ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11 , Iss 8 2022

USER BEHAVIOR CLUSTERING BASED ON WEBLOG DATA USING IMPROVED BALD EAGLE OPTIMIZATION ALGORITHM

M R Ramya [1], Dr. R RajaLakshmi [2], Dr. N. Chenthalir Indra [3]

[1]Registration No: 19123152282022, Research Scholar, Department of Computer Science, Research Centre S.T Hindu College, Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli-2.

[2]Assistant Professor, Department of Computer Science, Noorul Islam College of Arts and Science, Kumaracoil, Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli-2.

[3]Assistant Professor, Department of Computer Science, S.T. Hindu College, Nagercoil, Affiliated to Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli 2.

E-mail: [1]mrramya_1984@yahoo.co.in, [2]rajalakshminatesh@gmail.com, [3]chenthalirindra@gmail.com

Abstract

Web mining is the methodology of extracting useful information and patterns from the vast data available on the World Wide Web. It involves three main types: web content mining, web structure mining, and web usage mining. Web log data, generated by web servers/clients/proxy servers, contains a record of user interactions with a website, including page views, clicks, and other activities. By analysing this data, organizations can gain insights into user navigation patterns, popular websites, and areas for improvement. The clustering of user behaviour based on educational site and non-educational site in campus network is developed using the Improved Bald Eagle Optimization (IBEO) algorithm. The traditional Bald Eagle algorithm is inspired by the foraging behaviour of the bald eagle, Haliaeetus leucocephalus that employs various hunting strategies to capture its target without allowing it to escape. However, this algorithm can sometimes become trapped in a local optimum solution. To address this, the exploration aspect of the algorithm needs to be enhanced by incorporating randomness. This enhancement is achieved using chaotic mapping of the solution, specifically through sinusoidal chaotic mapping, to improve randomness. By integrating this approach into the conventional bald eagle optimization algorithm, the IBEO (Improved Bald Eagle Optimization) algorithm aims to achieve the global best solution to cluster the User behaviour Analysis using weblog data.

Keywords: Web Mining, Web log, Clustering, Bandwidth Allocation, Network, Optimization

1. INTRODUCTION

User behaviour analysis using web log data involves studying interactions and activities on a website by analysing log files produced by the web server or clients. This process seeks to identify patterns, preferences, and trends in user engagement, offering valuable insights for many applications [1]. These applications encompass website and network optimization, e-commerce, security, content personalization, and marketing. Descriptive statistics, sequential pattern mining, machine learning, and clustering are commonly used to analyse user behaviour [2]. However, challenges such as handling massive datasets, noisy log data, real-time analysis, scalability, and privacy protection hinder effective analysis.

The IBEO algorithm initiates by defining the population size and maximum iterations for the Haliaeetus. This evolutionary method then searches the feature space to optimize solutions, which are applied to cluster user behaviour based on web log data. The algorithm pre-processes weblogs and employs sinusoidal similarity to group user sessions, assuming that sessions with more viewed pages represent higher user engagement. In the proposed user behaviour analysis, the pre-processing is employed initially and then the user behaviour clustering is employed using the Improved Bald Eagle Optimization (IBEO) algorithm. In this, the sinusoidal mapping is combined with the existing bald eagle optimization algorithm for obtaining the global best solution.

2. LITERATURE REVIEW

It was the goal of [29] to learn more about how students interacted with one another on online learning environments. The breadth of this project included gathering and analysing enormous statistics produced by user behaviour, including login durations, browsing trends, and graduation rates from assessments. The behaviour of the students was examined for patterns and trends using a variety of web mining approaches, such as association rule mining, classification, and clustering. To improve learning experiences by tailoring them to the requirements of each individual and identifying successful teaching tactics were the goals. Development of predictive models to forecast student performance, early detection of learners who were having difficulty and material delivery optimization were among the accomplishments in this discipline.

Text analysis in the field of text mining requires complex techniques for handling numerous documents [28]. Among the most effective strategies in text mining, machine learning, and pattern recognition is text clustering. Using a reasonable text-clustering method, computers can organize a corpus of documents into conceptual clusters. This technique aims to

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11 , Iss 8 2022

maximize the reliability of text clustering algorithms, and minimize the number of uninformative traits. The proposed method achieves a high convergence rate, requires minimal computational time, and avoids getting trapped in local minima within a low-dimensional space. The process begins with inputting text data and performing pre-processing steps. Next, text feature selection involves choosing local optima from the text documents and then selecting the best global optima from these local optima using a hybrid Grey Wolf Optimization (GWO) and Grasshopper Optimization Algorithm (GOA). These selected optima are then clustered using the Fuzzy c-means (FCM) clustering algorithm, which enhances reliability and reduces computational time.

The paper [30] created a conceptual model in order to close the knowledge gap between conventional learning analytics and a sophisticated comprehension of how students engage with eLearning systems. The study used a two-pronged technique to do this. It started by adopting the metadata approach and carefully gathering information on students' online behaviours, such as their use of resources and involvement in forums. Secondly, the Community of Inquiry Model was utilized, which is a framework for classifying online learning according to three factors: instructional presence, social presence, and cognitive presence.

Real Life based usage analysis [31] exposed the learning behaviours tucked therein to transform the clicks, scrolling, and page visits into a rich needlepoint of information. They used a web browser plugin; a quiet observer that recorded every digital step pupils took on the online learning platform, to accomplish this. This abundance of information, which included page stay durations, resource downloads, and quizzes attempts, served as the foundation for the approach used in the study. By use of advanced statistical analysis, they extracted patterns from the clickstream and identified three different student groups based on their level of engagement: the enthusiastic explorers, the resource-conscious navigators, and the cautious students.

