International Journal on Emerging Technologies **11**(1): 161-173(2020)

ISSN No. (Print): 0975-8364
ISSN No. (Online): 2249-3255

Optimal Association Rule Mining for Web Page Prediction using Hybrid Heuristic Trained Neural Network

Roshan A. Gangurde[1] and Binod Kumar[2]
[1]Research Scholar, Department of Computer Science,
Savitribai Phule Pune University, Pune (Maharashtra), India.
[2]Professor & Director, Jayawant Institute of Computer Applications, Pune (Maharashtra), India.

(Corresponding author: Roshan A. Gangurde)
(Received 12 October 2019, Revised 14 December 2019, Accepted 16 December 2019)
(Published by Research Trend, Website: www.researchtrend.net)

ABSTRACT: Today, data mining, which is a branch of web mining acts a fundamental role in diverse applications like health care data extraction, education system, search engines for evaluating their performance rank over other systems. Web Page prediction (WPP) is a classification issue in which the prediction of web pages is accomplished that a user may visit according to the knowledge of the formerly visited pages. WPP problem can be extended and implemented to reduce the access time while surfing the websites. The need to anticipate the needs of the website users to improve accessibility and user engagement is more than apparent now a day. Association rule mining is one of the most significant fields in data mining and knowledge discovery in databases. This paper plans to implement a new web page prediction model using an improved machine learning algorithm. The proposed web page prediction involves three phases (a) Rule Mining, (b) Optimal Rule Selection, and (c) Prediction. Initially, the collected web data is subjected to rule mining process. It is performed using the renowned association rule mining called Apriori algorithm, which is adopted for mining the frequent item set and association rule learning over relational databases. The length of the rule extracted from the Apriori algorithm is long, and it is needed to be reduced for performing the prediction with unique informative rules. Hence, the optimal rule selection is adopted, which uses the hybrid optimization algorithm with the integration of Deer Hunting Optimization Algorithm (DHOA) and Chicken Swarm Optimization (CSO) called Deer Hunting Rooster-based CSO (DR-CSO). Further, the optimally selected rules are subjected to the Machine learning algorithm named Neural Network (NN) for predicting the browsing behavior of the user. Along with the optimal rule extraction, the proposed DR-CSO is used for performing the training in NN. The experimental and comparative results will prove the efficiency of the developed model over existing algorithms.

Keywords: Apriori Algorithm; Association Rule Mining; Deer Hunting Rooster-based Chicken Swarm Optimization; Neural Network; Optimal Rule Extraction; Web Log Data; Web Page Prediction.

Nomenclature: DHOA, Deer Hunting Optimization Algorithm; CSO, Chicken Swarm Optimization; DR-CSO, Deer Hunting Rooster-based CSO; NN, Neural Network; WPP, Web Prediction Problem; QoS, Quality of Service; MASE, Mean Absolute Scaled Error; SLR, Systematic Literature Review; EC, Example Classifier; SMAPE, Symmetric Mean Absolute Percentage Error; TLPM, Two-Level Prediction Model; MAE, Mean Absolute Error; SLC, Single Label Classification; DCNN, Deep Convolutional Neural Network; RMSE, Root Mean Square Error; WWW, World Wide Web; MSE, Mean Square Error; MEP, Mean Percentage Error.

I. INTRODUCTION

Nowadays, web prediction is a general problem where the researcher tries to predict the succeeding web pages, which the user might access on the basis of having knowledge of the earlier visited pages [1]. Moreover, this kind of information related to the user's history navigation behaviour in certain period is called a 'session'. The sessions that offer the source of information to train are taken from the weblogs of web servers and those have the series of pages that the users have accessed with the accessed date and time [2, 3]. The WPP is comprehensive and implemented in several industries like caching, recommendation systems, wireless applications, and search engines [4]. Thus, it is very significant for observing practical and scalable solutions, which enhances the training as well as prediction procedures. The enhancement of prediction procedure might decrease the user's visiting times during browsing, and it won't trouble the network traffic by shunning accessing inessential pages [5, 6].

Web mining approach has already been introduced and modernized to fulfil the needs of web users. The theory of web mining has utilized the approaches of data mining on internet, web services, and web sites for reference, clustering algorithm, association rule, and sequential pattern evaluation [7]. Web mining technique is to take out the information and pattern automatically [8]. The information such as consumption, transaction information, and browsing information of users are preserved on website, thus that type of information is completely suitable to implement data mining approaches for mining source, enhancing the QoS of web site, web log file, improving customer satisfaction, and reducing the user's latency [9, 10]. The objective of web usage mining is to identify the helpful data and improve the usage from the web log file for web applications like prediction, caching, pre-fetching, and personalization are the effective procedures to improve the browsing speed of the user, reducing the latency of the user and web server loading [11].

Many complicated approaches are introduced for improving the outcomes of the clustering techniques [12-14]. In recent times, many methods are developed, which uses web access series along with the information taken by the web page contents. With the help of N-grams from the web pages, efforts have been made for merging web access log files by web content [15]. Moreover, pre-processing and prediction are the significant challenges that involve controlling of more data, which is not suitable to the memory of the computer, selecting size of optimum sliding window, session recognition, and domain knowledge extraction. In addition, the challenges of prediction consisting of extensive training time, less prediction accuracy, and restricted memory [16]. Thus, the previously specified challenges are helpful for upcoming researches to predict web page effectively.

The fundamental contributions of the current research paper on web page prediction are given below.

– To acquire the standard benchmark datasets from different links for carrying out the optimal web page prediction

– To perform the association rule mining using a standard frequent item set mining

– To accomplish the optimal rule extraction and optimized machine learning algorithm with the assistance of a newly implemented meta-heuristic algorithm

– To develop a hybrid meta-heuristic algorithm with the integration of two renowned algorithm that can improve the prediction rate by optimal rule extraction and optimized machine learning algorithm

– To confirm and validate the performance of the proposed web page prediction model over the other existing approaches in terms of diverse error measures.

The organization of the paper is done according to the steps mentioned below: Literature review and features and challenges of existing web page prediction models are given in Section II. Section III describes the proposed web page prediction model: rule mining dependent framework. Moreover, the optimal rule extraction with optimized neural network-based web page prediction is shown in Section IV. The results and discussions are described in Section V. Lastly, the conclusion of the overall paper is shown in Section VI.

II. LITERATURE REVIEW

A. Related Works

Rekik *et al*., (2018) have suggested a SLR initially for recognizing the usage of earlier researches from the evaluation and established the influenced categories. Later, a procedure after gathering and extracting information from a list of researches has been introduced. An approach that was based on Apriori algorithm was allotted and applied for identifying the rules among the category and criteria of website. Moreover, the authors have presented a re-evaluation of soft computing evaluation techniques, which aimed to assist the research group for having a possibility in conventional researches and for upcoming developments. Finally, the acquired outputs have induced for future association rule mining and probe datasets [17].

Awad and Khalil (2012) has suggested a modified Markov method for reducing the problem of scalability in count of paths. Moreover, a new two-tier prediction model was proposed, which generated an EC on the basis of training examples and the adopted classifiers. The framework will enhance the prediction time by not compromising prediction accuracy. In order to evaluate, compare, and exhibit, a standard dataset was used to verify the efficiency of proposed approaches with differences of Markov approaches and association rule mining. The tests have shown that the improved Markov approach was efficient in reducing the count of paths with high accuracy [18].

Dimopoulos *et al.,* (2010) have focussed on solving the problem pertain to web page prediction by observing the user's access point's history and other data related to web page. The suggested technique has a benefit of having less computational complexity, and utilized a less amount of additional memory. The analysis was done on different web pages and website log files. The results have demonstrated that the suggested approach has high-quality performance, which was observed in several conditions [19].

Madjarov *et al.,* (2019) have utilized 10 various descriptions of web pages for carrying out the web page prediction. In order to assess the effect of various information sources effectively, predictive clustering trees as well as ensembles were employed. Two standard corpora such as 20-genre and SANTINIS-ML datasets were used for experimentation. Finally, the outcomes have shown that the usage of web genres produced the best predictive performance over two datasets, whole feature sets, hierarchies, and predictive models. Moreover, ensembles provided existing predictive performance and they were superior to single tree models [20].

Lee *et al.,* (2011) have suggested an effective prediction approach, TLPM, using a characteristic of original hierarchical property from web log information. TLPM have reduced the dimension of user group of web pages and improve the speediness of prediction by ample accuracy. The results have proven that the TLPM was improved the prediction performance while the count of web pages was growing [21].

Lopez-Sanchez *et al.,* (2019) have suggested a novel approach to categorize the web pages based on visual content. The overall model was attained by investigating the joint implementation of a transfer learning mechanism and metric learning approaches for constructing a DCNN for feature extraction while training data was also limited. Therefore, the acquired results have shown that the suggested approach was superior to existing handcrafted image descriptors and attained more classification accuracy [22].

