INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume3, Issue2)
Available online at www.ijariit.com

A Method for Webpage Quality Improvement Using Improved Grey Wolf Optimization (IGWO) Based Extreme Learning Machine

Jayanthi .B
Rathnavel Subramaniam College of Arts & Science,
Sulur, Coimbatore - 402, TN, India
jayanthi@rvsgroup.com

Dr. S. Suganya
Rathnavel Subramaniam College of Arts & Science,
Sulur, Coimbatore - 402, TN, India
suganya_cs@rvsgroup.com

***Abstract:** In emerging websites, there is a need to measure and evaluate the quality of the website and to make it better understanding. Several metrics based on the dimensions of the criteria are content quality, organization quality, design quality and user-friendly quality. These factors depend upon the word count, page size, and average link text count and so on. This research investigates an Improved Grey Wolf Optimization (IGWO) and Extreme Learning Machine (ELM) is proposed to correspond with quality related items. The proposed approach IGWO is compared to verify the performance of Grey Wolf Optimization (GWO) Support Vector Machine (SVM) model, traditional Extreme Learning Machine (ELM) and Ivory based models are made to test the web pages. The performance metrics captured in IGWO based ELM model can be used to predict the website designs are good or bad. The simulation results have proven the superiority of the proposed method over the other competitive counterparts.*

***Keywords:** Extreme Learning Machine (ELM), Grey Wolf Optimizer (GWO), Support Vector Machine (SVM), Webpage Quality improvement.*

I. INTRODUCTION

Web page access and usage information provide sources for data mining [3]. In order to find the useful information from the huge amount of data WWW [12] is very important. It contains a rich and dynamic collection of hyperlink information. It has great challenges for data mining application. In web-based systems [10] the difficulty is to locate the appropriate algorithm to researchers and learners. But the quality assessment is still challenging research area despite website increase. The new website quality features regulate to establish a new website quality metrics with more practical measurement criteria and suitable approaches for website quality evaluation needs. In order to improve the quality as well as accuracy, there is a need for an algorithm that satisfies websites authority and users with the help of search engines. The aim of this work is to examine the characterization of website quality criteria finds out an approach to work on website evaluation for the user perception. Especially developing the practical website quality metrics and identifies each quality characteristic, sub-characteristic and measurable measures. The Extreme Learning Method (ELM) is used for intelligent classification purpose for web page quality assessment. Recently, Extreme Learning Machine (ELM) has fascinated the attention of many researchers in different applications. ELM is an advancement of single layer feed forward neural network which is an improved version of the standard feed-forward neural network.

In researchers, Nature-inspired algorithms are preferred due to their ease and flexibility. The nature-inspired Metaheuristic algorithms are analyzed in terms of their diversity and adaptation, exploration and utilization, and attractions and diffusion mechanisms. Relating to these algorithms the success and challenges are based on their parameter tuning and parameter control. Grey Wolf Optimization (GWO), is recently developed heuristics motivated from the leadership hierarchy and hunting mechanism of gray wolves in nature and has been successfully applied to solving economic problems. The main drawback of GWO is noted that the initial population is generated in a random way. It may result in the lack of diversity for the wolf swarms during the search space.

In this work improved Grey Wolf Optimizer (IGWO) is used to generate the more suitable initial positions for GWO which is more efficient training algorithm. The developed has achieved superior classification performance to the other competitive counterparts. The feature selection process consists of four basic steps [10]: (1) generation: generate the candidate subset; (2) evaluation: evaluate the subset; (3) stopping criterion: decide when to stop; (4) validation: check whether the subset is valid. The advantage of ELM is obvious in shorter training time and in compact model size (i.e., computer memory to store the trained model) while the generalization of ELM is comparable to that of SVM. In this work, the performances of ELM (with or without prior duplication) in different aspects were evaluated by comparing the results with SVM.

