ISSN NO: 2770-4491
Date of Publication:11-05-2022

A new approach for internet traffic classification: Artificial Bee Colony algorithm-OSELM

Dhulfiqar Mahmood Tawfeeq Al-Saada [1] and Maryam Thabet Hussein Al-Khazraji [2]
mrymthabthsyn57@gmail.com
Iraqi Ministry of Health

Abstract: IP traffic classification is significant for Internet service providers and other private and public organizations, for example in various tasks such as bandwidth scheduling, network error detection, Internet service quality analysis, pricing for users who use specific Internet applications, tracking Internet traffic data and security for specific government agencies. Online sequential extreme learning machine is a method of online learning solving the problem of observing data of different sizes. However, the input weights and OSELM hidden layer biases are randomly determined. This results in an incorrect classification result. In the proposed method, the data are classified using anonline sequential extreme learning machine algorithm. Certainsoftware based on artificial bee colony algorithm (ABC -OSELM) was developed to select the parameters of the sequential fast online machine learning algorithm. The simulation results show that the proposed method has achieved a 7% improvement in accuracy criteria compared to the base paper.

Keywords: IP Traffic Classification, Artificial Bee Colony Algorithm (ABC), Fast Machine Learning (ELM), Online Sequential Extreme Learning Machine (OSELM).

1. Introduction

Classifying network traffic has been a vital issue since the advent of the Internet. The research literature began with the introduction of port-based approaches that classify network traffic based on the ports used into statistical and behavioral approaches. In these methods, network traffic is analyzed in detail using machine learning and deep learning approaches. Due to the ability to classify network traffic in identifying and classifying unknown network classes, this issue has strongly attracted the attention of Internet Service Providers (ISPs) to manage the overall performance of their network [1].

Basically machine learning can be classified into supervised and unsupervised learning. In unsupervised methods, classes are not specified at first. This method is also called clustering. In this approach, similar features are clustered together. Supervised classification or learning methods determine by analyzing the data whether the classes learned from the test data are estimated or not [2]. To choose the method, we face two important issues: the accuracy of the estimated value and the estimation time.

In recent years, deep learning methods (DLT) have been highly regarded by researchers due to their good classification accuracy. Most deep learning techniques are based on multi-layered neural network architectures that provide better understanding and learning conditions for penetration features. Of course, such architectures waste a lot of time due to the long training time. With this in mind, several researchers have attempted to use shallow NN / or NN single-layer architectures that provide the same classification result and take much less time [3].

In fact, the back-propagation process plays a vital role in training networks with repetitive updated weights. This can be a significant constraint on traditional ANN networks as well as other deep learning methods and waste a lot of time updating these weights. Therefore, ELM has been proposed to solve the mentioned problem. The ELM method does not accept input-based weights updates by assigning random values to them [4].

The ELM method has been widely criticized for its use of weighting methods proposed by various authors. Hence, the online sequential extreme learning machine (OSELM) is provided to solve this challenge. OSELM divides the samples into several categories; then identifies the set weights for each sample to determine the total weight of the samples [5].

ELM and OSELM have special mechanisms to reduce network training time. This mechanism expresses a hereditary topology transmitting hidden output weight values to the sample set. The training topology is done through weight-free updating and saves a considerable amount of time.

We need several other optimization techniques such as particle swarm optimization (PSO), genetic algorithm (GA) etc. to select the optimal parameter from the entire search space. The process of optimizing one or two parameters in the model is very fast, but optimizing more than two parameters leads to slow convergence. This fact is quite evident in kernel-based approaches, and convergence occurs very rapidly in relation to the RBF kernel,because the RBF kernel has only one parameter to optimize. Of course, the reason for the slow convergence in the wavelet kernel is the existence of three parameters for optimization. This challenge can be solved by several multidimensional optimization algorithms such as PSO [6].

In this article, we use the ELM method to classify Internet traffic. The OSELM approach is one of the ELM approaches applied to data. Also, the artificial bee colony algorithm (ABC) is used to select the parameters used in the algorithm (OSELM).

