

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/341426362>

# Biogeography optimization algorithm based next web page prediction using weblog and web content features

Article · June 2020

DOI: 10.11591/ijai.v9.i2.pp327-335

CITATION

1

READS

65

2 authors:



Roshan Gangurde

K. K. Wagh Institute of Engineering Education and Research

11 PUBLICATIONS 15 CITATIONS

[SEE PROFILE](#)



Dr Binod Kumar

Rajarshi Shahu College of Engineering

66 PUBLICATIONS 61 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Phd. Thesis Forum Review [View project](#)



Developing optimal predictive models using data mining [View project](#)

# Biogeography optimization algorithm based next web page prediction using weblog and web content features

Roshan A. Gangurde<sup>1</sup>, Binod Kumar<sup>2</sup>

<sup>1</sup>Department of Computer Science, Savitribai Phule Pune University, Maharashtra, India

<sup>2</sup>Jayawant Institute of Computer Applications, Pune, Maharashtra, India

## Article Info

### Article history:

Received Dec 25, 2019

Revised Feb 29, 2020

Accepted Mar 1, 2020

### Keywords:

Association rule mining

Biogeography optimization

Ontology

Prediction model

Web page recommendation

## ABSTRACT

Recommendation of web page as per users' interest is a broad and important area of research. Researcher adopts user behavior from actions present in cookies, logs and search queries. This paper has utilized a prior webpage fetching model using web page prediction. For this purpose, web content in form of text and weblog features are analyzed. As per dynamic user behavior, proposed model LWPP-BOA (Logistic Web Page Prediction By Biogeography Optimization Algorithm) predict page by using genetic algorithm. Based on user actions, weblog feature are developed in form of association rules, while web content gives a set of relevant text patterns. Page prediction as per random user behavior is enhanced by means of Biogeography Optimization Algorithm where crossover operation is performed as per immigration and emigration values. Here population updation depends on other parameters of chromosome except fitness value. Experiments are conducted on real dataset having web content and weblogs. Results are compared using precision, coverage, M-Metric, MAE and RMSE parameters and it indicates that the proposed work is better than other approaches already in use.

*This is an open access article under the [CC BY-SA](#) license.*



## Corresponding Author:

Roshan A. Gangurde,  
Research Scholar, Department of Computer Science,  
Savitribai Phule Pune University,  
Maharashtra, India.  
Email: roshanant@gmail.com

## 1. INTRODUCTION

The network is very big, dissimilar and ever changing. Exciting information extraction from Web information has become very trendy and as an outcome of that, web mining has fascinated lots of awareness in recent time. Web mining is a function of web information mining from huge web information repositories [1]. At this time, the setback of modeling and forecasting a user's web browsing performance on the internet has fascinated lot of researchers as it can be useful in developing web cache performance, Web Page Recommendations, Search Engine Optimizations, accepting and influencing buying samples, and practicing the web search experience of the users [2]. In e-commerce, the upcoming page forecasting is extremely essential and serious. The calculation supports corporations to handle the issues concerning to users such as their movements in purchasing and interests of the users in any specific products.

Web usage mining can be utilized to very cautiously examine web log records which are collected in web servers for pattern detection. Web mining chiefly concentrates on the preprocessing section and grouping section. Defining a cluster is hard and that is why, we find dissimilar kinds of algorithms connected to clustering. The existence of collection of records is the normal thing that connects all these

algorithms jointly. It comprises of three parts: preprocessing, pattern detection and pattern examination. Preprocessing is mandatory to change the raw information into a significant form helpful for well-organized processing. Pattern detection comprises methods to take out the pattern and includes numerical analysis, sequential pattern mining, path analysis, relationship rule mining, categorization, and bunching [3-4].

This paper has attempted to improve the accuracy of the next web page prediction where accuracy of prediction depends on the utilization of web mining features. In previous research, Markov model was utilized for weblog feature [5]. This needs to be replaced by other mathematical model where web content or web structure feature are also used. As web page prediction is highly dynamic in nature, a suitable genetic algorithm need to be proposed in collaboration with proper utilization of association rules to increase the accuracy of prediction model.

## 2. RELATED WORK

In [6], author proposed an approach hybrid utilizing Support Vector Machines, and the All-Kth Markov model, to determine calculation utilizing Dempster's rule. To boost the power of discrimination, they apply attribute extraction of SVM. Adding together, during forecast it occupies domain information to decrease the number of classifiers for the development of correctness and the decrease of calculation time.