Web mining techniques were employed by [32] to analyse a website belonging to an educational institution in order to determine user preferences and behaviour. Finding newly undiscovered patterns and suggesting changes was the goal. Analysing online access logs for daily, weekly, and monthly web metrics including hits, pages, and visits was one of the techniques used. Additionally examined were visitor trends, page visits, and time spent. Information on user behaviour, frequently visited locations, favoured browsers, and well-liked material was obtained as a result. The outcomes were good and provided insightful data about users' interactions with the website.

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11 , Iss 8 2022

3. DATA PRE-PROCESSING

Data pre-processing for web log data involving a student dataset encompasses multiple essential steps to read the data for analysis. Initially, relevant web log files are gathered to capture student interactions like page views and clicks. The data is then cleansed to eliminate irrelevant or redundant entries and handle any missing or corrupted information. Next, the raw data is structured by organizing logs based on timestamps, URLs, and user IDs. Sessions are identified by grouping actions according to user activity and time intervals [3]. Meaningful features such as page visit frequency and session duration are then extracted and aggregated to create a summary of interactions. If necessary, dimensionality is reduced, categorical data is encoded into numerical values, and numerical values are normalized to a consistent range. Finally, the data is validated for inconsistencies and anomalies, and relevant features are chosen based on the analysis goals.

Data cleaning is a crucial step in pre-processing web log data, and it involves several specific methods. First, handling missing values is essential; missing data can result from incomplete log entries or errors during data collection. In this methodology, missing or incomplete entries are removed to avoid skewing the analysis results. Second, removing duplicates is necessary to ensure the accuracy of the data. Duplicates may appear due to technical glitches, repeated requests, or errors, so eliminating these ensure that each user's behaviour is represented only once. Third, URL normalization addresses inconsistencies and variations in URLs. Since users might access the same content through different URLs due to factors like session IDs, query parameters, or variations in case normalizing URLs consolidates user behaviour data effectively. Together, these steps ensure that the web log data is clean, consistent, and ready for further analysis. After pre-processing the input data, the clustering of user behaviour is employed using the proposed IBEO algorithm.

4. PROPOSED METHODOLOGY

In the client-server architecture of the internet, a client makes a request, and the server responds, with protocols and messages exchanged during the session. The exponential growth of the web has led to a vast number of clients communicating with servers across millions of interconnected networks, significantly increasing latency and internet traffic. A proxy server can help reduce latency by acting as an intermediary between a web browser and a web server. It intercepts all queries to the web server to determine if it can fulfil the request itself. If user behaviour can be predicted, the proxy server can pre-emptively cache anticipated pages, further

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11 , Iss 8 2022

reducing latency. The first step in the web usage mining process is preparing the client log file. This involves transforming raw input data into a format suitable for analysis. The server log, which contains information such as IP addresses, user IDs, HTTP requests, date and time stamps, and URL requests, should be pre-processed to minimize mining time and improve the clustering of data into sessions.

User behaviour is identified using the Improved Bald Eagle Optimization (IBEO) algorithm. Inspired by the hunting tactics of the Bald Eagle (Haliaeetus leucocephalus), the traditional Bald Eagle algorithm often converges prematurely to suboptimal solutions. To address this, IBEO incorporates sinusoidal chaotic mapping to enhance exploration and increase the likelihood of finding the global optimum. The mathematical formulation of IBEO is detailed below.

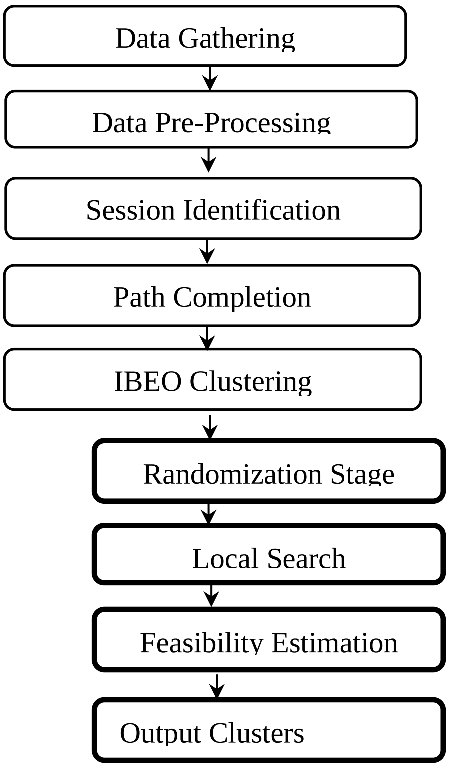

Figure 1: Workflow of the proposed IBEO based user behaviour clustering

4.1 IBEO ALGORITHM:

The IBEO algorithm starts by initializing a population of potential solutions, which are different ways to cluster user behaviour data. Its enhanced exploration capability, achieved through chaotic mapping, allows for a comprehensive search of the solution space to find effective clustering configurations. As the algorithm progresses, it focuses more on promising solutions to refine the clusters. The use of chaotic mapping introduces randomness, which