In Section 2, we review the previous work. In Section 3, we present the proposed method. Section 4 discusses the test results, and the final section presents the conclusions of the paper.

2.Literature review on IP traffic classification using deep learning techniques

In [6], ELM methods are used to classify Internet traffic. The kernel-based approach is one of the ELM approaches applied to data. In particular, software (GA-WK-ELM) based on genetic algorithm (GA) has been used to select parameters in the wavelet function (WK-ELM) based algorithm. The accuracy of applying the genetic algorithm is more than 95%. The average value criterion was used to compare the classification performance. In addition, system receiver operating characteristic (ROC) curves were plotted.

ISSN NO: 2770-4491
Date of Publication:11-05-2022

In [7], kernel-based ELM methods are used to classify ML internet traffic. In this study, the use of different activation functions has achieved more than 95% accuracy. In this paper, radial and polynomial base functions are used in the kernel-based ELM function. The classification performance of the functions was measured by changing the parameters used. RBF is used for the first parameter value. Subsequently, the polynomial function is used to change the classification. Decreasing the value of the parameter used for RBF increases the accuracy. Also, the value of the parameter is equal to 0.01 and the accuracy is 95.10%. If this parameter is equal to 0.001, the accuracy value is 96.27%. Also two parameter values are used for the polynomial function. Values 1 and 10 resulted in an accuracy of 93.7%.

In [8], the proposed classification depends on two groups: scope-based and reconstruction. The three proposed classifications belong to the reconstruction and the three classifications belong to the scope. The authors offer two types of learning, namely offline and online learning to classify a class (OCC). Among the 6 methods, 4 methods can be seen offline and 2 methods are online. Also among the 4 offline methods, 2 random feature mapping methods and also 2 core feature mapping methods can be seen. The authors also provide a comprehensive discussion of these methods as well as their comparisons.

In [9] proposes a two-stage classification mechanism based on machine learning and uses network flow (NetFlow) as input. The flow of each application is divided into separate classes by the k-mean algorithm. The C5.0 decision tree classifier is also used. In the validation section, fifteen streams were collected from popular Internet applications and independently categorized as k-means and assigned to flow classes.

In [10], ELM models are used to identify botnet tasks on the Internet by grouping traffic derived from the Domain Name System (DNS). The authors defined the basics of DNS servers as well as a set of data and processed them for patent detection. The proposed solutions are effective in accordance with ELM models and algorithms based on the implementation of single-layer feed-forward neural networks. Numerous numerical tests have been proposed to validate the proposed method. ELM models have a very high training speed, they work well on datasets, and their performance is comparable to other diagnostic solutions.

In [11], the authors proposed a new classification method for network encrypted traffic as well as an intrusion detection framework called Deep Full Range (DFR). Subsequently, they used three deep learning algorithms - stackedauto-encoder (SAE), long short-term memory (LSTM) and convolutional neural network (CNN). They want to use CNN to learn the characteristics of raw traffic based on location. The LSTM algorithm is used to learn time-related properties. The stackedauto-encoder is used to extract features from encrypted features.

In [12], the authors proposed a CNN-based algorithm called Minimum-Maximum Normalization Convolution Neural Networks (MMN-CNN). This algorithm can implicitly extract specifications by self-learning training data sets, which is able to avoid the challenge of artificially extracting features as well as resolving disputes related to feature selection in different classification algorithms. Compared to the traditional classification method, the experimental results show that the proposed CNN-based traffic classification method is able to reduce the classification time and achieve the desired accuracy.

In [13], the authors grouped network traffic using two deep learning models, ResNet and CNN. Image data was generated through preprocessing method and data set was generated based on these image packages for all 8 applications. They also used network search method to find optimal hyper-parameters to improve the performance of deep learning models. Through intensive testing, they can detect that deep learning models can fine-tune network traffic and run CNN ResNet. Therefore, using the ResNet model, it is possible to group network traffic with high accuracy as well as better QoS.