Devi et al., (2016) [1] proposed an implementation of a framework for website quality evaluation. Gonzalez et al., (2015), [3] presented a practical approach to website quality; it is based on the relationships between the website and its stakeholders. Malhotra and Chug (2014),[6] proposed a method of data handling (GMDH), genetic algorithms (GA) and probabilistic neural network (PNN) models the prediction of software maintainability and it was found that GMDH models predict more accurately than the other machine learning models. The major algorithms include Elishet.al achieved the prediction accuracy in software maintenance by using the Tree Net model using stochastic gradient boosting. Shi Y. And R. Eberhart discussed a modified particle swarm optimizer. Cao et al., (2012), [15] proposes an improved learning algorithm for classification which is referred to as voting based extreme learning machine. The proposed method integrates the voting method into the extreme learning machine for web page quality classification applications. Bienstock (2015), [4] the idea of service quality is increasingly important to managers of companies with a web presence. It is one of the important topics in measuring e-service quality to understand the customer's value in service transactions. The major goal is to conceptualize the e-service quality by exploring the process, outcome, and recovery dimensions of service quality Wang et al., (2015),[12] analyzed and evaluated the web services quality based on the web log analysis, they have constructed an operable evaluation indexed system that includes website construction, website function service and website benefit and cost function. Udo et al (2010), [13] reported the website quality dimensions based on his internet experience as e-customer of web service, expectations, and perceptions. They developed a quality constructs of web service and their relationships are analyzed based on behavioral intentions and customer satisfaction in an e-business environment.

TABLE I.
ATTRIBUTES FOR LINK AND TEXT ELEMENT

Metric attributes	Description
Link Element	
Total Link	Total Link on Page
Text Link	Total Text Link
Text Element	
Word Count	Total words on page
Total Body Words	Number of words in sentence
Total Sentence	Number of sentence in paragraph
Total paragraph	Number of paragraphs in body text
Total cluster count	Number of text cluster on page

The metrics of the link element, text element, image element, color element and reading complexity are listed in table I and each attribute are discussed briefly.

The remainder of this paper is organized as follows. Section 2 gives basic concepts of ELM and GWO. The detailed implementation of the IGWO method will be explained in Section 3. In Section 4 the experimental results and discussions of the proposed approach are presented. Finally, the conclusions are summarized in Section 5.

II. BASIC CONCEPTS

In this section basic concept considered for this research work is as follows:

A. Extreme Learning Method

Extreme learning machine (ELM) mainly applied for Single Hidden Layer Feedforward Neural Networks (SLFNs) it is the process of randomly selecting the input weights and systematically determines the output weights of SLFNs. This algorithm tends to the best generalization performance at extremely fast learning speed. ELM contains the three layers they are input layer, a hidden layer, and an output layer. ELM has several significant features which differ from traditional learning algorithms applied for feed-forward neural networks. The learning speed of ELM is extremely fast. The learning speed of ELM could be completed in seconds or less than seconds for many traditional applications. In the past, it seems that there exists a virtual speed barrier in which most of the classic learning algorithms cannot process and it is not an unusual way to take long time for train a feed-forward network using classic learning algorithms for uncomplicated applications. The ELM has better simplification performance compared with gradient-based learning algorithms such as backpropagation. The gradient-based learning algorithms and some other learning algorithms may face many issues such as local minima, improper learning rate and overfitting, etc. For avoiding these issues, some methods are implemented such as weight decay and stopping methods are often used in these classical learning algorithms.

In real applications, the number of hidden N nodes will always be less than the number of training samples N and the training error cannot be made exactly zero but can be a nonzero training error ϵ. The hidden node parameters $a_i$and b_i(input weights and biases or centers and impact factors) of ELM need not be tuned during training and may simply assigned with random values according to continuous sampling distribution. If the number of neurons in the hidden layer is equal to the number of samples, then H is square and invertible. Otherwise, the system of equations needs to be solved by numerical methods, concretely by solving

$$\left\|H(w_1, \dots, w_M, b_1, \dots, b_M)\hat{\beta} - T\right\| = min_\beta \|\beta - T\|$$

The result that minimizes the norm of this least squares equation is

$$\hat{\beta} = H^\dagger T$$

Where $H^\dagger$was the Moore-Penrose generalized inverse of matrix H.