Li *et al.,* (2017) have offered an optimized classification approach for performing the enormous web page categorization by semantic networks. Here, Wikipedia dataset was utilized and initialized some category of entity words as class words. In order to compute the class weight of all Weighted Entity Words, a weight assessment model based on breadth and depth of Wikipedia network was employed. In this model, the keywords in the web page were taken out from the main text as well as from the title. Bayesian classifier was utilized for assessing the page class probability. Finally, the results have shown that the suggested approach acquired excellent strength, reliability, and scalability for massive web pages [23].

Wu *et al*., (2019) have suggested a novel approach for Public Intelligence mining trends and user prediction was built on the basis of fractional order differential. Moreover, this approach was recognized with an optimization approach.

The parameters were given the analytical solution of the approach and the predicted outcome equation was employed as a new technique to public intelligence mining prediction, and Internet user intention. The outcomes have proven that the suggested method was able to recognize the parameters in a precise manner [24].

B. Review

Even though there are many web page prediction approaches for forecasting the future web pages used by the user, still there are many obstacles to fulfill those challenges, and new methodologies need to be introduced in an efficient manner. By using Naive Bayes classifier, it is practically complicated to classify a web page. Some of the features and challenges of conventional web page prediction models are described in Table 1. Among them, Apriori algorithm [17] is implemented to create robust association rules, and it is the most powerful association rule miner. But, there are few disadvantages such as it need more database scans, and complex to identify product groupings. Markov model [18] controls the excess memory requirements when using huge datasets, and reduces the number of paths with high accuracy. Yet, it has more number of unstructured parameters. Hierarchical clustering [19] has high performance, and it is easy to implement. Though, there are few drawbacks such as it has cubic time complexity in several cases so it is slower. Predictive clustering trees [20] are used to predict structured outputs consisting of SLC etc., and it uses randomized version of selection of attributes. Still, there are few conflicts like it is not scalable, and it doesn't work well with noise and outlier. TLPM [21] improves the prediction performance when the count of web pages is rising, and it can diminish the size of candidate set of web pages and improves the speediness of prediction with satisfactory accuracy. However, it needs to improve the efficiency and accuracy. Metric learning approaches utilized the final layers of DCNN [22] for making it more appropriate for distance oriented classification, and it consistently improved the accuracy. Though, it produces worst for training the features directly in the classifier. Bayesian classifier [23] is utilized to evaluate the page class probability, and it consists of specifying a prior and integration. Although, there are some disadvantages like it is not feasible. Fractional differential method [24] is more reliable with original data, and has high performance. But, the error seems to be increased. Hence, the above mentioned obstacles are highly induced for developing the further developments effectively.

Table 1: Features and challenges of existing web page prediction models.

Author [citation]	Methodology	Benefits	Drawbacks
Rekik *et al.* [17]	Apriori algorithm	– It is implemented to create robust association rules. – It is the most powerful association rule miner.	– Need more database scans. – Complex to identify product groupings.
Awad and Khalil [18]	Markov model	– Reduces number of paths with high accuracy. – It controls the excess memory requirements when using huge datasets.	– Has more number of unstructured parameters.
Dimopoulos *et al.,* [19]	Hierarchical Clustering	– Has high performance. – It is easy to implement.	Has cubic time complexity in several cases so it is slower.
Madjarov *et al.,* [20]	Predictive clustering trees	– It is used to predict structured outputs consisting of SLC etc. – It uses randomized version of selection of attributes.	– It is not scalable. – It won't well in the presence of noise and outlier.
Lee *et al.,* [21]	TLPM	– It improves the prediction performance when the count of web pages is rising. – Reduced speediness and reasonable accuracy	– Need to improve efficiency and accuracy.
Lopez-Sanchez *et al.,* [22]	DCNN	– More suitable for distance oriented classification – It consistently improved the accuracy.	– It might produce poor performance when trained the features directly.
Li *et al.*, [23]	Bayesian classifier	– It is utilized to evaluate the page class probability. – It consists of specifying a prior and integration.	– It is not feasible.
Wu *et al.*, [24]	Fractional differential method	– It is more reliable with original data. – Has high performance.	– The error is increased by one to two orders of magnitude.

III. PROPOSED WEB PAGE PREDICTION MODEL: RULE MINING DEPENDENT FRAMEWORK

A. Architectural Model

With the rapid growth of web applications and services, the data used in the web page has been overloaded and hence, controlling the dynamic web page information is a challenging task in current days. Nowadays, the extraction of such interesting and search-related data is very significant so that web mining concept has attracted more. Moreover, web mining focuses on pattern evaluation from the WWW. As there is an improvement in the technology, a common person using the web connectivity results in increment of information on WWW. The web page prediction is done based on the user behaviours, as which page the user is going to visit in the next session. The architecture of proposed web page prediction model is depicted in Fig. 1.

In the proposed model, web page prediction is done by four steps such as (a) Data Acquisition, (b) Rule Extraction, (c) Optimal Rule Extraction, and (d) Prediction. In the data acquisition phase, the web log data is collected from the three benchmark sets, like FIFA, KOSARAK, and MSNBC_SPMF. The next phase is rule extraction, in which the rules are extracted using Apriori algorithm from the datasets. "Apriori algorithm is an influential model for mining regular data for association rules".

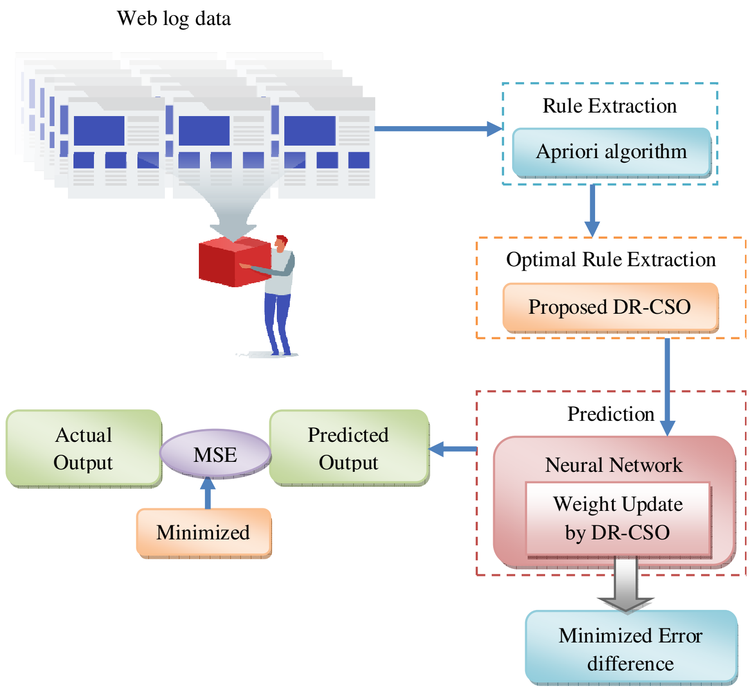


Fig. 1. Proposed architecture of web page prediction model.

Moreover, it is employed for decreasing the set of rules by determining the frequent rules from the database. Since the length of the extracted rules is long, optimal rule extraction is employed as the next step. In optimal rule extraction, the best set of rules is extracted from the generated rules in the rule extraction phase. The optimal rules extraction is done by the proposed DR-CSO model, which is the hybridized form of DHOA and CSO algorithms. Further, the optimal set of rules is subjected to the NN classifier for predicting the output. Here, the two objectives of the proposed model are obtaining the minimum MSE between the predicted and actual outcome for optimal rule extraction, and the minimum error difference between the predicted and the actual outcome for optimal weight update by DR-CSO algorithm. In addition, the training of the NN is done with the normalized data, which is between 0 and 1. Thus the final predicted value provides the browsing behaviours of different users from the datasets.

B. Rule Extraction

In order to extract the rules from the three datasets, Apriori algorithm [25] is employed. It is one of the traditional models considered for functioning on databases, which consists of transactions. In association rule mining, Apriori algorithm is the one of the classic and famous models for deriving rules and patterns from the datasets. Moreover, it is a decisive algorithm, which employs generation of users to determine frequent item sets. One of the advantages of Apriori algorithm is its simple implementation. Some of the general terms used in Apriori algorithm are described below.

Item set: It is a group of items present in the datasets. Assume the set of items as $Is_1, Is_2, \cdots Is_n$. When there are *m* items in the database, there will be 2^m possible set.

Transaction: It is the entry of the database, which includes the complete information regarding all the items and it is indicated by Tr. The subset of Tr is denoted as $Tr = \{Is_1, Is_2, \cdots Is_n\}$.