In [7], author explained internet mining algorithm that targets at modifying the analysis of the draft's production of connection rule mining. This algorithm is being extremely utilized in internet mining. The end results achieved established the strength of the algorithm projected in this paper.

In [8], author concentrated on Web Usage Mining wherein the internet user routing patterns and their utilization of web resources are discovered. The dissimilar phases occupied in this mining procedure and with the proportional study between the pattern detection algorithms: Apriori and FP-growth algorithm.

In [9], author executed the preprocessing methods to change the log file into client sessions which are appropriate for mining and decrease the range of the session file by sorting the least demanded pages utilizing the preprocessing method. Information Preprocessing is one of the significant missions before inserting mining algorithms. It changes the raw record file into client session. In this vocation, they have in brief established record file preprocessing and applied it in a CTI record file. Also, they created the review of the client session file. They have utilized filtering method to eradicate slightest demanded resources.

In [10], author projected utilizing GRUs to study further meaningful aggregation for client browsing history (browsed news), and suggested news articles with hidden feature model. The end result show a considerable development compared with the conventional word-based approach. The system has been fully deployed to online construction services and helping over ten million distinctive user's everyday.

In [11], author planned a sequential DSSM model which incorporates RNNs into DSSM for suggestions. Based on conventional DSSM, TDSSM put back the left network with point static characteristics, and the right network with two sub-networks to modeling client static characteristics (with MLP) and client sequential features (with RNNs).

In [12], author projected an incorporated structure with CNNs and RNNs for modified key structure (in videos) suggestion, in which CNNs are utilized to study characteristic demonstrations from key frame images and RNNs are utilized to practice the textual characteristics.

## 3. PROPOSED METHODOLOGY

The complete work is broadly divided in two phases. The first phase is developing ontology by using Web Content (WC) and Weblog (WL) features as shown using Figure 1. Second phase is testing where input is users' recent visited web pages and ontology. So output of second model is predicted web page. The block diagram of proposed LWPP-BOA model is as shown in Figure 2. Table 1 describes the symbols used in this research paper.

### Dataset pre-processing

There are many digital features of the web portal. In order to predict the next web page, the proposed model works on web content which is in the form written text and weblog is in form of user previous sequence of actions. So raw dataset need to be pre-processed first, by removing the noisy data.

### Web content pre-processing

As text content on web page have words which need pre-processing by removing stop-words. So set of stop-words are removed and filtered words are further processed to collect web keywords. Further, each web page has its own set of keywords that depends on type of content present in the web page.

There may be chances of common keywords that may exist between pages. So let pages be  $P_1, P_2, P_m$ , where  $m$  is total number of pages in the site. Now  $P_m$  page has web content  $\{w_1, w_2, s_1, w_3, s_2, s_1, \dots, w_n\}$   $n$  is total number of words in  $P_m$  page. After removal of stop-words  $\{s_1, s_2, \dots\}$ , important words will be  $\{w_1, w_2, w_3, \dots, w_n\}$  [13]. Finally to get keywords, Term Frequency is evaluated from the available words and the words which cross minimum frequency act as keywords of that page. Here, sequence of words of the web page does not affect the pre-processing outcomes.

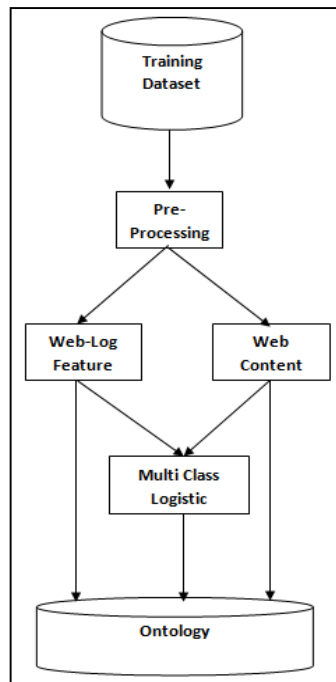


Figure 1. Block Diagram of Developing Ontology

Symbol used	Meaning of Symbol
$U_i$	$i^{\text{th}}$ Web User
$P_m$	$m^{\text{th}}$ Web Page
$K_t$	$t^{\text{th}}$ Web Page Keyword
WL	Weblog
$AR_r$	$r^{\text{th}}$ Association Rule
FFC	Feed Forward Counter
SC	Set Counter
$L_p$	Logistic Probability
H	Habitat
$\lambda$	Emigration Rate
$\alpha$	Immigration Rate
$M_p$	Mutation Probability
H	Population Size

#### Pre-processing of weblog

In this step, unnecessary columns from the Weblog dataset are removed except IP address and access URL request. Here each visited page is identified by a unique page number. So Weblog WL has a set of visited pages  $(P_1, P_2, P_3, \dots, P_m)$  of a single user  $U_i$ .