ISSN PRINT 2319 1775 Online 2320 7876
Research paper © 2012 IJFANS. All Rights Reserved, UGC CARE Listed (Group -I) Journal Volume 11 , Iss 8 2022

helps prevent the algorithm from getting stuck in local optima and ensures that it doesn't prematurely settle on suboptimal clustering configurations. Each candidate solution is evaluated using a fitness function, such as the Mean squared errors (MSE) within clusters or the silhouette score. The IBEO algorithm iteratively improves these solutions based on this evaluation. After a predetermined number of iterations or when convergence is reached, the algorithm identifies the best clustering configuration. This configuration represents the most effective grouping of users based on their web log data. The combination of chaotic mapping and randomness in IBEO enhances both exploration and exploitation, making the clustering process more effective and efficient. The algorithm is robust to noisy data, which is common in real-world web log datasets, and can handle large datasets, making it suitable for extensive user behaviour clustering. Additionally, the improved search mechanism of IBEO helps avoid local optima, bringing the clustering solution closer to the global optimum.

4.1.1 Initialization Stage:

The proposed IBEO algorithm for the user behaviour clustering begins with initialization, wherein the population size of the Haliaeetus and the maximal iteration considered are assigned initially. The Haliaeetus searches the assigned characteristic space to identify the prey and updates the solution. The solution identified by the Haliaeetus is utilized for clustering the user behaviour based on web log data. The Haliaeetus search agents are arbitrarily located in the search space and the solution evaluated is defined as:

$$W_{b,\beta} = W_a + \alpha * g(W_m - W_\beta) \tag{4.1}$$

The arbitrary value that has the boundary limits $1.5\, and\, 2$ is defined as α and the solution accomplished by the search agent in the initial stage is denoted as $W_{b,\beta}$. The average solution acquired by search agents in this phase is indicated as W_m and the best solution obtained is represented as W_a. The present iteration is denoted as β and g is the random factor that ranges between $[0,1]$.

4.1.2 Randomization Stage:

In this stage, the Haliaeetus search agents utilizes the spiral flight for identifying the prey. The Haliaeetus searches in various directions to identify the prey, wherein it utilizes the spiral flight for detection purpose. Thus, the solution updated in this phase is expressed as:

$$W_{b,\beta} = W_\beta + s(\beta) * (W_\beta - W_{\beta+1}) + e(\beta) * (W_\beta - W_m) \tag{4.2}$$

where,

$$e(\beta)=\frac{eg(\beta)}{\max(|eg|)}, \qquad s(\beta)=\frac{sg(\beta)}{\max(|sg|)} \tag{4.3}$$

$$eg(\beta)=g(\beta)*\sin(\varphi(\beta)), \qquad sg(\beta)=g(\beta)*\cos(\varphi(\beta)) \tag{4.4}$$

$$\varphi(\beta)=n*\pi*q \tag{4.5}$$

$$g(\phi)=\varphi(\beta)+Q*q \tag{4.6}$$

The factor n has the limit of $5 to 10$ that details the corner of the characteristic space and the factor Q has the range of $0.5 and 2$ that is utilized for setting the number of circles in the spiral movement. The direction of the movement is desired by the factor $s(\beta)$ and $e(\beta)$, the angular parameter is denoted as $\varphi(\beta)$, and the radial parameter is defined as $g(\beta)$. The components $eg(\beta)$ and $sg(\beta)$ are calculated based on polar coordinates, and the random number is denoted as q.

Here, the randomization capability of the further enhanced by considering the Sinusoidal chaotic mapping. It is expressed as:

$$W_{b,\beta}=A\cdot W_b^2\sin\left(\pi W_{b,\beta-1}\right) \tag{4.7}$$

where, the solution updating using the sinusoidal chaotic mapping is denoted as $W_{b,\beta}$, the control factor utilized for enhancing the randomization criteria is defined as A and the solution obtained in the previous iteration is denoted as $W_{b,\beta-1}$.

After incorporating the sinusoidal chaotic mapping, the solution evaluated by the IBEO algorithm is defined as:

$$\left(W_{b,\beta}\right)_{\mathrm{IBaE}}=0.5\left(W_{b,\beta}\right)_{Bald}+0.5\left(W_{b,\beta}\right)_{chaotic\ \sin} \tag{4.8}$$

$$W_{b,\beta}=0.5\left[W_\beta+s(\beta)*\left(W_\beta-W_{\beta+1}\right)+e(\beta)*\left(W_\beta-W_m\right)\right]+ 0.5\left[A\cdot W_b^2\sin\left(\pi W_{b,\beta-1}\right)\right] \tag{4.9}$$

Hence, the IBEO based position updating concerning the sinusoidal mapping, the randomness is enhanced and the solution trapping at local solution is eliminated.

Algorithm 4.1: Pseudo code for IBEO

Pseudo-code for IBEO Algorithm

1	Set the initial values of parameters like β, W and b are initialized
2	Position the search agents randomly within the solution space.
3	Evaluate the feasibility of each solution based on predefined criteria.
4	Assess the solutions generated during the randomization phase to determine their quality.
5	Evaluate the solutions obtained through the local search phase to refine and improve them.
6	Perform a secondary feasibility check to ensure the solutions meet all necessary criteria.
7	Determine the global best solution based on the evaluations and return it as the final result.
8	End