Database: It is nothing but a group of transactions, which is indicated by $Dt = \{Tr_1, Tr_2, \cdots, Tr_n\}$.

Support: It computes the transactions that have item sets, which calculates both sides of consequences in association rules, and it is denoted as Su. The corresponding equation is denoted in Eqn. (1), which agrees with statistical implication. When the support is small, the rule is not considered or it is preferred less.

Support ($C \rightarrow D$) = Number of transaction including both *C* and *D* (1)

Minimum Support: "It is the minimum threshold that must be fulfilled by an item to be the regular item in the dataset", which is represented as *Mns*.

Confidence: It measures the strength of the rules, and it is indicated by Cf, which is calculated using Eqn. (2).

Confidence ($C \rightarrow D$) =

$$\frac{\text{Number of transaction including both } C \text{ and } D}{\text{Transaction } C} \quad (2)$$

Frequent Item set: The item set that fulfils the conditions of being equal or more than the minimum support are termed as frequent item set, which is represented by Fr_{Is}, where I_S is the item set. It is not frequent item set, if item set doesn't meet the conditions.

Candidate Item set: The items that are to be employed to process are named candidate item set, and it is denoted as Ca_{Is}. From item set of length *n*, Apriori algorithm provides candidate item set of length $n+1$.

Association rule: In general terms, it is utilized to determine an association among two items or the items that are used to associate two items is called association rule. Assume $C \leq I_S$, and $D \leq I_S$, and $C \cap D = \phi$, the rule might be $C \rightarrow D$.

Apriori algorithm determines frequent item set from web browsing database and further obtains association rule from those sets. In order to generate frequent item set, candidate generation phase is utilized. The item set with high or same frequency to the minimum support threshold is considered when the frequent item set is determined. If the confidence is same or more to minimum confidence, the generation of association rules is considered. In Apriori algorithm, the candidate generation follows two stages such as join and prune steps.

The union of two frequent item sets of size n in C_n and D_n, which has the initial $n-1$ elements in similar are considered in join step, which is shown in Eqn. (3).

$$Jo_{n+1} = C_n \cup D_n \tag{3}$$

All the item sets of size n in Jo_{n+1} are verified either the item sets are frequent or not that pass the requirement generation Ca_{n+1} is done in prune step. In Apriori algorithm, join step follows prune step. Association rules are determined when the minimum transaction support and minimum confidence conditions are fulfilled. The performance of Apriori improves if the size of the candidate is decreased.

(a) Group of frequent 1-item set Ca_1 is found.

(b) By scanning the entire database, the minimum threshold support Mns is measured.

(c) By comparing Ca_1 and Mns, frequent item is found, thus generating Fr_{Is}.

(d) Later, candidate generation step takes place again for generating Ca_2.

(e) In order to generate candidate item set, repeat the steps from (b)-(d).

(f) Repeat all the above steps until no frequent item set or candidate set is generated.

The pseudo code of Apriori algorithm is shown in Algorithm 1.

Algorithm 1: Pseudo code of Apriori algorithm [25]

$Ca_1 = \{Candidate\,itemset - 1\}$;
$Fr_1 = \{Frequent\,itemset, cf \in Ca_1 | cf \cdot count \geq Mns\}$;
for $\{K = 1; Fr_K = !; K++\}$ do begin
  $Ca_{K+1} = apriori - gen(Fr_K)$;
  For all transactions $Tr \in database$ do begin
    $Ca_{Tr} = subset(Ca_{K+1} + 1)$;
    For all candidates $cf \in Ca_{Tr}$ do
      $cf \cdot count++$;
    End
    $Fr_{K+1} = \{Ca \in Ca_K | c \cdot count > +Mns\}$;
  End
End
Answer $U_K Fr_K$;

Thus after applying Apriori algorithm, rules are extracted from the web database, and the extracted rules are indicated as R_i, where $i = 1, 2, \cdots, N_R$, and the total number of extracted rules are denoted as N_R.

C. Objective Model and Solution Encoding

The objective model of the developed web page prediction using DR-CSO is focussed in two sections like optimal rule extraction and optimal trained NN-based prediction.

The first objective is to select the optimal rules using the proposed DR-CSO in such a way that the MSE between the predicted and actual outcome is minimum. The mathematical equation of the first objective function is given by Eqn. (4), and MSE is given by Eqn. (5), where the actual value is denoted as V_l and the predicted value is denoted as $\hat{V}_l$. Moreover, the term n defines number of samples.

$$FR1 = \underset{\{R_1, R_2, \cdots, R_{N_R}\}}{\arg Min} (MSE) \tag{4}$$

$$MSE = \frac{1}{n} \sum_{l=1}^{n} \left(V_l - \hat{V}_l\right) \tag{5}$$

Accordingly, the solution encoding for optimal rule extraction is shown in Fig. 2, which is done using the proposed DR-CSO. Here, the maximum and minimum bounding limit of the solution is $(1, N_R)$, and the number of optimally selected rules is fixed as 100.

R_1	R_3	R_6	R_2	……	R_{N_R}

Fig. 2. Solution encoding for optimal rule extraction.

The second objective is to optimize the weight of NN for performing the better training. The optimal training is done in such a manner that the error difference between the actual and the predicted outcome should be minimum. Accordingly, the second objective is shown in Eqn. (6), and the formulation showing the error computation in NN is shown in Eqn. (7).

$$FR2 = \underset{\{A_{wu}^{NN}\}}{\arg Min} (Error) \tag{6}$$

$$Error = \sum_{h=1}^{L(o)} \left| L_h - \hat{L}_h \right| \tag{7}$$

In Eqn. (7), $\hat{L}_h$ is the predicted outcome, and L_h is the actual outcome of NN. Fig. 3 shows the solution encoding for optimal training in NN, where N_A denotes the number of weights from total layers of NN.

A_1^{NN}	A_2^{NN}	……	$A_{N_A}^{NN}$

Fig. 3. Solution Encoding for weight update in NN.

IV. OPTIMAL RULE EXTRACTION WITH OPTIMIZED NEURAL NETWORK-BASED WEB PAGE PREDICTION

A. Optimal Rule Extraction

The extracted rule from the Apriori algorithm is denoted as R_i, which is subjected to the new hybrid optimization algorithm DR-CSO for obtaining the optimal rules. Here, the optimally extracted rules are denoted as R_i^*, where $i = 1, 2, \cdots, N_R^*$, and the total number of optimally extracted rules is denoted as N_R^*.

B. Neural Network

In most of the applications, NN [26] is utilized because it is compatible over other classifiers. Moreover, its structure includes output, input and hidden layers. In order to compute the result of the whole network, the hidden layer's output is essential.

The output of hidden layer and output layer is expressed in Eqns. (8) and (9), respectively. In Eqn. (8), the activation function is given as acv, the bias weight to the hidden neuron is denoted as $\tilde{A}^{(M)}_{(\hat{G}e)}$. The count of input neurons is represented as $In(o)$. The input and hidden neurons are denoted as g and e, respectively. The weight from the input neuron to the hidden neuron is given by $\tilde{A}^{(M)}_{(ge)}$, and the input features are denoted as R_i^*.

$$\overline{M}^{(M)} = acv\left(\tilde{A}^{(M)}_{(\hat{G}e)} + \sum_{g=1}^{In(o)} \tilde{A}^{(M)}_{(ge)} R_i^*\right) \tag{8}$$

$$\hat{L}_h = acv\left(\tilde{A}^{(L)}_{(\hat{G}h)} + \sum_{e=1}^{HD} \tilde{A}^{(L)}_{(eh)} \overline{M}^{(M)}\right) \tag{9}$$

In Eqn. (9), the bias weight to the output neuron is indicated by $\tilde{A}^{(L)}_{(\hat{G}h)}$, the number of hidden neurons is denoted as HD. The weight from the hidden neuron to the output neuron is represented as $\tilde{A}^{(L)}_{(eh)}$. The weight function $A^{NN}_{wu} = \left\{\tilde{A}^{(M)}_{(\hat{G}e)}, \tilde{A}^{(L)}_{(\hat{G}h)}, \tilde{A}^{(M)}_{(ge)}, \tilde{A}^{(L)}_{(eh)}\right\}$ is selected optimally to provide better training for NN with minimum error difference as shown in Eqn. (6). The error difference should be minimized by weight update using proposed DR-CSO algorithm.