The three important properties are

- Minimum training error.
- Smallest norm of weights and best generalization performance.
- The minimum norm least-square solution of Hβ = T is unique,$\hat{\beta} = H^\dagger T$

Give a training set $N = \{X1 \middle| t1 \;\; X1 \in R^n, t1 \in R^m, t1 = 1 \dots\dots N\}$activation function g(x) and hidden neuron N , do the following • Assigning random value to the input weight wi and the bias bi,i=1,……….N • Find the hidden layer output matrix H. • Calculate the output weight β

Figure. 1 Algorithm of ELM

B. Grey Wolf Optimization Algorithm

Grey Wolf Optimizer (GWO) is a typical swarm intelligence algorithm which is stimulated from the leadership hierarchy and hunting mechanism of gray wolves in nature. Grey wolves are apex predators with an average group size of 5 to 12. In GWO hierarchy, alpha (α) is considered the most dominating member among the group. The subordinates to α are beta (β) and delta (δ) helps to control the majority of wolves in the hierarchy which is considered to be omega (ω). The ω wolves are of lowest ranking in the hierarchy. The mathematical model of hunting mechanism of grey wolves consists of the following: (i) Tracking, chasing, and approaching the prey. (ii) Pursuing, encircling, and harassing the prey until it stops moving. (iii) Attacking the prey.

Step 1: Initialize the beacon frame grey wolf population X_i (i = 1, 2... n), including the pending data.

Step 2: Initialize a, A, and C and exchange the real time data in CFP period.

Step 3: Calculate the fitness of each search agent (coordinate the data)

X_α=the best search agent

X_β=the second best search agent

X_δ=the third best search agent

Step 4: while (t < Max number of iterations)

For each search agent (value) update the position of the current search agent (check the real time data position).

Step 5: Update $\alpha, A\ and\ C$, and Calculate the fitness value for all search and start CAP

Step 6: Update $X_\alpha, X_{\beta,} X_\delta, make\ t = t + 1;$

Figure. 2 Algorithm for Grey Wolf Optimizer

Encircling Prey: Grey wolves encircle the prey during the hunt. The mathematical model of the encircling behavior is given below.

$$\mathrm{D} = |\mathrm{C}.\mathrm{X_p}\ (\mathrm{t}) - \mathrm{A}.\mathrm{X}(\mathrm{t})$$

$$\mathrm{X}(\mathrm{t}+1) = \mathrm{X_p}\ (\mathrm{t}) - \mathrm{A}.\mathrm{D}$$

Where't' is the current iteration, A and C are coefficient vectors, Xp is the position vector of the prey, and X indicates the position vector of a gray wolf.

Hunting: Hunting of prey is usually guided by α and β, and δ will participate occasionally. The best candidate solutions, α, β, and δ, have better information about the potential location of prey. The other search agents (ω) update their positions according to the position of three best search agents.

Attacking Prey. In order to mathematically model for approaching the prey, we decrease the value of $\vec{a}$. The Fluctuation range of $\vec{A}$ is also decreased by $\vec{a}$. $\vec{A}$ is a random value in the interval $[-a, a]$ where a is decreased linearly from 2 to 0 over the course of iterations. When random values of $\vec{A}$ are in [−1, 1], the next position of a search agent can be in any position between its current position and the position of the prey. The value $|A| < 1$ forces the wolves to attack the prey. After the attack again they search for the prey in the next iteration, wherein they again find the next best solution α among all wolves. This process repeats till the termination criterion is fulfilled.

III. PROPOSED METHODOLOGY

The proposed a new computational framework Improved gray wolf optimization (EGWO) based ELM for feature selection approach is used for the purpose of finding the optimal feature subset for data. In the first stage, IGWO is used to filter out the redundant and irrelevant information by adaptively searching for the best feature combination in the data. In the proposed approach, genetic algorithm(GA) was firstly adopted to generate the diversified initial positions, and then gray wolf optimization (GWO) was used to update the current positions of the population in the discrete searching space, thus after getting the optimal feature subset for the better classification purpose based on EGWO. In the second stage, the effective and efficient ELM classifier is conducted based on the optimal feature subset obtained in the first stage. Figure 3 presents a detailed flowchart of the proposed GWO-ELM framework.