4.1.3 Local Search:

Based on the solution identified in the randomization stage, the capturing of target is employed in the local search stage. The Haliaeetus utilizes the swooping behaviour for capturing the target. The solution obtained by the Haliaeetus in the local search stage is formulated as:

$$W_{b,\beta} = q * W_a + e1(\beta) * (W_\beta - l1 * W_m) + s1(\beta) * (W_\beta - l2 * W_a) \quad (4.10)$$

where,

$$e1(\beta) = \frac{eg(\beta)}{\max(|eg|)}, \qquad s1(\beta) = \frac{sg(\beta)}{\max(|sg|)} \quad (4.11)$$

$$eg(\beta) = g(\beta) * \sinh(\varphi(\beta)), \qquad sg(\beta) = g(\beta) * \cosh(\varphi(\beta)) \quad (4.12)$$

$$\varphi(\tau) = n * \pi * q \quad (4.13)$$

$$g(\beta) = \varphi(\beta) \quad (4.14)$$

Here, the factors $l1, l2$ exists within the limit $[1,2]$. The updation employed through the Haliaeetus search agent is checked for its appropriateness using the fitness function.

4.1.4 Re-estimation of feasibility:

The feasibility of the solution is evaluated by calculating the cosine similarity measure and is expressed as:

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11 , Iss 8 2022

$$Fit = \frac{P \cdot Q}{\|P\| \cdot \|Q\|} \tag{4.15}$$

where, fitness function is indicated as, the two vectors required target and the acquired values are indicated as P and Q respectively. The highest value signifies the similarity between the users and lowest vale signifies the dissimilarity among the users.

4.1.5 Stoppage:

The stoppage criteria of the algorithm are devised with the attainment of required solution or the maximal count of iteration. The pseudo-code for the algorithm is presented in Algorithm 4.1. Thus, using the clustering technique, the user behaviour analysis is devised for employing the student learning management[33]. Thus, using the clustering technique, the user behaviour analysis is devised for employing the student learning management.

4.2 EXPERIMENTAL RESULT AND COMPARATIVE ANALYSIS

The proposed user behaviour clustering approach is compared with existing approaches like Particle Swarm Optimization (PSO) [28][40], Bald Eagle Optimization (BEO) [29], and Grey Wolf Optimization (GWO) [30] for illustrating the enhanced outcome of proposed IBEO. The comparison based on various scenarios like data size and number of users is illustrated in this section

4.2.1 Comparison based on Number of Users

The evaluation of the user behaviour cluster analysis based on various numbers of users with various assessment measures is elaborated in this section.

Clustering Accuracy: The clustering accuracy derived by the proposed IBEO is 0.8224, still the existing PSO, BEO and GWO is 0.7446, 0.6838, and 0.8067 respectively with 250 users. With 1000 users, the clustering accuracy estimated is 0.9305 still the existing PSO, BEO and GWO is 0.8158, 0.8096, and 0.8541 respectively. Thus, for varying the number of users, the IBEO accomplished superior outcome compared to the other conventional algorithms. Here, the incorporation of sine chaotic mapping assists the proposed model to solve the issue of local optimal trapping. The clustering accuracy for various numbers of users is portrayed in Figure 4.1 and its detailed evaluation is illustrated in Table 4.1.

Table 4.1: Clustering Accuracy based on Number of Users

Number of Users / Methods	PSO	BEO	GWO	Proposed

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11 , Iss 8 2022

250	0.7446	0.6838	0.8067	0.8224
500	0.8159	0.7771	0.8355	0.9134
750	0.7844	0.7315	0.8233	0.8932
1000	0.8158	0.8096	0.8541	0.9305

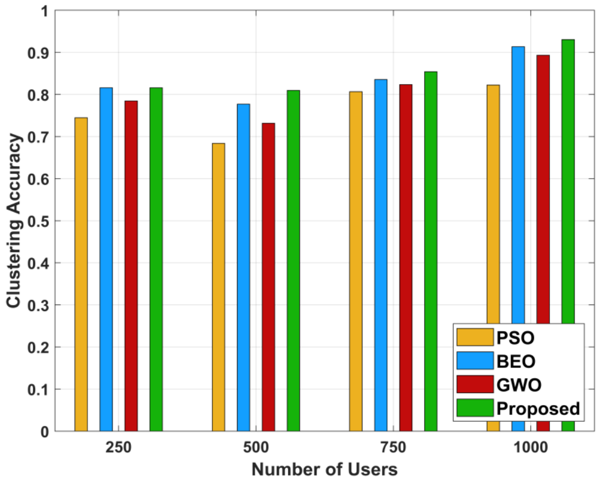

Figure 4.1: Clustering Accuracy based on Number of Users

Clustering Time: The clustering time derived by the proposed IBEO is 2.4478 ms, still the existing PSO, BEO and GWO is 3.8202 ms, 3.4654 ms, and 3.0431 ms respectively with 250 users. With 1000 users, the clustering time estimated is 3.034 ms, still the existing PSO, BEO and GWO is 4.7175 ms, 4.2705 ms, and 4.2463 ms respectively. Thus, for varying the number of users, the IBEO accomplished minimal clustering time compared to the other conventional algorithms. The clustering time for various numbers of users is portrayed in Figure 4.2 and its detailed evaluation is illustrated in Table 4.2.