#### Feature generation from weblog and web content

Rules obtained from weblog are generated by setting the counter for all set of patterns from the pages [14]. So for  $n$  pages, set of counters SC is represented by (1).

$$SC = \frac{n(n+1)}{2} \quad (1)$$

Hence, for  $n=3$ ,  $SC=6$  means  $SC=\{(1), (2), (3), (1,2), (1,3), (1,2,3)\}$  so cardinality of  $|SC|$  is 6.

Association rule generation by above counter technique reduced number of dataset scan to one pass [15]. Hence this rule generation is termed as Feed Forward Counter (FFC).

$$AR \leftarrow FFC(WL, SC) \quad (2)$$

#### Similarity matrix (SM)

Keywords obtained from the pages after pre-processing are arranged in matrix where number of row and column are equal to number of pages. So matrix  $SM_{m \times m}$  have cell values which is the count of similar words between 2 pages in row and column. For example, page  $P_1 \{K_1, K_2, K_3\}$  and  $P_2 \{K_4, K_1, K_5\}$  have similarity value 1 as keyword  $K_1$  is common in both the pages.

### Logistic regression

LHS page patterns of the Association rules are collected to generate the regression value as the probability of RHS page to be found. This can be understood as follows: Consider the rules,  $P_1, P_5 \rightarrow P_{10}$  and  $P_1, P_5 \rightarrow P_{12}$ . Both these rules have common set of LHS page pattern ( $P_1, P_5$ ) and different RHS pattern. So regression will generate probability of next page by using confidence value of the rule and similarity matrix. So input to regression is one feature which is the multiplication of similarity matrix and confidence value. In above example, two categories were found, page  $P_{10}$  and  $P_{12}$ .

So input to the logistic function is  $C_1 = \text{Conf}(P_1, P_5 \rightarrow P_{10})$ ,  $C_2 = \text{Conf}(P_1, P_5 \rightarrow P_{12})$

The input parameter for regression calculation is shown with the help of example as shown in Table 2.

Table 2. Input parameters for regression calculation

Category 1 (X) $P_{10}$	Category 2 (X) $P_{12}$
$SM[P_1, P_{10}] \times C_1$	$SM[P_1, P_{12}] \times C_2$
$SM[P_5, P_{10}] \times C_1$	$SM[P_5, P_{12}] \times C_2$
$SM[P_1, P_5] \times C_1$	$SM[P_1, P_5] \times C_2$

Output of this logistic function gives an intercept  $\beta_0$  and  $\beta_1$ . Finally by putting values of predictors, one can get probability values as shown in (3):

$$L_p = \frac{1}{1 + e^{-(\beta_0 + X_1 \times \beta_1 + X_2 \times \beta_2)}} \quad (3)$$

### Filter rule

Those rules which have very low support values should be filtered from the generated rules, as these rules act as noise. Hence one threshold value needs to be set as per dataset rules. Association rules crossing minimum support value are used for page prediction in genetic algorithm.

### Testing phase

Here, the dataset is again preprocessed for the weblog portion in order to get the logs that are used for testing the built model. Pre-Processing steps are similar as done in previous steps of model. The only difference here is that pre-processed logs are break such that LHS pages of the association rule are in the testing part and the RHS page of the association rule is used for the evaluation of result.

### Biogeography optimization algorithm (BOA)

Species in nature adopt changes as per suitable environmental conditions. So change of habit is one type of change adopted by species time to time. Based on this, author proposed a mathematical algorithm in early 1960, wherein the main concern of this model is to understand the migration of species from one habitat to another [16]. Biogeography was a trending research area at that time. So in 2008, author proposed a generalized genetic algorithm to resolve similar type of issues [17]. Some of basic terms related to this work are:

#### Habitat suitability index (HSI)

This is termed as fitness value of the habitat, means higher value shows that poor place to live while low value means good place to live in terms of resources, life, etc.