A. Search-Based Optimization Technique Based On Genetic Algorithm

Genetic Algorithm (GA) is a search-based optimization technique based on the principles of Genetics and Natural Selection. A Genetic Algorithm is a biological system proposed in Charles Darwin's evolution theory. It is a high-level simulation. The GA starts with a set of solutions (represented by chromosomes) called population. This process is repeated until some condition is satisfied such as achievement of the best solution. Hence the population is improved over generations to accomplish the best

solution. The GA uses three main cycles at each step to create a new population. The operator selects the individuals in the population called parents to contribute in the next. The following are general steps implemented when using GA algorithms first it generates a random initial population and creates the new population by applying the selection operators to select pairs. The number of pairs will be the population size divided by two, so the population size will remain constant between generations. By applying the crossover operator to the pairs of the s of the new population and apply the mutation operator to each pair in the new population. Suppose if the selection is not fitted than the go to step two processes.
The Genetic algorithm is explained for proposed method is given in figure2.

1. Start: Generate random population of n chromosomes strings
2. Evaluate the fitness f(x) of each chromosome x in the entire solution space.
3. Create a new population by repeating following steps until the new population is complete
4. Select two parent chromosomes (strings) from a population according to their fitness (the better fitness, the bigger chance to be selected)
5. With a crossover probability cross over the parents to form a new offspring (children). If no crossover was performed, offspring is an exact copy of parents.(tree Construction, power configuration)
6. With a mutation probability mutate new offspring at each locus (GTS).
7. Place new offspring in a new population (valid tree construction)
8. Use new generated population for a further run of algorithm
9. If the end condition is satisfied, stop, and return the best solution in current population
10. Go to step 2

Figure. 3 Algorithm for Genetic Algorithm

B. Improved Grey Wolf Optimization

The proposed IGWO works in order to generate the initial positions of the population by using GA. The current positions of the population in the discrete searching space are updated by GWO. In the second stage, the effective and efficient ELM classifier is organize based on the optimal feature subset obtained in the first stage. Figure 3 presents a detailed flowchart of the proposed IGWO. The IGWO is mainly used to adaptively search the feature space for best feature combination. The best feature combination is the one with maximum classification accuracy and a minimum number of selected features. The fitness function used in IGWO to evaluate the selected features is shown as the following equation:

$$Fitness = \alpha P + \beta \frac{N - L}{N}$$

where P is the accuracy of the classification model, L is the length of selected feature subset, N is the total number of features in the dataset, and α and β are two parameters corresponding to the weight of classification accuracy and feature selection quality, $\alpha \in [0,1]$and $\beta = 1-\alpha$. A flag vector for feature selection is shown in Figure 4.The vector consisting of a series of binary values of 0 and1 represents a subset of features, that is, an actual feature vector, which has been normalized. The ith feature is selected if the value of the ith bit equals one; otherwise, this feature will not be selected (i = 1,2,...,n). The size of a feature subset is the number of bits, whose values are one in the vector. The pseudo code of the IGWO algorithm is presented as shown in pseudo code 2.

The proposed framework consists of two main stages which are feature selection and classification, respectively. An improved gray wolf optimization, IGWO is proposed for selecting the most informative features in the specific data.

```
Begin
Initialize the parameters popsize, maxiter, ub and lb
where popsize: size of population,
maxiter: maximum number of iterations,
ub: upper bound(s) of the variables,
lb: lower bound(s) of the variables;
Generate the initial positions of grey wolves with ub and
lb; Initialize a, →using GA
A, and →C;
Calculate the fitness of each grey wolf;
alpha = the grey wolf with the first maximum fitness;
beta = the grey wolf with the second maximum fitness;
delta = the grey wolf with the third maximum fitness;
While k < maxiter
for i = 1: popsize
Update the position of the current grey wolf by Eq. (14);
end for
Update a, →
A, and →C;
Calculate the fitness of all grey wolves;
Update alpha, beta, and delta;
k = k + 1;
end while
Return alpha;
End
```

Figure. 4 Algorithm for Improved Grey Wolf Optimization

The proposed method is compared against well-known feature selection methods including GA and GWO on the two disease diagnosis problems using a set of criteria to evaluate different aspects of the proposed framework. The simulation results have verified that the proposed IGWO method not only adaptively converges more quickly, producing much better solution quality, but also gains less number of selected features, achieving high classification performance.

