. Res., 2: 77-79.

Hung, Y.S., K.L.B. Chen, C.T. Yang and G.F. Deng, 2013. Web usage mining for analysing elder self-care behavior patterns. Expert Syst. Applic., 40: 775-783. DOI: 10.1016/j.eswa.2012.08.037

Jarukasemratana, S. and T. Murata, 2013. Web caching replacement algorithm based on web usage data. New Generat. Comput., 31: 311-329. DOI: 10.1007/s00354-013-0404-z

Karaboga, D. and C. Ozturk, 2011. A novel clustering approach: Artificial Bee Colony (ABC) algorithm. Applied Soft Comput., 11: 652-657. DOI: 10.1016/j.asoc.2009.12.025

Kumar, K.S. and M. Hemalatha, 2014. An innovative potential on rule optimization using fuzzy artificial bee colony. Res. J. Applied Sci. Eng. Technol., 7: 2627-2633.

Mohanraj, V. and M. Chandrasekaran, 2011. An ontology based approach to implement the online recommendation system. J. Comput. Sci., 7: 573-581. DOI: 10.3844/jcssp.2011.573.581

Mohanraj, V., M. Chandrasekaran, J. Senthilkumar, S. Arumugam and Y. Suresh, 2012. Ontology driven bee's foraging approach based self adaptive online recommendation system. J. Syst. Software, 85: 2439-2450. DOI: 10.1016/j.jss.2011.12.018

Omkar, S., J. Senthilnath, R. Khandelwal, G.N. Naik and S. Gopalakrishnan, 2011. Artificial Bee Colony (ABC) for multi-objective design optimization of composite structures. Applied Soft Comput., 11: 489-499. DOI: 10.1016/j.asoc.2009.12.008

Paramasivam, S.P. and D.S. Srinivasan, 2013. Perfections and psychiatring user profile in web sites using web usage mining and clustering sessions. IJRCCT, 2: 78-85.

Rathipriya, R., K. Thangavel and J. Bagyamani, 2011. Binary particle swarm optimization based biclustering of web usage data. Int. J. Comput. Applic., 25: 43-49. DOI: 10.5120/3001-4036

Salehi, M., I.N. Kmalabadi and M.B.G. Ghoushchi, 2012. A new recommendation approach based on implicit attributes of learning material. IERI Procedia, 2: 571-576. DOI: 10.1016/j.ieri.2012.06.136

Senkul, P. and S. Salin, 2012. Improving pattern quality in web usage mining by using semantic information. Knowl. Inform. Syst., 30: 527-541. DOI: 10.1007/s10115-011-0386-4

Suresh, K., R. MadanaMohana, A. RamaMohan 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

Upadhyay, M.A. and M.B. Purswani, 2013. Web usage mining has pattern discovery. Int. J. Scientific Res. Publicat., 3: 1-4.

Yin, P.Y. and Y.M. Guo, 2013. Optimization of multi-criteria website structure based on enhanced tabu search and web usage mining. Applied Math. Comput., 219: 11082-11095. DOI: 10.1016/j.amc.2013.05.033