#### Immigration and emigration rate

Some of basic terms of Immigration  $\lambda$  and Emigration  $\alpha$  was done by:

$$\lambda_R = (1 - \alpha_R) \quad (4)$$

$$\alpha_R = \frac{R}{h} \quad (5)$$

where R is rank of habitat in terms of HSI value, while h is total number of habitats.

#### Generate habitats

Possible set of solutions which are termed as habitat in this algorithm are generated in this step. Each habitat is set of possible pages obtained from association rule where left side pages of the rule are the visited pages

of the user. Hence habitat is combination of  $H=\{P_1, P_5, P_m\}$ , where population has total  $h$  number of habitats. Hence population generation function in this algorithm is shown by (6).

$$H \leftarrow \text{Habitat}(m, h, AR) \quad (6)$$

The block diagram of the proposed LWPP-BOA (Logistic Web Page Prediction by Biogeography Optimization Algorithm) model is shown in Figure 2.

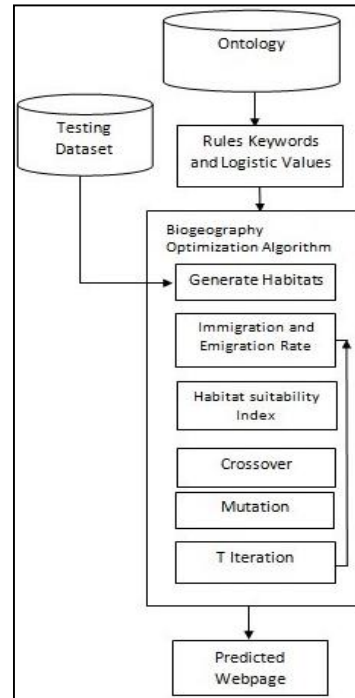


Figure 2. Block diagram of proposed LWPP-BOA model.

#### Fitness function

Habitat Suitability Index (HSI) of any habitat depends on the summation of logistic values obtained from the ontology stored during training of the model. Hence rank  $i$  of the habitat depend on the logistic probability value  $L_p$  of each predicted page as per visited pages for testing log.

$$HSI = \text{Rank}(L_p, T_l, H)$$

#### Crossover

Emigration of pages in the form of species from one habitat to other depends on emigration rate. While permitting the species to enter in a habitat depends on immigration rate. Hence for crossover from one habitat to other, both type of rate need to find. So crossover depends on following condition.

```

Loop x=1:h
  If Crossover_Limit > λR
    Loop y=1:h
      If Crossover_Limit > αR
        m ← Rand()
        H[x, m] ← H[y, m]
      EndIf
    EndLoop
  EndIf
EndLoop
  
```

Where Crossover\_Limit is random number range between 0-1,  $x$  and  $y$  is habitat position specify immigration, emigration operation.

### Mutation

In this work, after crossover, mutation is also performed. So chances of new solutions get increases. For this, author has involved mutation probability where mutation is performed in selected habitats as per HSI value.

$$M_R = \frac{HSI_R}{\text{sum}(h)} \quad (7)$$

$$M_p = \frac{M_R}{\text{Max}(M_R)} \quad (8)$$

Hence habitat which crosses a constant mutation\_cross\_limit which has range in 0-1,  $M_R$  gives a mutation rank for the habitat as per HSI value. Higher the value of HSI, higher will be the mutation rank. Hence the habitat which has higher mutation rank will have higher Mutation Probability. So habitat which has lower Mutation Probability as compared to Mutation\_Cross\_Threshold undergoes to mutation.

### Final Solution

In this work, after sufficient number of iterations, best possible habitats are obtained and set of those pages are recommended pages by the proposed model of Biogeography Optimization Algorithm.