IV. SIMULATION RESULTS

In this section, the result of proposed model based on the web page element metrics dataset is evaluated. In this research, due to easy clustering, the web page elements dataset is suitable with IGWO and ELM model and overall valuation are made. The clustering is carried out with the wolf to calculate the accuracy of the model. The sample demonstration web page is shown in figure 4. Based on the experimental results the level of the overall reading complexity value is not dependent upon the page sizes.

The proposed IGWO model is developed by means of above study to predict good and not good web pages from the source of education, finance, news, and health. The metrics are computed by using automated, it produces raw metrics dataset. The mainly associated attributes using feature selection technique and to apply it as the input in the ELM model is extracted. The six attributes are chosen here are a total word, reading complexity, total link, total image, total color and text cluster as input dataset to train and test in this model. Through estimated metrics web pages are classified into good pages and not good pages. The captured metrics in the ELM is used to calculate the quality of website design. ELM can be used as a classifier based on the dataset of Web page metrics. In this system, MATLAB is used to implement metric computation tool and to build a prediction model.

Table II
COMPARISON OF ACCURACY OBTAINED IN DIFFERENT METHODS

Methods	Accuracy (%)
Ivory	75.8%
SVM	78.4%
ELM	88.8%
GWO-ELM	92.2%
IGWO	95.2%

Table II shows the accuracy for several comparison methods. Thus the proposed method of improved gray wolf optimizer based ELM has high accuracy when compare with other traditional methods such as GWO, SVM, ELM, and Ivory. Figure 5 shows the accuracy of GWO, SVM, ivory, ELM and proposed IGWO. Thus the proposed method of IGWO has higher accuracy and performs the better function for website quality when compared with other models.

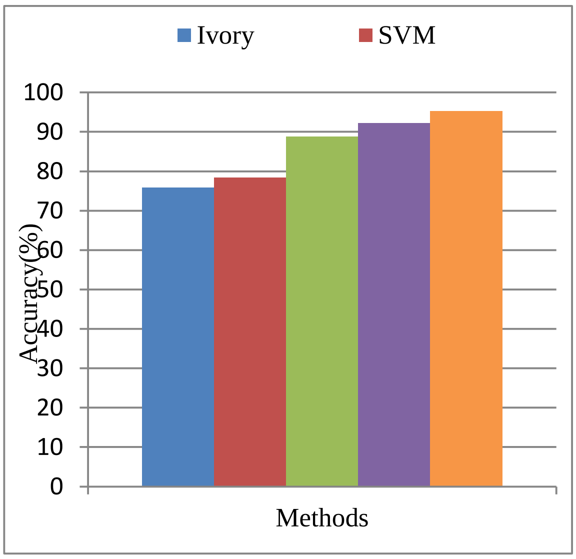

Figure. 5 Accuracy for Ivory, SVM, ELM, GWO-ELM and IGWO

CONCLUSION

In this paper, an Improved Grey Wolf Optimization (IGWO) methodology is described in detail. To evaluate the proposed IGWO as best algorithm for quality analysis, the comparison is made by GWO, Ivory, SVM and traditional ELM technique. The analysis is made on the total link, text link, word count, total body words and total sentence, based on these factors accuracy is analyzed. In future, extend this work by several optimization techniques apart from the quality factor. Since websites are developing recently by incorporating latest web technologies, hence, the quality metrics may be adjusted to adapt the challenges and to formulate the web quality model for web connecting devices. The simulation results have demonstrated that the proposed IGWO method adaptive converges more quickly and also produce much better solution quality, but also gains less number of selected features, achieving high classification performance with high accuracy.

REFERENCES

[1] Devi, K. and Sharma, A.K., 2016. Implementation of a Framework for Website Quality Evaluation: Himachal Pradesh University Website. Indian Journal of Science and Technology, 9(40).

[2] Bilgihan, A., & Bujisic, M. (2015). The effect of website features in online relationship marketing: A case of online hotel booking. Electronic Commerce Research and Applications, 14(4), 222-232.

[3] Gonzalez, M. E., Quesada, G., Davis, J., & Mora-Monge, C. (2015). Application of Quality Management Tools in the Evaluation of Websites: The Case of Sports Organizations. The Quality Management Journal, 22(1), 30.