C. Conventional CSO

The inspiration of CSO [27] is relying on the chicken swarm's behaviour. It includes roosters, chicks, and hens. The chicken's behaviour is idealized by some set of rules. There are multiple clusters; in each cluster there exists a "rooster, couple of hens, and chicks", in which the roosters are dominant. In order to specify the individuality of chickens, the chicken swarm is divided into multiple clusters that are based on the chicken's fitness values. The one who is having the best fitness value is called rooster, and the one with worst fitness is called chicks. The remaining chickens are termed as hens. These hens choose their own group. In general, the mother-child relationship has been made among hens and chicks. The relationships of dominant, mother-child are not changed. After several time steps Tis, the status of each is updated. Moreover, chickens follow the rooster existing in the cluster for searching the food because they are prevented to eat their own food. The chicks search their food near to their mother.

Let the roosters, hens and chicks are termed as *RT, HN*, and *CS*, respectively. The mother hens are denoted as *MHN*. All the B virtual chickens are denoted by the locations $PH^{Tis}_{a,b}(a \in [1,\cdots,B], b \in [1,\cdots,Dms])$ searching for food in dimensional space Dms. The roosters with extensive fitness values have the benefit for searching the food in large areas. The corresponding equation is given in Eqns. (10) and (11).

$$PH^{Tis+1}_{a,b} = C^{Tis}_{a,b} * (1 + Rad(0, \sigma^2)) \tag{10}$$

$$\sigma^2 = \begin{cases} 1, \text{if } ftn_a \le ftn_c, \\ \exp\left(\dfrac{(ftn_c - ftn_a)}{|ftn_a| + \varepsilon}\right), \text{ otherwise}, c \in [1, B], c \ne a \end{cases} \tag{11}$$

From the above equations, the Gaussian distribution with mean 0 and standard deviation σ^2 is denoted as $Rad(0, \sigma^2)$. In order to elude zero division error, ε is considered as it is smallest constant. Moreover, the index of the rooster is given by a, which is selected randomly from the cluster and the fitness value of PH is denoted as ftn. The numerical equation of the hens will follow the rooster, and the chicks follow the mother hens for searching the food as represented in Eqns. (12), (13) and (14). Here, the random number lies in between 0 and 1. The terms $rd1 \in [1,\cdots,B]$ and $rd2 \in [1,\cdots,B]$ are selected from a^{th} hen's cluster and the other is selected randomly from the swarm, respectively. In addition, the fitness values $ftn_a > ftn_{rd1}$, $ftn_a > ftn_{rd2}$, thus $D2 < 1 < D1$. The chicks revolve in the surroundings of their mother for accessing the food and the equation is denoted in Eqn. (15).

$$PH^{Tis+1}_{a,b} = PH^{Tis}_{a,b} + D1 * Rad * \left(PH^{Tis}_{rd1,b} - PH^{Tis}_{a,b}\right) + D2 * Rad * \left(PH^{Tis}_{rd2,b} - PH^{Tis}_{a,b}\right) \tag{12}$$

$$D1 = \exp\left((ftn_a - ftn_{rd1}) / (abs(ftn_a) + \varepsilon)\right) \tag{13}$$

$$D2 = \exp\left((ftn_{rd2} - ftn_a)\right) \tag{14}$$

$$PH^{Tis+1}_{a,b} = PH^{Tis}_{a,b} + Flm * (PH^{Tis}_{k,a} - PH^{Tis}_{a,b}) \tag{15}$$

In Eqn. (15), the location of a^{th} chick's mother is given as PH, where $m \in [1, B]$. In order to find the food, the chick would follow their mother, which is represented as $Flm(Flm \in (0,2))$. The term Flm is taken randomly from 0 to 2 of the individual chick.

D. Conventional DHOA

DHOA [28] is inspired by the human's hunting behaviour in order to chase the deer. The objective of DHOA is to acquire the optimal position of the deer to shoot. It has some specifications, which are more difficult for human to hunt the deer. Deer's visual power is five times improved than humans. Initially, the hunter's population is initialized by PH and the corresponding equation is denoted in Eqn. (16). Here, the total number of hunters is denoted as N.

$$PH = \{PH_1, PH_2, \cdots, PH_N\} \quad 1 < p < N \tag{16}$$

The position angle and the wide angle of the deer are initialized for defining the best positions of the hunters. Assume, the circle as a search space, the wind angle follows the circumference of the circle and it is denoted in Eqn. (17). The position angle is expressed in Eqn. (18).

$$\theta_{Tis} = 2\Pi rnd \tag{17}$$

$$\phi_{Tis} = \theta + \Pi \tag{18}$$

From the above equations, the random number is denoted as rnd. The leader and the successor position are taken into account and they are represented as PH^{ldr} and PH^{scr}, respectively. The leader's position is considered as the best location of the hunter, and the successor's position is considered as the next best location of the hunter. Next, each hunter tries to be in the best position, for that the update procedure begins. At the same time, the encircling behaviour is indicated in Eqn. (19), in which D and E are the coefficient vectors. The term s is the random number on the basis of wind angle, and the values lies in between 0 and 2. The coefficient vectors are represented in Eqn. (20) and (21).

$$PH_{Tis+1} = PH^{ldr} - D \cdot s \cdot \left|E \times PH^{ldr} - PH^{Tis}\right| \tag{19}$$

$$D = \frac{1}{4}\log\left(Tis + \frac{1}{Tis_{\max}}\right)t \tag{20}$$

$$E = 2 \cdot r \tag{21}$$

In Eqns. (20) and (21), the term t is a parameter that lies in between -1 and 1. The random number is denoted as r ranging from [0, 1]. The position angle is taken into account in the update rule, and the search space enhancement is done. The angle of visualization of prey is represented by Eqn. (22). The updating procedure of the position angle is given in Eqn. (23), which is computed on the basis of variance among the wind and the visual angle. For updating the position angle to the next iteration, the mathematical formula is shown in Eqn. (24).

$$dva_{Tis} = \frac{\Pi}{8} \times rnd \tag{22}$$

$$\mathrm{var}_{Tis} = \theta_{Tis} - dva_{Tis} \tag{23}$$

$$\phi_{Tis+1} = \phi_{Tis} + \mathrm{var}_{Tis} \tag{24}$$

Similarly, the position update is done by assuming the position angle as denoted in Eqn. (25). The encircling behaviour in exploration stage is considered by regulating the vector E. Therefore, position update is done on the basis of successor's position rather than first best solution. Eqn. (26) specifies the global search equation.

$$PH_{Tis+1} = PH^{ldr} - s \cdot \left|\cos(q) \times PH^{ldr} - PH_{Tis}\right| \tag{25}$$

$$PH_{Tis+1} = PH^{scr} - D \cdot s \cdot \left|E \times PH^{scr} - PH_{Tis}\right| \tag{26}$$

The position update is done in each iteration up to the best solution is acquired.

E. Proposed DR-CSO

The optimal rule extraction and optimal training in NN is done by the proposed hybrid DR-CSO algorithm. The DR-CSO is the hybridized form of existing CSO and DHOA. The conventional CSO algorithm has high potential ability as it can attain the best optimization results with high accuracy. However, this algorithm may fall into the local optima easily. In order to overwhelm the problem, and to further improve the performance, the concept of DHOA is introduced into it. Since the sensitivity of DHOA is extremely good when compared over other algorithms, it might works well for improving the performance of CSO. The proposed DR-CSO operates based on the new fitness function F_{Tis} and the previous fitness function F_{Tis-1} in the rooster update. Here, if the $F_{Tis-1} > F_{Tis}$, the solution is updated using the rooster update from Eqn. (10) of CSO. In the other condition, the roster update is done by the update formula Eqn. (19) of DHOA. The pseudo code of the proposed DR-CSO is shown in Algorithm 2.

Algorithm 2: Pseudo code of proposed DR-CSO Algorithm

```
Initialize the chickens population as B and describe the related parameters
Validate the fitness value of B chickens, Tis is initialized to 0
while (Tis < Max_Generation)
   if (Tis%TS == 0)
      Rank the fitness values of chickens and create a hierarchal order in swarm
      Separate the swarm into various groups, and find the relationship between chicks and mother hens present in
      the group.
   end if
   for a = 1 : B
      if a == rooster
           If F_{Tis-1} > F_{Tis}
           Update roosters position by Eqn. (10)
           Else if
           Update roosters position by DHOA using Eqn. (19)
      end if
      if a == hen
           Update hen's position by Eqn. (12)
      end if
      if a == chick
           Update chick's position by Eqn. (15)
      end if
      Validate the new solution
      If the new solution is best than the earlier one, update it
   end for
end while
```

V. RESULTS AND DISCUSSIONS

A. Experimental Procedure

The proposed optimal association rule mining- based web page prediction was developed in MATLAB 2018a, and the performance evaluation was accomplished. The standard datasets used for experimentation were FIFA, KOSARAK, and MSNBC_SPMF. The number of iterations considered for analysis was 25, and the population size was 10. The performance of the suggested DR-CSO algorithm was compared with the conventional FF [29], GWO [30], WOA [31], CSO [27], and DHOA [28] algorithms by analysing the error measures such as "MEP, SMAPE, MASE, MAE, RMSE, L1-Norm, L2-Norm, and L-Infinity Norm".