Table 4.2: Clustering Time based on Number of Users

Number of Users / Methods	**Proposed**	**BEO**	**GWO**	**PSO**
250	2.4478	3.4654	3.0431	3.8202
500	3.5236	4.4381	3.5773	4.6882
750	4.1793	5.4026	4.2764	5.4721
1000	3.034	4.2705	4.2463	4.7175

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11 , Iss 8 2022

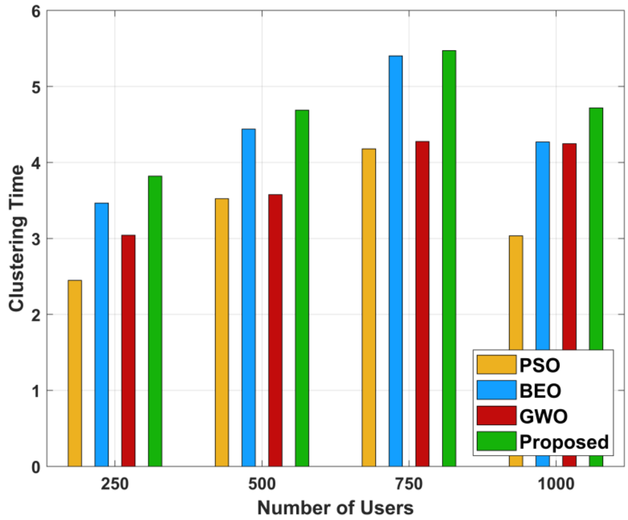

Figure 4.2: Clustering Time based on Number of Users

F1-Score: The F1-Score derived by the proposed IBEO is 0.7905, still the existing PSO, BEO and GWO is 0.6827, 0.7883, and 0.7509 respectively with 250 users. With 1000 users, the F1-Score estimated is 0.9327, still the existing PSO, BEO and GWO is 0.7394, 0.9252, and 0.8841 respectively. Thus, for varying the number of users, the IBEO accomplished better F1-Score compared to the other conventional algorithms. The F1-Score for various numbers of users is portrayed in Figure 4.3 and its detailed evaluation is illustrated in Table 4.3.

Table 4.3: F1-Score based on Number of Users

Number of Users / Methods	**PSO**	**BEO**	**GWO**	**IBEO**
250	0.6827	0.7883	0.7509	0.7905
500	0.7416	0.8481	0.7664	0.8967
750	0.7481	0.8767	0.797	0.9229
1000	0.7394	0.9252	0.8841	0.9327

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11 , Iss 8 2022

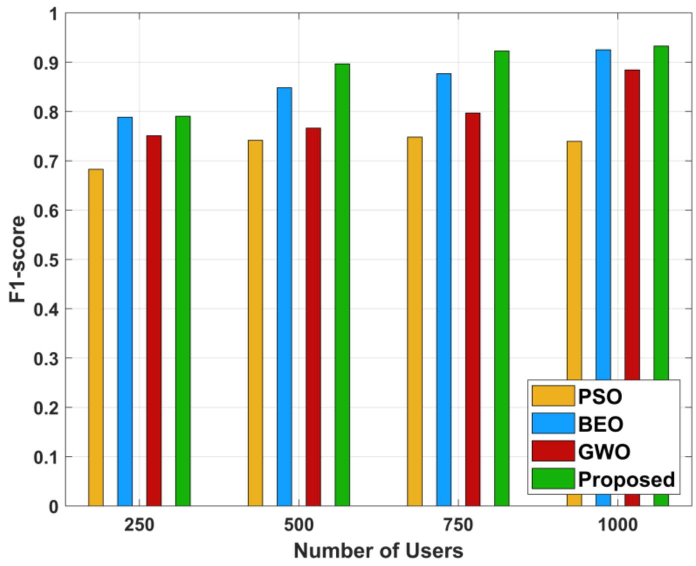

Figure 4.3: F1-Score based on Number of Users

Precision: The Precision derived by the proposed IBEO is 0.924, still the existing PSO, BEO and GWO is 0.8442, 0.9044, and 0.9033 respectively with 250 users. With 1000 users, the Precision estimated is 0.9324, still the existing PSO, BEO and GWO is 0.8684, 0.9169, and 0.9059 respectively. Thus, for varying the number of users, the IBEO accomplished better Precision value compared to the other conventional algorithms. The Precision for various numbers of users is portrayed in Figure 4.4 and its detailed evaluation is illustrated in Table 4.4.

Table 4.4: Precision based on Number of Users

Number of Users / Methods	PSO	BEO	GWO	IBEO
250	0.8442	0.9044	0.9033	0.924
500	0.7934	0.9012	0.8444	0.9088
750	0.8531	0.8922	0.8708	0.912
1000	0.8684	0.9169	0.9059	0.9324

IJFANS International Journal of Food And Nutritional Sciences Official Publication of International Association of Food and Nutrition Scientists