### Proposed Algorithm

---

Input: WF // WF: Web Features  
Output: NPP // NPP: Next Predicted Page

1. [WL WC] ← Pre-Processing(WF)
2.  $SC = \frac{n(n+1)}{2}$  // n number of pages in WL
3.  $AR \leftarrow \text{FFC}(WL, SC)$
4.  $SM \leftarrow \text{Similarity\_Matrix}(WC)$
5. Loop 1: r // r: number of rules
6.  $L_p[r] \leftarrow \text{Logistic\_Regression}(AR, SM)$
7. EndLoop
8.  $T_i \leftarrow \text{User\_Visited\_Page}$  // Testing of Model
9. Initialize BOA Parameters m, h,  $\lambda$ ,  $\alpha$
10.  $H \leftarrow \text{Generate\_Habitat}(m, h, AR)$
11. Loop 1: iteration
12.  $HSI = \text{Rank}(L_p, T_i, H)$
13.  $H \leftarrow \text{Crossover}(H, \lambda, \alpha)$
14.  $H \leftarrow \text{Mutation}(H, HSI, h)$
15. EndLoop
16.  $HSI = \text{Rank}(L_p, T_i, H)$
17.  $NPP \leftarrow \text{Max}(HSI)$

---

## 4. EXPERIMENT AND RESULT

### 4.1. Data sets

In this work, authors have used real time dataset from “Project Tunnel” website [18]. The dataset contains weblog for the month of April 2019. This weblog has 20,000 sessions of 6,240 users. Number of pages in the website is 278.

### 4.2. Experiment setup

Experimental setup was developed on MATLAB software where number of inbuilt function increases the easiness of implementation. Comparison of proposed model is done on the basis of PASO algorithm proposed in [19], TermNetWP algorithm proposed in [12], WPPM (Web Page Prediction Model) model where only weblog feature is used with PAPS algorithm [13], LWPPM (Logistic Web Page Prediction Model) model with weblog and web content features using logistic regression and PASO algorithm.

Evaluation Parameters:

Precision

Precision = Approximate\_Correct\_pages / All\_predictions

Coverage

Coverage = Approximate\_Correct\_pages / All\_Visited\_Pages

M-metric

M-metric = (2 x Precision x Coverage) / (Precision + Coverage)

Execution Time: Total time required for the execution of the algorithm for predicting of the page based on the different sizes of dataset.

Mean Absolute Error (MAE) is average difference between the predicted page chromosome value to actual page chromosome fitness value.

$$MAE = \frac{\sum_j^n F\_Predict_j - F\_Real_j}{n} \quad (9)$$

In (9), n is total number of pages predicted,  $F\_Predict_j$  is the  $j^{th}$  predicted page chromosome value and  $F\_Real_j$  is the fitness value of chromosome of page actually opened by user for  $j^{th}$  prediction.

Root Mean Square Error (RMSE) highlights the superior deviations.

$$RMAE = \sqrt{\frac{\sum_j^n F\_Predict_j - F\_Real_j}{n}} \quad (10)$$

#### 4.3. Result

Table 3 shows that the proposed LWPP-BOA model has perfectly utilized weblog and web content features with logistic regression relation, hence precision value get increases. It is also observed that use of Biogeography Optimization Algorithm for page prediction gives better results as compared to TermNetWP [12] and PASO [19] models on different dataset sizes.

Table 3. Precision based comparison of page recommendation models

Testing Dataset Size Percentage	LWPP-BOA	LWPPM	WPPM (Without Content Feature) [13]	PASO [19]	TermNetWP [12]
20	0.8542	0.8333	0.6354	0.5625	0.5169
30	0.8392	0.8322	0.5385	0.5245	0.4729
40	0.7592	0.7277	0.4850	0.4469	0.4655
50	0.7185	0.6807	0.4726	0.4328	0.4318

Table 4 shows that use of BOA algorithm in LWPP-BOA reduces the execution time value, as crossover and mutation operations depends on immigration, emigration and mutation rate values. So it reduced the number of operations among habitats (chromosomes). While in Swarm Optimization, each iteration performs all operation in each set of chromosome.

Table 4. Execution time (seconds) based comparison of page recommendation models

Testing Dataset Size Percentage	LWPP-BOA	LWPPM	WPPM (Without Content Feature) [13]	PASO [19]	TermNetWP [12]
20	3.21	10.34	20.35	22.36	4.4066
30	3.56	15.63	27.68	30.52	4.203
40	4.67	21.06	33.23	37.21	5.2145
50	6.98	25.71	42.56	47.45	6.2465

From Tables 5 and 6, it is observed that proposed LWPP-BOA model has perfectly utilized weblog and web content features with logistic regression relation hence coverage value get increases. It is also observed that use of Biogeography Optimization Algorithm for web page prediction gives better results as compared to TermNetWP [12] and PASO [19] models on different dataset sizes. Thus increase in precision and coverage ultimately increases the M-metric value as well.