[4] Collier, J. E., & Bienstock, C. C. (2015). A conceptual framework for measuring e-service quality. In Creating and Delivering Value in Marketing (pp. 158-162). Springer International Publishing.

[5] Lin, Hsiu-Fen. "An application of fuzzy AHP for evaluating course website quality." Computers & Education 54, no. 4 (2010): 877-888.

[6] Malhotra, R. and Chug, A., 2014. Application of group method of data handling model for software maintainability prediction using object-oriented systems. International Journal of System Assurance Engineering and Management, 5(2), pp.165-173.

[7] Wells, John D., Joseph S. Valacich, and Traci J. Hess. "What Signals Are You Sending? How Website Quality Influences Perceptions of Product Quality and Purchase Intentions." MIS quarterly 35, no. 2 (2011): 373-396.

[8] Udo, Godwin J., Kallol K. Bagchi, and Peeter J. Kirs. "An assessment of customer's e-service quality perception, satisfaction, and intention." International Journal of Information Management 30, no. 6 (2010): 481-492.

[9] O'Cass, Aron, and Jamie Carlson. "An e-retailing assessment of perceived website-service innovativeness: Implications for website quality evaluations, trust, loyalty, and word of mouth." Australasian Marketing Journal (AMJ) 20, no. 1 (2012): 28-36.

[10] Bharati, P., & Chaudhury, A. (2015). Product customization on the web: an empirical study of factors impacting choice board user satisfaction. Zheng, Z., Zhang, Y., & Lyu, M. R. (2014). Investigating QoS of real-world web services. IEEE Transactions on Services Computing, 7(1), 32-39.

[11] 'Bharati, P. and Chaudhury, A.(2006),"Product Customization on the Web: An Empirical Study of Factors Impacting Choiceboard User Satisfaction," Information Resources Management Journal, 19(2), 69-81.

[12] Wang, L., Law, R., Guillet, B.D., Hung, K., and Fong, D.K.C., 2015. The impact of hotel website quality on online booking intentions: eTrust as a mediator.International Journal of Hospitality Management, 47, pp.108-115.

[13] Udo, G.J., Bagchi, K.K. and Kirs, P.J., 2010. An assessment of customers'e-service quality perception, satisfaction, and intention. International Journal of Information Management, 30(6), pp.481-492.

[14] Ahmed, O. H., Sullivan, S. J., Schneiders, A. G., & McCrory, P. R. (2011). Concussion information online: evaluation of information quality, content, and readability of concussion-related websites. British journal of sports medicine, bjsports81620.

[15] Cao, J., Lin, Z., Huang, G.B. and Liu, N., 2012. Voting based extreme learning machine. Information Sciences, 185(1), pp.66-77. Emary, E., Zawbaa, H.M., Grosan, C. and Hassenian, A.E., 2015. Feature subset selection approach by gray-wolf optimization. In Afro-European Conference for Industrial Advancement (pp. 1-13). Springer International Publishing.

[16] Lee, K.S., and Geem, Z.W., 2005. A new meta-heuristic algorithm for continuous engineering optimization: harmony search theory and practice.Computer methods in applied mechanics and engineering, 194(36), pp.3902-3933.

[17] Goldberg, D.E. and Holland, J.H., 1988. Genetic algorithms and machine learning. Machine learning, 3(2), pp.95-99.

[18] Moscato, P., 1989. On evolution, search, optimization, genetic algorithms and martial arts: Towards memetic algorithms. Caltech concurrent computation program, C3P Report, 826, p.1989.

[19] Komaki, G.M., and Kayvanfar, V., 2015. Grey Wolf Optimizer algorithm for the two-stage assembly flow shop scheduling problem with release time. Journal of Computational Science, 8, pp.109-120.

[20] Afsar, M.M. and Tayarani-N, M.H., 2014. Clustering in sensor networks: A literature survey. Journal of Network and Computer Applications, 46, pp.198-226.

[21] Huang, Guang-Bin, Xiaojian Ding, and Hongming Zhou. "Optimization method based extreme learning machine for classification." Neurocomputing 74, no. 1 (2010): 155-163.