B. Error Measures

In the proposed model, eight error measures are considered for analysing the web page prediction. Based on these measures, the efficiency of the proposed model is measured. The description of each measure is defined below.

1. MEP: It is "defined as the average of percentage errors by which forecasts of a model differ from actual

values of the quantity being forecast". Here, ft is the forecasted value, at is the actual value, v is the number of fitted points, and the value of computation is added for each fitted point is denoted by u .

$$MEP = \frac{100\%}{v} \sum_{u=1}^{v} \frac{at - ft}{at} \quad (27)$$

2. SMAPE: "SMAPE is an accuracy measure based on percentage errors".

$$SMAPE = \frac{100\%}{v} \sum_{u=1}^{v} \frac{|ft - at|}{\frac{(|at| + |ft|)}{2}} \quad (28)$$

3. MASE: "It is the MAE of the forecast values, divided by the MAE of the in-sample one-step naive forecast".

$$MASE = mean\left(\frac{|ft|}{\frac{1}{v-1} \sum_{u=1}^{v} |at_u - at_{u-1}|} \right) \quad (29)$$

4. MAE: "It is a measure of difference between two continuous variables".

$$MAE = \frac{\sum_{u=1}^{v} |ft_u - at_u|}{v} \quad (30)$$

5. **RMSE:** RMSE "is a frequently used measure of the differences between values predicted by a model or an estimator and the values observed".

$$RMSE = \sqrt{\frac{\sum_{u=1}^{v} (at_{u2} - ft_{u1})^2}{v}} \quad (31)$$

6. L1-Norm: "L1 Norm is the sum of the magnitudes of the vectors in a space". In this, L is a matrix, and $u = 1, 2, \cdots, n$, and n is the size of the matrix.

$$L_1 = \sum_{u} |L_u| \quad (32)$$

7. L2-Norm: "It is the shortest distance to go from one point to another". It is also known as Euclidean norm.

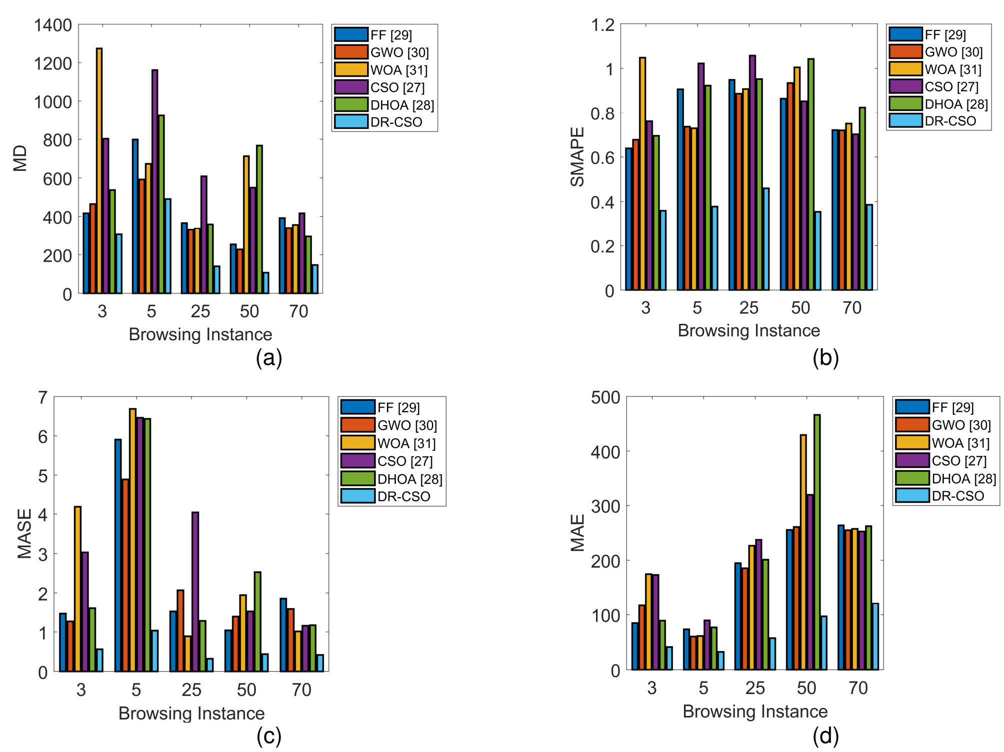
$$L2 = \left(\sum_{u=1}^{v} L_u^2 \right)^{\frac{1}{2}} \quad (33)$$

8. L-Infinity Norm: "The length of a vector can be calculated using the maximum norm".

$$L_{\inf} = \max_{1 \le u \le v} |L_u| \quad (34)$$

C. Performance Analysis using FIFA

The performance analysis of the proposed DR-CSO and the existing web page prediction models with FIFA dataset with respect to browsing instance is shown in Fig. 4. In Fig. 4, with three browsing instances, the 4th instance is predicted, and till 5 instances, the user is visiting and the 6th instance is going to be predicted in the suggested web page prediction analysis. Similarly, 26th, 51st, and 71st instance are also predicted using the suggested DR-CSO approach and other existing approaches. Here, the MEP of the developed DR-CSO technique is 46.4% better than DHOA, 64.2% better than CSO, 60.5% better than WOA, 58.3% better than GWO, and 62.5% better than FF for predicting 71th browsing instance from Fig. 4 (a). Moreover, for predicting 3rd browsing instance, the measure SMAPE of the modified DR-CSO is 45.7% enhanced than DHOA, 49.3% enhanced than CSO, 63.8% enhanced than WOA, 44.1% enhanced than GWO, and 40.6% enhanced than FF, which is shown in Fig. 4 (b). In Fig. 4 (e), the RMSE of the introduced DR-CSO is 30.2% improved than DHOA, and GWO, 25% improved than CSO, 26.8% improved than WOA, and 28.5% improved than FF for predicting the browsing instance 71. Thus, it is confirmed that the suggested DR-CSO model is well suitable for web page prediction.



(a) (b) (c) (d)