In [14], the authors proposed a new method for increasing data using LSTM networks to generate traffic flow patterns as well as kernel density estimation (KDE) to multiply the numerical features of each class. Initially, they used the LSTM network to learn and generate flow packets for private classes. Next, they completed the sequence featuresby generating random values according to the characteristic distribution estimated using KDE. Finally, they applied CRNN training to large-scale datasets. Data evaluation was performed through recall and F1 criteria and accuracy for each class. The results show that the proposed design is suitable for network traffic flow data sets and also enhances the performance of deep learning algorithms.

In [15], the researchers proposed the end-to-end SDN-HGW framework to support Distributed Network Quality Management (QoS). The DataNet offers the core SDN-HGW framework, which is an encrypted data-based learning package. The proposed data network has been improved by 3 methods: CNN, SAE, MLP. An open data set with more than 20,000 packages of fifteen applications has been used to test and develop the proposed data network. The experimental results showed that the improved data network can be used in the proposed SDN-HGW framework with proper packet classification as well as high computational efficiency for real-time processing in the smart home network.

In [16], the authors proposed an end-to-end classification method for encrypted traffic based on 1-D complexity neural networks. This method combines feature extraction processes, feature selection as well as classification in an integrated end-to-end framework. To our knowledge, this is the first time an end-to-end method has been used to classify encrypted traffic. This method is validated with the ISCX VPN-nonVPN public traffic data set. A total of 4 experiments were performed on the exact model. Eleven of the 12 evaluation criteria prove the effectiveness of the proposed method.

In [17], the authors proposed a Deep Learning (DL) Method as a consistent strategy for designing traffic classifications. Features are automatically extracted and reflect complex mobile traffic patterns. Currently, there are various DL techniques for TC analysis such as a performance evaluation desk. The proposed method was evaluated on 3 datasets related to the actual activities of human users for encrypted traffic (TC) classification of mobile users.

In [18], the authors presented a new DNN. This method combines a return network and a convolution network to improve the accuracy of classification results. Convolution network is used to extract packet properties. The return network is applied to three consecutive packets of input current properties and is trained to select current properties. The proposed model is superior to the basic paper that requires the first N pack of flow. In addition, this model offers more flexibility in practice. They also compared

ISSN NO: 2770-4491
Date of Publication:11-05-2022

their proposed model with current deep learning-based work to classify encrypted traffic. Experimental results show that this model is better than works in terms of impact and efficiency.

In [19], the authors presented Deep Package, a special framework based on the use of deep learning algorithms for traffic classification that automatically extracts features from network traffic. To our knowledge, the deep package is the first traffic classification system using deep learning algorithms called 1D-CNN, SAE, which can control two sub-functions of describing traffic as well as application identification. The results show that the deep package in two sub-tasks, namely traffic description and also identifying applications so far, performs similar actions on the "ISCX VPN-nonVPN" traffic data set.

In [20], the authors proposed a traffic classification framework based on convolutional neural networks (CNN) called Seq2Img. Their main idea is to use an embedded non-parametric core compression method to convert primary flow sequences to images that completely capture the static and dynamic behaviors of different applications and prevent the use of manual features. Then proposed CNN is applied to the generated images to classify traffic. Numerous experiments have been performed on actual network traffic and the results justify the efficiency of the proposed method.

3. Proposed method

In the proposed method, the ELM algorithm is used to classify Internet traffic. Online OSELM approach is one of the ELM-based approaches to data management. In particular, software based on the artificial bee colony algorithm (ABC-OSELM) has been developed to select the parameters used in the algorithm (OSELM). The purpose of the proposed method is to classify the Internet traffic information received from computer networks and to create security policies by network administrators and improve the quality of services. As shown in Figure 1, the proposed solution consists of several components.