Table 5. Coverage based comparison of page recommendation models

Testing Dataset Size Percentage	LWPP-BOA	LWPPM	WPPM (Without Content Feature) [13]	PASO [19]	TermNetWP [12]
20	0.4316	0.4211	0.1347	0.1192	0.2421
30	0.4225	0.4190	0.1095	0.1067	0.2148
40	0.3816	0.3658	0.0858	0.0938	0.2132
50	0.3608	0.3418	0.0958	0.0809	0.2004



Table 6. M-Metric based comparison of page recommendation models

Testing Dataset Size Percentage	LWPP-BOA	LWPPM	WPPM (Without Content Feature) [13]	PASO [19]	TermNetWP [12]
20	0.5734	0.5594	0.2222	0.1967	0.3297
30	0.5621	0.5574	0.1820	0.1773	0.2954
40	0.5079	0.4869	0.1438	0.1574	0.2924
50	0.4803	0.4551	0.1615	0.1363	0.2742

Table 7 and Table 8 shows that use of BOA algorithm in LWPP-BOA reduced MAE and RMAE value, as crossover and mutation operation depends on immigration, emigration and mutation rate values. Hence more relevant pages per chromosomes are identified. While in Swarm Optimization, each iteration perform all operation in each set of chromosome so chance of getting relevant page get reduces, as elements in the crossover are depend on fitness value.

Table 7. MAE based comparison of page recommendation models

Testing Dataset Size Percentage	LWPP-BOA	LWPPM	WPPM (Without Content Feature) [13]	PASO [19]	TermNetWP [12]
20	0.1116	0.2348	0.2337	0.2884	1.2
30	0.1169	0.1864	0.2728	0.2883	1.1831
40	0.1171	0.3094	0.3005	0.2519	0.9684
50	0.2125	0.2782	0.2681	0.3836	0.7975

Table 8. RMAE based comparison of page recommendation models

Testing Dataset Size Percentage	LWPP-BOA	LWPPM	WPPM (Without Content Feature) [13]	PASO [19]	TermNetWP [12]
20	0.3341	0.4595	0.4834	0.5370	1.0954
30	0.3420	0.4317	0.5223	0.5370	1.0877
40	0.3421	0.5563	0.5482	0.5019	0.984
50	0.4610	0.5274	0.5178	0.6194	0.893

The result section has shown that proposed model has reduced the execution time as well as improved the precision of next web page prediction. In this research, authors tried to effectively utilize the weblog feature as well as web content features which resulted into reduced MAE and RMAE values. Here proposed model has overcome limitations of previous work where association rules take large time. Web content feature utilization helped to improve the regression model as previous approaches has used semantic collection of words which reduced the work efficiency.

## 5. CONCLUSION

In order to improve user familiarity of websites, number of parameters is available but each of them has its own limitation, so researcher focus on user behavior. Hence number of user features is proposed by researcher of this field which enhanced the web page prediction work. This paper has also involved a Multimodal Logistic regression method for developing a concrete feature using web content and weblog dataset. As this find a probability of the sequence of pages as per users' past series of visits in weblog and type of text contents present on pages. One more benefit of this concrete feature of logistic regression is high reduction in execution time of proposed model. Further random adoption of page prediction is maintained by Biogeographic Optimization Algorithm. Experiment is done on real time dataset of projecttunnel website. Comparison of proposed model is done with existing PASO method and it is observed that proposed model is better as it enhanced the prediction accuracy by 61.24%, reduced Mean Average Error by 53.95%. In future, page prediction algorithm may improve this work by introducing cookies data as well.

## REFERENCES

- [1] Kosala, Raymond, and Hendrik Blockeel. "Web mining research: A survey." ACM Sigkdd Explorations Newsletter 2, no. 1, pp. 1-15, 2000.
- [2] Sharma, Kavita, GulshanShrivastava, and Vikas Kumar. "Web mining: Today and tomorrow." In Electronics Computer Technology (ICECT), 2011 3rd International Conference on, vol. 1, pp. 399-403. IEEE, 2011.
- [3] Srivastava, J., R. Cooley, M. Deshpande Mukund, and P. N. Tan. "Web usage mining: discovery and application of usage patterns from web data." Proceedings of SIGKDD explorations 1, no. 2, 2002.