IJFANS INTERNATIONAL JOURNAL OF FOOD AND NUTRITIONAL SCIENCES

ISSN PRINT 2319 1775 Online 2320 7876

 UGC CARE Listed (Group -I) Journal Volume 11 , Iss 8 2022

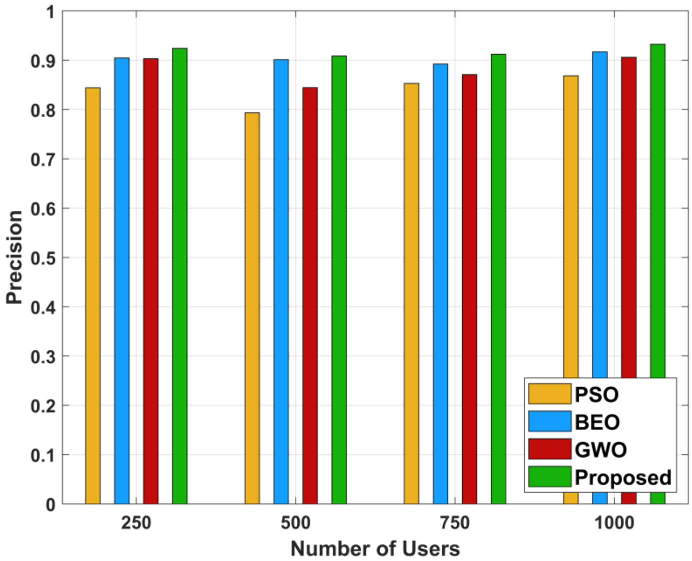

Figure 4.4: Precision based on Number of Users

Recall: The Recall derived by the proposed IBEO is 0.8845, still the existing PSO, BEO and GWO is 0.8101, 0.875, and 0.8251 respectively with 250 users. With 1000 users, the Recall estimated is 0.9343, still the existing PSO, BEO and GWO is 0.9021, 0.9307, and 0.9061 respectively. Thus, for varying the number of users, the IBEO accomplished better Recall value compared to the other conventional algorithms. The Recall for various numbers of users is portrayed in Figure 4. 5 and its detailed evaluation is illustrated in Table 4.5.

Table 4.5: Recall based on Number of Users

Number of Users / Methods	**PSO**	**BEO**	**GWO**	**IBEO**
250	0.8101	0.875	0.8251	0.8845
500	0.869	0.9122	0.8961	0.9052
750	0.8418	0.9257	0.9003	0.9129
1000	0.9021	0.9307	0.9061	0.9343

IJFANS International Journal of Food And Nutritional Sciences
Official Publication of International Association of Food and Nutrition Scientists

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11 , Iss 8 2022

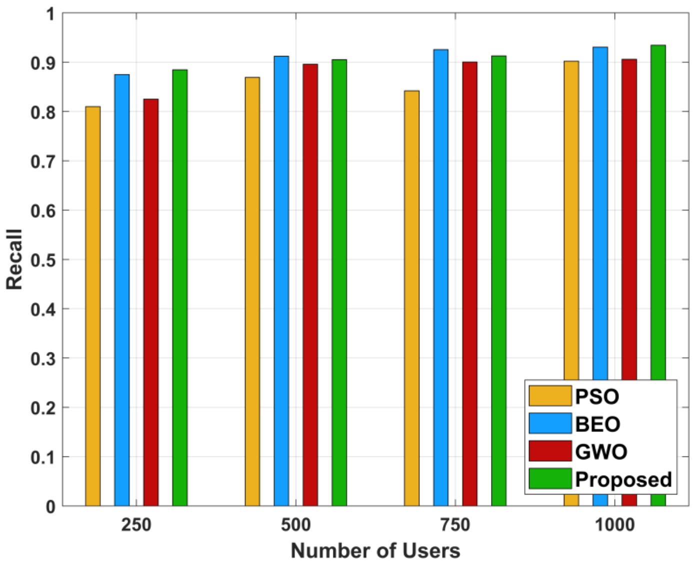

Figure 4. 5**:** Recall based on Number of Users

The Figure 4.6 illustrates the comparison of various metrics, highlighting the superiority of the IBEO algorithm.

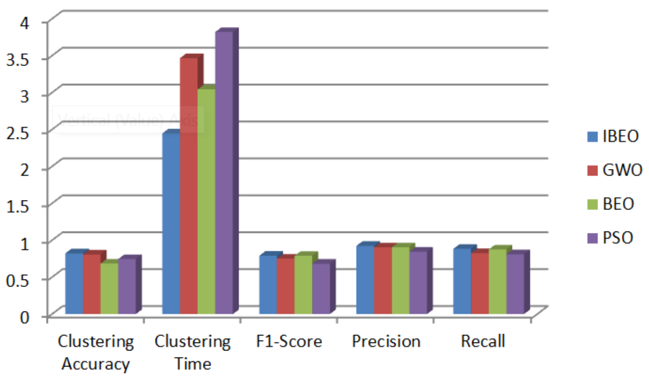

Figure 4.6: Comparison of Proposed algorithm with various measures for 250 users

4.2.2 Clustering of users based on browsing behaviours:

Clustering students based on their searching behaviour into Educational and Non-Educational site categories can provide insights into their learning patterns, interests, and potential areas of support. Educators can allocate resources more efficiently by understanding the needs and behaviours of different student clusters. Students' behaviour may change over time, requiring continuous monitoring and re-clustering. Accurately classifying sites as educational or non-educational can be challenging, especially with sites that serve multiple purposes.

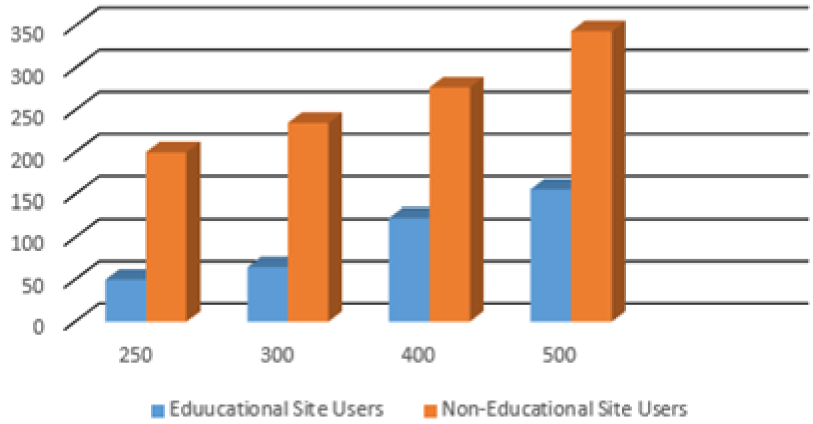

Figure 4.7: Clustering of users as Educational and Non – Educational sites.