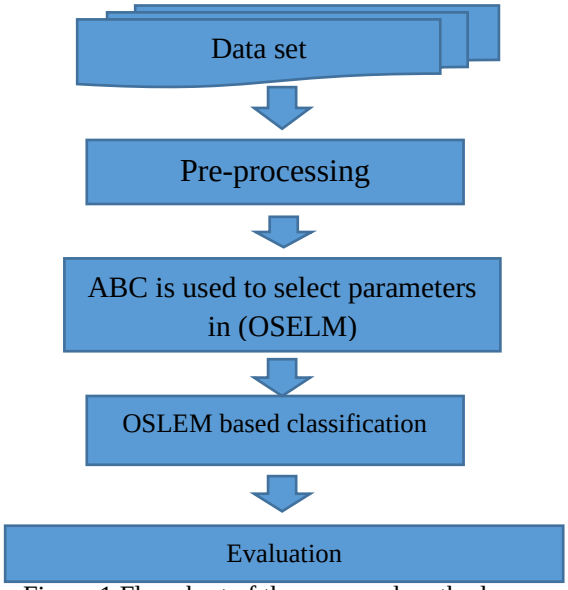

Figure 1 Flowchart of the proposed method

3.1 Pre-processing stage

Pre-processing is a task of data mining to convert raw data into an understandable format. In general, real-world data are inconsistent, incomplete, lack specific tendencies / behaviors, and likely contain many errors. The pre-processing stage solves these problems as a proven method. Pre-processing prepares raw data for further processing.

3.1.1 Cleansingthe data

Data can have many missing and irrelevant parts. To address this challenge, data cleansing is done. This process includes missing data, noise data management, and more.

3.1.2 Data conversion

This process involves converting data into accurate and convenient formats for the extraction process. The discretization process is a way of converting data to replace the numerical properties of raw values with conceptual / interval data.

3.1.3 Database categorizingfor training and testing

The third step is to categorize the data set into training and testing subsets. In the present work, 80% of the data are considered as training sets and 20% as test sets.

3.2 Artificial bee colony (ABC) algorithm for selecting parameters in OSELM

The ABC method is a swarm intelligence algorithm proposed by Karaboga to solve optimization problems in various fields. The ABC algorithm introduces three types of bees for a colony [21]:Employed bees (E), Onlooker bees (O) and Scout bees. The position of food sources is determined first (N). The population of E bees is equal to the number of food sources. Each employed bee is assigned to a food source. E bees use food sources and transmit nectar information to O bees. O bees use food and neighborhood sources based on information transmitted from E bees. The finished food source employed bee becomes a scoutbee. Scoutbees then begin the process of searching for new food sources. Nectar content information indicates the quality of the response obtained from the food source. Increasing the amount of nectar increases the likelihood of O bees choosing a particular food source [22].

3.2.1 Definition

After the scaling process, the ABC algorithm sets the best weights for the OSELM. In this regard, we use the bee colony algorithm to achieve optimal values.

ISSN NO: 2770-4491
Date of Publication:11-05-2022

3.2.2 The stage of creating a bee colony

First, a set of random numbers in the following formula is considered for the initial values of the parameters:

Parameter value = $Prange_{min}$ + ($Prange_{max}$ − $Prange_{min}$)× r

(1)

where the initial values are between the range $Prange_{max}$ - $Prange_{min}$ and $Prange_{min}$ and $Prange_{max}$ are the lower and upper bounds of the parameters, respectively. R is also a random number in the range [0, 1]. When bees are looking for food, the position of the bee is mapped to a vector equivalent to the OSELM weight vector. To determine the quality of the vector, we must evaluate it. One of the quality criteria of the solution provided by bees is the accuracy of the OSELM classification on test data. The fitness function is defined in Formula 2:

Fitness value = $OSELM_{accuracy}$

(2)

Equation 3 is used to update the location of employed bees. Equation 4 is used to determine the probability of the position of each food.