- [4] E.Tuba, R.Jovanovic, R.C.Hrosik, A. Alihodzic and M.Tuba, "Web Intelligence Data Clustering by Bare Bone Fireworks Algorithm Combined with K-Means". In Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics, (article 7). ACM. 2018.
- [5] Deepti Razdan, "The Next Page Access Prediction Using Markov Model", *International Journal of Electronics Communication and Computer Technology (IJECCCT)*, Volume 1, Issue 1, September 2011, ISSN: 2249-7838.
- [6] M. Awad, L. Khan, and B. Thuraisingham. "Predicting WWW surfing using multiple evidence combination," *VLDB J.*, vol. 17, no. 3, pp. 401-417, 2008.
- [7] A. Ladekar, P. Pawar, D. Raikar and J. Chaudhari, "Web Log Based Analysis of User's Browsing Behavior", *IJCSIT - International Journal of Computer Science and Information Technologies*, Vol. 6 (2), 2015.
- [8] S. Parvatikar and B. Joshi, "Analysis of User Behavior through Web Usage Mining," ICAST - International Conference on Advances in Science and Technology, 2014.
- [9] A. Deepa, and P. Raajan, "An efficient preprocessing methodology of log file for Web usage mining," NCRIMIAMI - National Conference on Research Issues in Image Analysis and Mining Intelligence, 2015.
- [10] Shumpei Okura, Yukihiro Tagami, Shingo Ono, and Akira Tajima. Embedding-based News Recommendation for Millions of Users. In SIGKDD. 2017.
- [11] Yang Song, Ali MamdouhElkahky, and Xiaodong He. Multi-rate deep learning for temporal recommendation. In SIGIR. 909-912, 2016.
- [12] Xu Chen, Yongfeng Zhang, Qingyao Ai, Hongteng Xu, Junchi Yan, and Zheng Qin. Personalized Key Frame Recommendation. In SIGIR. 2017.
- [13] ThiThanh Sang Nguyen, Hai Yan Lu, Jie Lu. "Web-page Recommendation based on Web Usage and Domain Knowledge," 1041-4347/13/\$31.00 © 2013 IEEE.
- [14] Roshan A. Gangurde and Binod Kumar. "Next Web Page Prediction using Genetic Algorithm and Feed Forward Association Rule based on Web-Log Features". *International Journal of Performability Engineering*, 16(1): 10-18, 2020.
- [15] Amit Bhagat, Sanjay Sharma, K. R. Pardasani. "Ontological Frequent Patterns Mining by potential use of Neural Network". *International Journal of Computer Applications*, Vol 36, No.10, pp. 0975-8887, 2011.
- [16] MacArthur R., Wilson E. *The Theory of Biogeography*. Princeton, NJ, USA: Princeton University Press; 1967.
- [17] Simon D. Biogeography-based optimization. *IEEE Transactions on Evolutionary Computation*. 12(6): 702-713, 2008. doi: 10.1109/tevc.2008.919004.
- [18] Dataset source: <https://projecttunnel.com/pages/Support>
- [19] R. Manikandan. "A novel approach on Particle Agent Swarm Optimization (PASO) in semantic mining for web page recommender system of multimedia data: a health care perspective". Springer Science+Business Media, LLC, part of Springer Nature 10 January 2019.

## BIOGRAPHIES OF AUTHORS



Mr. Roshan A. Gangurde is a Research Scholar at Department of Computer Science, Savitribai Phule Pune University, Pune, Maharashtra, India. His areas of interest are Data Mining, Business Analytics and Machine Learning.



Dr. Binod Kumar is Professor & Director at Jayawant Institute of Computer Applications (JSPM's Group), affiliated to Savitribai Phule Pune University, India. He has obtained PhD (Computer Sc.) in 2010, M.Phil. (Computer Sc.) in 2006, MCA (NIT Jamshedpur) in 1998 and M.Sc.(BHU) & B.Sc.(BHU) in 1995 & 1993 respectively. He is having more than 21 years of experience in various capacities in research, teaching and academic administration. He got exposure as Associate Professor at School of Engineering and Computer Technology, Quest International University, Malaysia. He has been a guide and an evaluator for MCA / M.Phil. dissertations and Ph.D thesis of students of various Universities. He is recognized Ph.D guide in Computer Science & Computer Management under Savitribai Phule Pune University. At present five Ph.D. candidates are registered under his supervision.