5 CONCLUSIONS

In conclusion, applying the Improved Bald Eagle Optimization (IBEO) algorithm for clustering students' web log data has demonstrated significant improvements across several key performance metrics. By effectively categorizing web usage into educational and non-educational sites, the IBEO algorithm has achieved superior clustering accuracy, ensuring that data points are correctly grouped with a high degree of reliability. The enhanced precision and recall metrics further indicate the algorithm's ability to accurately identify and capture relevant patterns in students' online behaviour, minimizing both false positives and false negatives. Additionally, the algorithm maintains a balanced F1-Score, reflecting its proficiency in handling datasets with varying class distributions. Moreover, the clustering time is optimized, making the IBEO algorithm not only accurate but also efficient, and suitable for large-scale and real-time applications. Overall, the IBEO algorithm provides a robust and high-performing solution for analysing students' web log data, offering valuable insights into user behaviour with a level of accuracy and efficiency that surpasses existing approaches.

REFERENCES

[1] Satyam R and Deepthi D 2017 'WEB USAGE MINING FOR UNDERSTANDING USER BEHAVIOR' International Journal of Computer Engineering & Technology (IJCET) Volume 8, no. 6, pp. 12–22.

[2] Guozhang Li , Alfred R, & Wang X 2021 'Student Behavior Analysis and Research Model Based on Clustering Technology' Hindawi Mobile Information Systems Volume, Article ID 9163517, https://doi.org/10.1155/2021/9163517

[3] Romero, C., & Ventura, S. 2020, 'Educational data mining and learning analytics: An updated survey', Wiley interdisciplinary reviews: Data mining and knowledge discovery, vol.10 no. 3, e1355.

[4] Liñán, L. C., & Pérez, Á. A. J. 2015. 'Educational Data Mining and Learning Analytics: differences, similarities, and time evolution', RUSC. Universities and Knowledge Society Journal,vol. 12 no. 3, pp. 98-112.

[5] Abbas, J, Aman, J, Nurunnabi, M & Bano, S 2019, 'The impact of social media on learning behaviour for sustainable education: Evidence of students from selected universities in Pakistan', Sustainability, vol. 11, no. 6, pp. 1683.

[6] Abdelkader, HE, Gad, AG, Abohany, AA & Sorour, SE 2022, 'An efficient data mining technique for assessing satisfaction level with online learning for higher education students during the COVID-19', IEEE Access, vol. 10, pp. 6286-6303.

[7] Adnan, M, Habib, A, Ashraf, J, Mussadiq, S, Raza, AA, Abid, M & Khan, SU 2021, 'Predicting at-risk students at different percentages of course length for early intervention using machine learning models, Ieee Access, vol. 9, pp. 7519-7539.

[8] Harika, B & Sudha, PT 2020, 'Data Extraction from Web Server Logs using Web Usage Mining', International Journal of Innovative Research In Technology, vol. 6, no. 11, pp. 469-474.

[9] Arpaci, I 2016, 'Understanding and predicting students' intention to use mobile cloud storage services', Computers in Human Behaviour, vol. 58, pp. 150-157.

[10] Asadianfam, S, Kolivand, H & Asadianfam, S 2020, 'A new approach for web usage mining using case based reasoning', SN Applied Sciences, vol. 2, no. 7, pp. 1251.

[11] Bataineh, AQ, Abu-AlSondos, IA, Frangieh, RH, Salameh, AA & Alnajjar, IA 2024, 'Predictive Modeling in Marketing Analytics: A Comparative Study of Algorithms and Applications in E-Commerce Sector', Kurdish Studies, vol. 12, no. 1, pp. 499-515.

[12] Bernaschina, C, Brambilla, M, Koka, T, Mauri, A & Umuhoza, E 2017, 'Integrating modeling languages and web logs for enhanced user behaviour analytics', In Proceedings of the 26th International Conference on World Wide Web Companion, pp. 171-175.

[13] Canay, O & Kocabıcak, U 2024, 'CAWAL: A novel unified analytics framework for enterprise web applications and multi-server environments', Information Processing & Management, vol. 61, no. 3, pp. 103617.

[14] Cetin, G & Fadali, S 2021, 'Optimal resource allocation in networked control systems using viterbi algorithm', Bulletin of Electrical Engineering and Informatics, vol. 10, no. 3, pp. 1524-1535.

ISSN PRINT 2319 1775 Online 2320 7876
 UGC CARE Listed (Group -I) Journal Volume 11 , Iss 8 2022
[15] Chander, KP, Sharma, SSVN, Nagaprasad, S, Anjaneyulu, M & Ajantha Devi, V 2020, 'Analysis of Efficient Classification Algorithms in Web Mining', In Data Engineering and Communication Technology: Proceedings of 3rd ICDECT-2K19, Springer Singapore, pp. 319-332.

[16] Chatterjee, RP, Deb, K, Banerjee, S, Das, A & Bag, R 2019, 'Web Mining Using k-Means Clustering and Latest Substring Association Rule for E-Commerce', J. Mech. Contin. Math. Sci., vol. 14, pp. 28-44.