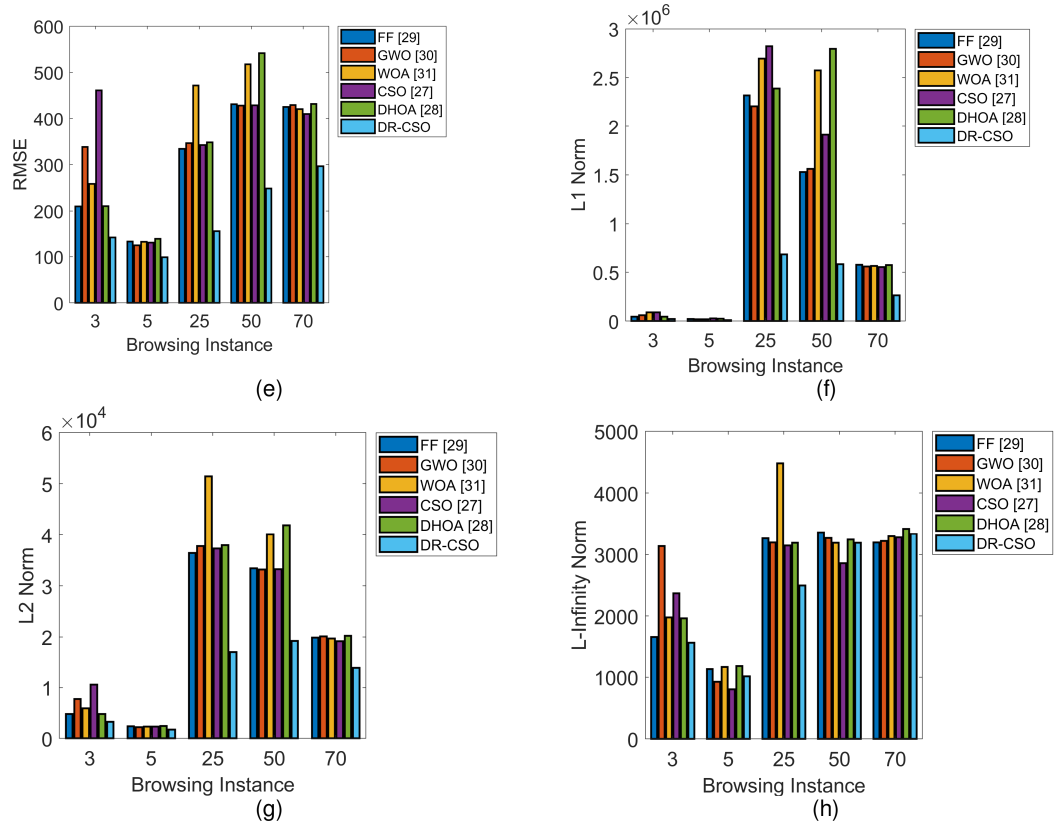
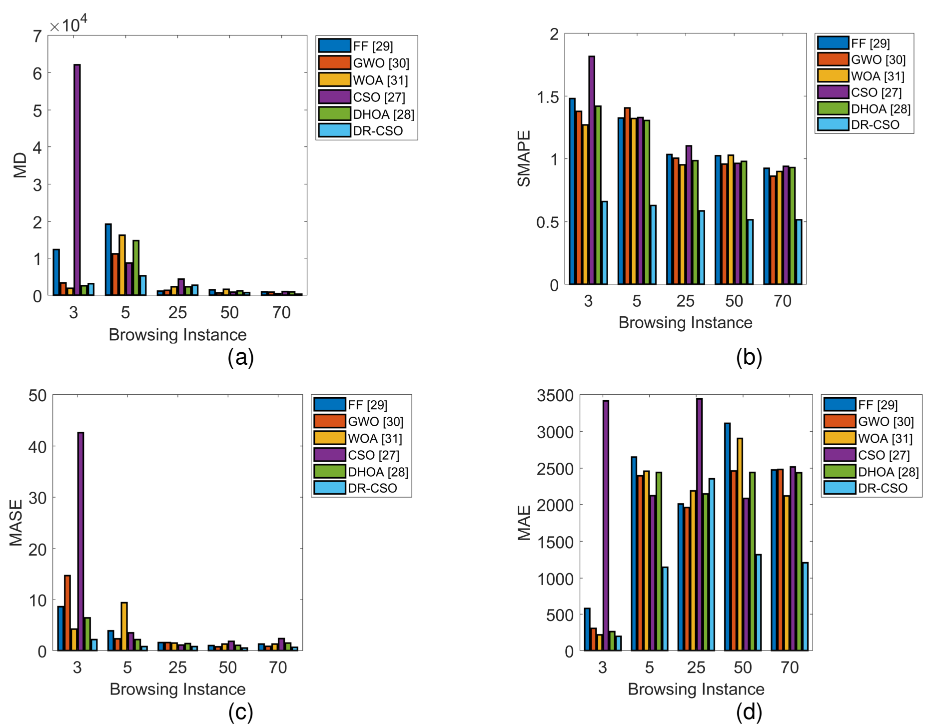


Fig. 4. Performance analysis of the proposed and conventional web page prediction models using FIFA for the measures (a) MEP, (b) SMAPE, (c) MASE, (d) MAE, (e) RMSE, (f) L1-Norm, (g) L2-Norm, and (h) L-Infinity Norm.

D. Performance Analysis using KOSARAK

Here, the performance analysis of suggested DR-CSO and traditional methods using the KOSARAK are shown in Fig. 5. From Fig. 5 (b), the SMAPE of the developed DR-CSO model is 43.3% superior to DHOA and FF, 43.9% superior to CSO, 42% superior to WOA, and 36.2% superior to GWO for predicting 70th instance. In addition, for forecasting 51th instance, the RMSE of the developed DR-CSO is 21.8% better than DHOA, 16.1% better than CSO, 34.2% better than WOA, 34.3% better than GWO, and 39% better than FF, which is shown in Fig. 5 (e). Thus, the web page prediction is predicted efficiently with the help of developed DR-CSO algorithm.



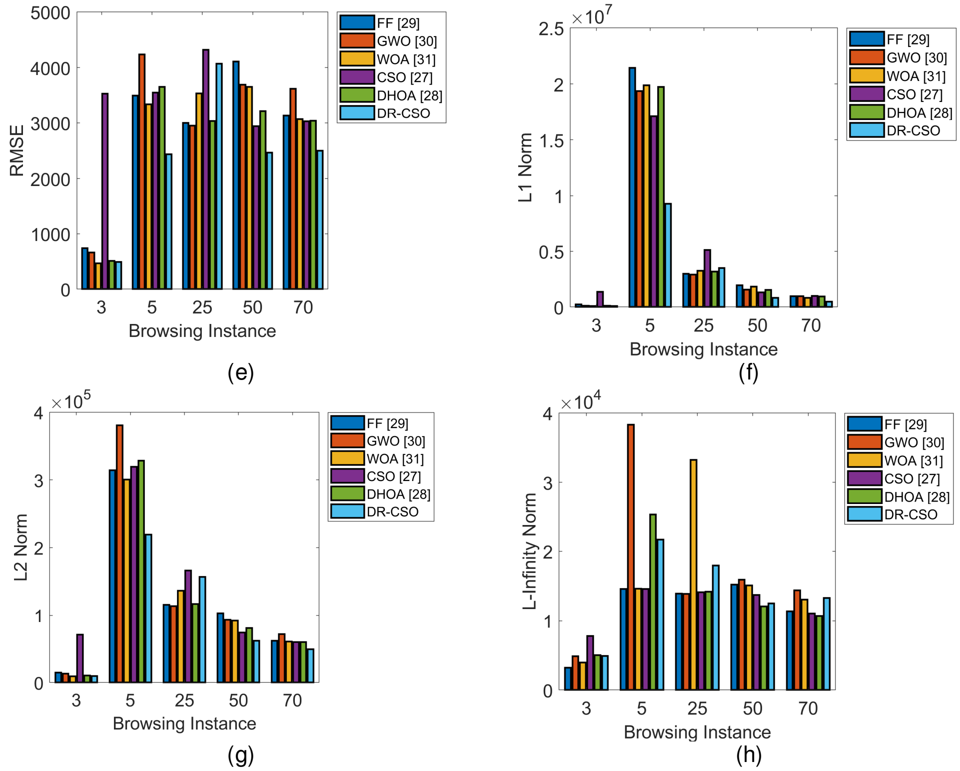
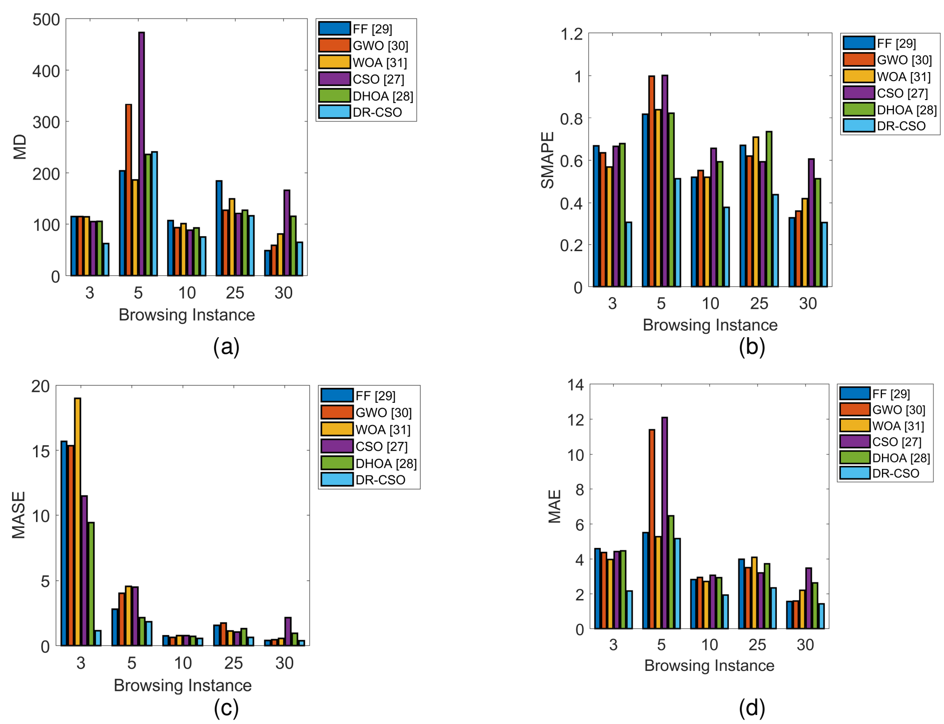


Fig. 5. Performance analysis of the proposed and conventional web page prediction models using KOSARAK for the measures (a) MEP, (b) SMAPE, (c) MASE, (d) MAE, (e) RMSE, (f) L1-Norm, (g) L2-Norm, and (h) L-Infinity Norm.

E. Performance Analysis using MSNBC_SPMF

The performance evaluation of the optimal models with MSNBC-SPMF database concerning the browsing instance is given in Fig. 6. For predicting the 30th browsing instance, the SMAPE of the suggested DR-CSO is 40%, 50%, 26.8%, 21%, and 14.2% improved than DHOA, CSO, WOA, GWO, and FF, respectively, which is shown in Fig. 6 (b). From Fig. 6 (e), the RMSE of the developed DR-CSO is 12.5% enhanced than DHOA, 14.6% enhanced than CSO, 7.8% enhanced than WOA, 16.6% enhanced than GWO, and 14.6% enhanced than FF for predicting 10th browsing instance. Similarly, the L2-Norm of the implemented DR-CSO is 14.2% better than DHOA, 5.2% better than CSO, 28% better than WOA, 10% better than GWO, and 14.2% better than FF for predicting the 25th browsing instance. Finally, it has been proven that the introduced DR-CSO approach is outperforming the classic models for web page prediction.



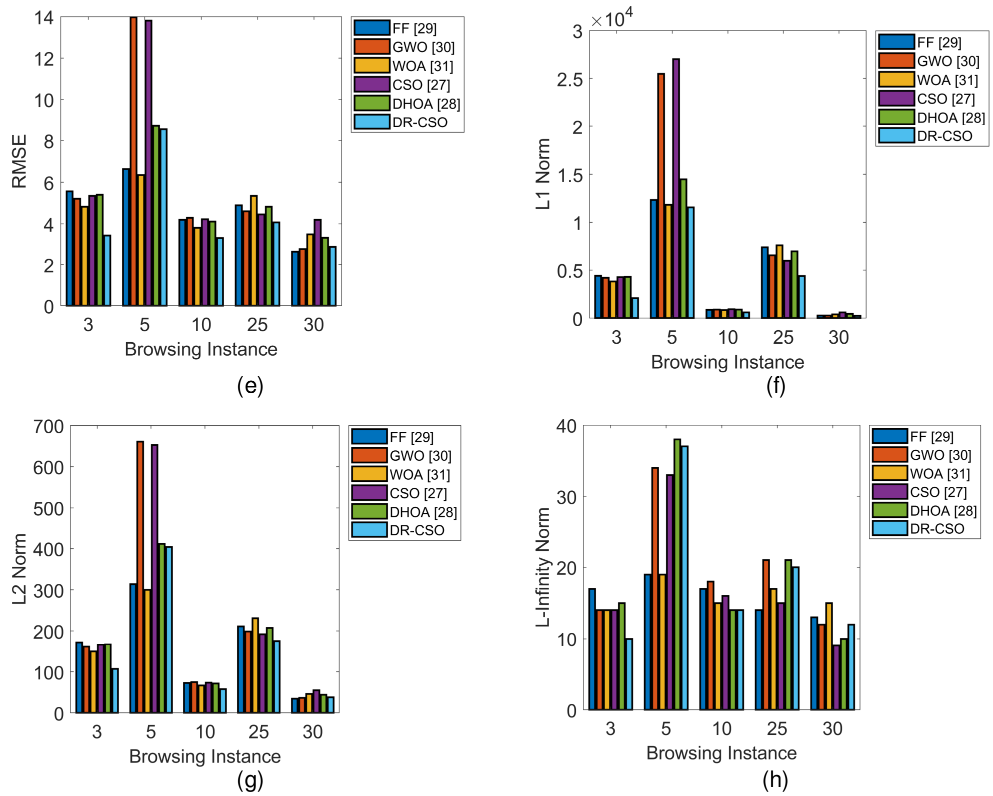


Fig. 6. Performance analysis of the proposed and conventional web page prediction models using MSNBC_SPMF for the measures (a) MEP, (b) SMAPE, (c) MASE, (d) MAE, (e) RMSE, (f) L1-Norm, (g) L2-Norm, and (h) L-Infinity Norm.

F. Overall Performance Analysis

The overall performance analysis of the proposed and the conventional web page prediction models with three datasets such as FIFA, KOSARAK, and MSNBC_SPMF are given in Table 2-4. From Table 2, the SMAPE of the suggested DR-CSO is 51.6% enhanced than FF, 48.1% enhanced than GWO, 49.4% enhanced than WOA, 56.6% enhanced than CSO, and 51.7% enhanced than DHOA. The RMSE of the implemented DR-CSO is 53.2%, 54.8%, 66.8%, 54.3%, and 55.1% superior to FF, GWO, WOA, CSO, and DHOA, respectively. In Table 3, the performance of the proposed and the suggested model are tabulated using KOSARAK dataset. Thus, the SMAPE of the presented DR-CSO is 55.4% better than FF, 52.1% better than GWO, 48% better than WOA, 63.6% better than CSO, and 53.5% better than DHOA. Moreover, the performance of the proposed DR-CSO model for the measure RMSE is 56.1% improved than FF, 26% improved than GWO, 4.8% improved than WOA, 86.1% improved than CSO, and 4.4% improved than DHOA. The overall performance analysis of the developed DR-CSO and the existing models regarding MSNBC_SPMF dataset is shown in Table 4. In Table 4, the SMAPE of the suggested DR-CSO is 27.2% improved than FF, 31.4% improved than GWO, 27.1% improved than WOA, 42.3% improved than CSO, and 36.1% improved than DHOA. In addition, the performance of developed DR-CSO approach for the measure RMSE is 20.7% better than FF, 22.5% better than GWO, 12.6% better than WOA, 21.3% better than CSO, and 19.1% better than DHOA. The L-Infinity Norm of the suggested DR-CSO is 17.6% superior to FF, 22.2% superior to GWO, 6.6% superior to WOA, and 12.5% superior to CSO. Therefore, it is concluded that the suggested DR-CSO is superior to conventional models for predicting the web page.

Table 2: Overall performance analysis of proposed and conventional models for web page prediction using FIFA.

Error Measures	FF [29]	GWO [30]	WOA [31]	CSO [27]	DHOA [28]	DR-CSO
MEP	365.92	330.13	337.73	608.89	358.64	139.89
SMAPE	0.94878	0.88584	0.90762	1.0579	0.95186	0.45904
MASE	1.5213	2.0683	0.89324	4.0412	1.281	0.32175
MAE	194.85	185.59	226.62	237.3	200.92	57.284
RMSE	333.97	346.17	471.7	342.05	347.99	156.17
L1-Norm	2.32×10^{06}	2.21×10^{06}	2.70×10^{06}	2.82×10^{06}	2.39×10^{06}	6.81×10^{05}
L2-Norm	36421	37751	51441	37303	37950	17031
L-Infinity Norm	3263	3194	4481	3143	3190	2495

Table 3: Overall performance analysis of proposed and conventional models for web page prediction using KOSARAK.

Error Measures	FF [29]	GWO [30]	WOA [31]	CSO [27]	DHOA [28]	DR-CSO
MEP	12333	3329.1	1902.6	62039	2596.5	3095
SMAPE	1.4802	1.3785	1.2687	1.8159	1.4212	0.65961
MASE	8.6314	14.802	4.2082	42.572	6.4168	2.1766
MAE	580.6	305.53	216.94	3411.2	260.44	195.66
RMSE	740.35	663.18	467.71	3533.5	512.91	490.31
L1-Norm	2.32×10^{05}	1.22×10^{05}	8.68×10^{05}	1.36×10^{06}	1.04×10^{05}	7.83×10^{04}
L2-Norm	14807	13264	9354.3	70670	10258	9806.2
L-Infinity Norm	3190	4843	3938	7750	4998	4888

Table 4: Overall performance analysis of proposed and conventional models for web page prediction using MSNBC_SPMF.

Error Measures	FF [29]	GWO [30]	WOA [31]	CSO [27]	DHOA [28]	DR-CSO
MEP	106.96	93.033	100.94	87.999	92.792	75.137
SMAPE	0.52061	0.55224	0.52004	0.65658	0.59336	0.37876
MASE	0.75015	0.63802	0.77492	0.77075	0.72115	0.5606
MAE	2.823	2.9508	2.7148	3.0754	2.9344	1.9344
RMSE	4.1894	4.2911	3.8051	4.2241	4.108	3.3221
L1-Norm	8.61×10^{02}	9.00×10^{02}	8.28×10^{02}	9.38×10^{02}	8.95×10^{02}	5.90×10^{02}
L2-Norm	73.164	74.94	66.453	73.77	71.743	58.017
L-Infinity Norm	17	18	15	16	14	14

VI. CONCLUSION

The paper has presented a new approach for predicting web page using three benchmark datasets. The suggested web page prediction model has included three steps such as rule mining, optimal rule selection, and prediction. At first, the web data gathered from three datasets such as FIFA, KOSARAK, and MSNBC_SPMF were subjected to rule mining procedure. Here, the rules were extracted using the association rule mining method named Apriori algorithm. In order to reduce the dimension of the extracted rules, optimal rule selection technique was utilized using hybrid optimization algorithm named DR-CSO. Further, the optimally selected rules were given to a machine learning algorithm called NN for predicting the user's browsing behaviour. In addition, the suggested DR-CSO was employed to perform training in NN. The overall performance regarding RMSE by the implemented DR-CSO was 53.2%, 54.8%, 66.8%, 54.3%, and 55.1% superior to FF, GWO, WOA, CSO, and DHOA, respectively using FIFA dataset. The performance of the proposed DR-CSO model using KOSARAK for the measure RMSE was 56.1% improved than FF, 26% improved than GWO, 4.8% improved than WOA, 86.1% improved than CSO, and 4.4% improved than DHOA. In addition, the performance of developed DR-CSO approach for the measure RMSE was 20.7% better than FF, 22.5% better than GWO, 12.6% better than WOA, 21.3% better than CSO, and 19.1% better than DHOA using MSNBC_SPMF dataset. Hence, it is confirmed that the developed DR-CSO-based optimal rule selection and optimal prediction models are well suitable for web page prediction.