[17] Chayanukro, S, Mahmuddin, M & Husni, H 2021, 'Understanding and assembling user behaviours using features of Moodle data for eLearning usage from performance of course-student weblog', In Journal of Physics: Conference Series, IOP Publishing, vol. 1869, no. 1, pp. 012087.

[18] Chen, F, Jing, C, Zhang, H & Lv, X 2022, 'Wifi Log-Based Student Behaviour Analysis and Visualization System', The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, vol. 43, pp. 493-499.

[19] Chitraa, V & Thanamani, AS 'Web Log Data Analysis by Enhanced Fuzzy C Means Clustering', International Journal on Computational Sciences & Applications (IJCSA), vol. 4, no. 2.

[20] Choudhary, L & Swami, S 2023, 'Exploring the Landscape of Web Data Mining: An In-depth Research Analysis', Current Journal of Applied Science and Technology, vol. 42, no. 24, pp. 32-42.

[21] Chu, X, Ilyas, IF, Krishnan, S & Wang, J 2016, 'Data cleaning: Overview and emerging challenges', In Proceedings of the 2016 international conference on management of data, pp. 2201-22**06**.

[22] Cui, J, Liu, Y & Nallanathan, A 2019, 'Multi-agent reinforcement learning-based resource allocation for UAV networks', IEEE Transactions on Wireless Communications, vol. 19, no. 2, pp. 729-743.

[23] Dash, S, Luhach, AK, Chilamkurti, N, Baek, S & Nam, Y 2019, 'A Neuro-fuzzy approach for user behaviour classification and prediction', Journal of Cloud Computing, vol. 8, no. 1, pp. 1-15.

[24] Delgado, S, Moran, F, Jose, JCS & Burgos, D 2021, 'Analysis of students' behaviour through user clustering in online learning settings, based on Self Organizing Maps neural networks', IEEE ACCESS Digital Object Identifier, vol. 99.

[25] Demuyakor, J 2020, 'Coronavirus (COVID-19) and online learning in higher institutions of education: A survey of the perceptions of Ghanaian international students in China', Online Journal of Communication and Media Technologies, vol. 10, no. 3, p. e202018.

[26] Dhanaraj, RK, Ramakrishnan, V, Poongodi, M, Krishnasamy, L, Hamdi, M, Kotecha, K & Vijayakumar, V 2021, 'Random forest bagging and x-means clustered antipattern detection from sql query log for accessing secure mobile data', Wireless Communications and Mobile Computing, vol. 2021, pp. 1-9.

[27] Wang, D., Tan, D., & Liu, L. 2018. Particle swarm optimization algorithm: an overview. Soft computing, vol 22,pp. 387-408.

[28] R. Purushothaman, S. P. Rajagopalan, Gopinath Dhandapani. 2020. Hybridizing Gray Wolf Optimization (GWO) with Grasshopper Optimization Algorithm (GOA) for text feature selection and Clustering, https://doi.org/10.1016/j.asoc.2020.106651

[29] Psaromiligkos, Y., Orfanidou, M., Kytagias, C., & Zafiri, E. 2011. Mining log data for the analysis of learners' behaviour in web-based learning management systems. Operational Research, vol. 11, pp. 187-200.

[30] Man, M., Azhan, M. H. N., & Hamzah, W. M. A. F. W. 2019. Conceptual model for profiling student behavior experience in e-learning. International Journal of Emerging Technologies in Learning, vol. 14 no. 21 pp. 189 - 197.

[31] Sheard, J., Ceddia, J., Hurst, J., & Tuovinen, J. 2003. Inferring student learning behaviour from website interactions: A usage analysis. Education and Information technologies, vol. 8, pp. 245-266.

[32] Sharma, A. K., & Gupta, P. C. 2013. Predicting the Behaviour and Interest of the Website Users through Web Log Analysis. International Journal of Computer Applications, vol. 64 no.7. pp 398- 401.

[33] Dixit, A & Sharma, S 2023, 'Web server access log pattern analysis using GWO-based clustering. Journal of Integrated Science and Technology, vol. 11, no. 2, pp. 479-479.

[34] Doherty, C 2022, 'An investigation into the relationship between multimedia lecture design and learners' engagement behaviours using web log analysis', Plos one, vol. 17, no. 8, pp. e0273007.

[35] Doherty, C 2023, 'Using web log analysis to evaluate healthcare students' engagement behaviours with multimedia lectures on YouTube. Plos one, vol. 18, no. 4, pp. e0284133.

[36] Ehikioya, SA & Zeng, J 2021, 'Mining web content usage patterns of electronic commerce transactions for enhanced customer services', Engineering Reports, vol. 3, no. 11, pp. e12411.

ISSN PRINT 2319 1775 Online 2320 7876

Research paper UGC CARE Listed (Group -I) Journal Volume 11 , Iss 8 2022

[37] Etminani, K, Akbarzadeh-Totonchi, MR & Yanehsari, NR 2009, 'Web Usage Mining: users' navigational patterns extraction from web logs using ant-based clustering method', In IFSA/EUSFLAT Conf., pp. 396-401.

[38] Flanagan, B, Majumdar, R & Ogata, H 2022, 'Early-warning prediction of student performance and engagement in open book assessment by reading behaviour analysis', International Journal of Educational Technology in Higher Education, vol. 19, no. 1, pp. 1-23.