VII. FUTURE SCOPE

By hybridizing advanced metaheuristic algorithms and improving hardware setup as a future framework for validating the tests, we can further boost the objective function.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the contributions made by anonymous reviewers whose valuable comments and insightful suggestions resulted in the improvement of this paper's preliminary version.

Conflict of Interest. The authors declare that there is no conflict of interest of any kind on this research work.

REFERENCES

[1]. Sen, E., Toroslu, I. H., & Karagoz, P. (2016). Improving the prediction of page access by using semantically enhanced clustering. *Journal of Intelligent Information Systems*, *47*(1), 165-192.

[2]. Anandhi, D., & Ahmed, M. I. (2017). Prediction of user's type and navigation pattern using clustering and classification algorithms. *Cluster Computing*, *22*(5), 10481-10490.

[3]. Dutta, R., Kundu, A., Dattagupta, R., & Mukhopadhyay, D. (2009). An approach to web page prediction using markov model and web page ranking. *Journal of Convergence Information Technology*, *4*(4), 61-67.

[4]. Yuliang, W., Qi, Z., Fang, L., Xixian, H., Guodong, X., & Bailing, W. (2019). A novel approach for Web page modeling in personal information extraction. *World Wide Web*, *22*(2), 603-620.

[5]. Nguyen, T. T. S., Lu, H. Y., & Lu, J. (2013). Web-page recommendation based on web usage and domain knowledge. *IEEE Transactions on Knowledge and Data Engineering*, *26*(10), 2574-2587.

[6]. Zhu, J., Xie, Q., Yu, S. I., & Wong, W. H. (2016). Exploiting link structure for web page genre identification. *Data mining and knowledge discovery*, *30*(3), 550-575.

[7]. Swarnakar, S., Thakur, A., Misra, D., Debopriya, P., Pakira, M., & Roy, S. (2016). Enhanced model of web page prediction using page rank and markov model. *International Journal of Computer Applications*, *140*(7), 30-34.

[8]. Liu, J., Fang, C., & Ansari, N. (2015). Request dependency graph: A model for web usage mining in large-scale web of things. *IEEE Internet of Things Journal*, *3*(4), 598-608.

[9]. Ma, Y., Wang, S., Hung, P. C., Hsu, C. H., Sun, Q., & Yang, F. (2015). A highly accurate prediction algorithm for unknown web service QoS values. *IEEE Transactions on Services Computing*, *9*(4), 511-523.

[10]. Xu, J., Zheng, Z., & Lyu, M. R. (2016). Web service personalized quality of service prediction via reputation-based matrix factorization. *IEEE transactions on reliability*, *65*(1), 28-37.

[11]. Geetharamani, R., Revathy, P., & Jacob, S. G. (2015). Prediction of users webpage access behaviour using association rule mining. *Sadhana*, *40*(8), 2353-2365.
[12]. Liu, C., Zhao, Z., & Wen, G. (2019). Adaptive neural network control with optimal number of hidden nodes for trajectory tracking of robot manipulators. *Neurocomputing*, *350*, 136-145.
[13]. A. Diaz-Manríquez, A. B. Ríos-Alvarado, J. H. Barrón-Zambrano, T. Y. Guerrero-Melendez and J. C. Elizondo-Leal, (2018). An Automatic Document Classifier System Based on Genetic Algorithm and Taxonomy. IEEE Access, *6*, 21552-21559.
[14]. Zaki, J. F., Ali-Eldin, A., Hussein, S. E., Saraya, S. F., & Areed, F. F. (2019). Traffic congestion prediction based on Hidden Markov Models and contrast measure. *Ain Shams Engineering Journal*, 1-17.
[15]. Liu, H., & Kešelj, V. (2007). Combined mining of Web server logs and web contents for classifying user navigation patterns and predicting users' future requests. *Data & Knowledge Engineering*, *61*(2), 304-330.
[16]. Pinto, P., Theodoro, I., Arrais, M., & Oliveira, J. (2017). Data mining and social web semantics: a case study on the use of hashtags and memes in Online Social Networks. *IEEE Latin America Transactions*, *15*(12), 2276-2281.
[17]. Rekik, R., Kallel, I., Casillas, J., & Alimi, A. M. (2018). Assessing web sites quality: A systematic literature review by text and association rules mining. *International Journal of Information Management*, *38*(1), 201-216.
[18]. Awad, M. A., & Khalil, I. (2012). Prediction of user's web-browsing behavior: Application of markov model. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, *42*(4), 1131-1142.
[19]. Dimopoulos, C., Makris, C., Panagis, Y., Theodoridis, E., & Tsakalidis, A. (2010). A web page usage prediction scheme using sequence indexing and clustering techniques. *Data & Knowledge Engineering*, *69*(4), 371-382.
[20]. Madjarov, G., Vidulin, V., Dimitrovski, I., & Kocev, D. (2019). Web genre classification with methods for structured output prediction. *Information Sciences*, *503*, 551-573.
[21]. Lee, C. H., Lo, Y. L., & Fu, Y. H. (2011). A novel prediction model based on hierarchical characteristic of web site. *Expert Systems with Applications*, *38*(4), 3422-3430.
[22]. Lopez-Sanchez, D., Arrieta, A. G., & Corchado, J. M. (2019). Visual content-based web page categorization with deep transfer learning and metric learning. *Neurocomputing*, *338*, 418-431.
[23]. Li, H., Xu, Z., Li, T., Sun, G., & Choo, K. K. R. (2017). An optimized approach for massive web page classification using entity similarity based on semantic network. *Future Generation Computer Systems*, *76*, 510-518.
[24]. Wu, S., Zhang, Q., Chen, W., Liu, J., & Liiu, L. (2019). Research on trend prediction of internet user intention understanding and public intelligence mining based on fractional differential method. *Chaos, Solitons & Fractals*, *128*, 331-338.
[25]. Jesus Silva, Noel Varela, Luz Adrian, Borrero López, & Rafael Humberto Rojas Millán (2019). Association Rules Extraction for Customer Segmentation in the SMEs Sector Using the Apriori Algorithm. *Procedia Computer Science*, *151*, 1207-1212.
[26]. Fernández-Navarro, F., Carbonero-Ruz, M., Alonso, D. B., & Torres-Jiménez, M. (2017). Global sensitivity estimates for neural network classifiers. *IEEE transactions on neural networks and learning systems*, *28*(12), 2592-2604.
[27]. Meng, X., Liu, Y., Gao, X., & Zhang, H. (2014). A new bio-inspired algorithm: chicken swarm optimization. In *International conference in swarm intelligence* (pp. 86-94). Springer, Cham.
[28]. G Brammya, S Praveena, N S Ninu Preetha, R Ramya, B R Rajakumar, D Binu, (2019). Deer Hunting Optimization Algorithm: A New Nature-Inspired Meta-heuristic Paradigm, *The Computer Journal*, bxy133, https://doi.org/10.1093/comjnl/bxy133
[29]. Gandomi, A. H., Yang, X. S., Talatahari, S., & Alavi, A. H. (2013). Firefly algorithm with chaos. *Communications in Nonlinear Science and Numerical Simulation*, *18*(1), 89-98.
[30]. Mirjalili, S., Mirjalili, S. M., & Lewis, A. (2014). Grey wolf optimizer. *Advances in engineering software*, *69*, 46-61.
[31]. Mirjalili, S., & Lewis, A. (2016). The whale optimization algorithm. *Advances in engineering software*, *95*, 51-67.

How to cite this article: Gangurde, Roshan A. and Kumar, Binod (2020). Optimal Association Rule Mining for Web Page Prediction using Hybrid Heuristic Trained Neural Network. *International Journal on Emerging Technologies*, *11*(1): 161–173.