$V_{ij} = x_{ij} + \emptyset_{ij}(x_{ij} - x_{kj})$

(3)

where X_{ij} is the j dimension of the old i food source. Also, xkj is a randomly selected food source. $Q_{ij}$is a random number between [-1, 1].

$$P_i = \frac{Fit_i}{\sum_{n=1}^{N} food_{fit_n}}$$

(4)

$Fit_i$is also the fitness of my food source. Each onlooker bee goes to the selected source and continues the random search process for better fitness in the neighboring area using the formula 3.

3.3 OSLEM based classification

The best values used in the ABC approach are selected using the trial and error method to execute the classification with the most accurate values. Data were grouped using ABC -OSELM method with high accuracy. The choice of parameters used in OSELM is critical to the classification function. Better classification performance is guaranteed by changing the values of these parameters. Data were successfully grouped using ABC -OSELM method with high accuracy.

Algorithm 1: Online Sequential Extreme Learning Machine
Input: f (x) Sample of network connection Output: f (y) Connection Class (normal or abnormal)
levels: 1. Encode the connection features to real values in Ni neurons 2. Divide the data samples into several categories 3. For each category of samples 4. Do the following: 5. Adjust random weights between input and hidden WiFi 6. To calculate hidden neurons, multiply the weights by the input values: 7. Hidden neurons = ∑ Ni × Wi + input bias 8. Initialize the matrix based on the following equation, set the weights between the hidden Wh and the outputs: 9.$H = \begin{pmatrix} g(w_1 x_1 + b_1) & \dots & g(w_{\tilde{N}} x_1 + b_{\tilde{N}}) \\ \vdots & \ddots & \vdots \\ g(w_1 x_N + b_1) & \dots & g(w_{\tilde{N}} x_{\tilde{N}} + b_{\tilde{N}}) \end{pmatrix}$ 10. To calculate the output neurons, multiply the hidden weights of Wh by the hidden values of Nh: 11. Predicted output = ∑ Nh × Wh + output bias 12. Apply the matrix result to all the hidden weights of the samples in this class 13. Repeat steps 5 to 13 for the next class.

As shown in the pseudo-code above, OSELM rejects inherited weights generated by the matrix at hidden weights. Instead, OSELM divides the samples into different classes. In this way, the inheritance is placed on each class. This process can greatly improve classification accuracy, even if the inherited hidden weight does not fit some data samples. The choice of parameters used in OSELM plays a vital role in classification performance. The ABC-OSELM method basically consists of three steps. The data used to classify the traffic network are recorded in the first step. The second step is the ABC-OSELM method, which uses ABC to select the best parameter values for OSELM. In the third stage, the classification process is performed and the final decision is made.

4. Evaluation

4.1 Simulation platform

This section evaluates the proposed algorithm using computational experiments. Experiments are performed on the ISCX dataset. In order to evaluate the performance, the proposed ABC-OSELM is compared with the basic paper algorithm [6]. The proposed algorithm is also implemented in Matlab 2020b environment.

ISSN NO: 2770-4491
Date of Publication:11-05-2022

4.2Data Set

Many studies have been conducted on the classification of encrypted traffic and the proprietary traffic of security companies. Draper-Gil et al. [23] presented the ISCX dataset including seven types of regular encrypted traffic and seven types of encapsulated traffic. Such programs are very diverse and extensive. In a traffic data set, there are two data formats, including raw traffic and flow characteristics (such as the pcap format). Numerous researchers have used the ISCX dataset as a test dataset. ISCX data stream flow properties have 14 class labels. Raw traffic has no tags. Therefore, we tagged pcap files in the dataset according to their study descriptions. Multiple files such as "Facebook_video.pcap" can be labeled as "flow" or "browser", and all files related to "browser" and "VPN-Browser" have the same challenge [24]. The researchers were unable to resolve the issue after sending emails to the authors. So we decided not to tag such cases. Finally, the ISCX-tagged dataset consists of twelve classes, including six regular encrypted traffic classes and six protocol-encapsulated traffic classes. Table 1 describes the ISCX dataset.

Table 1 ISCX VPN-NONVPN data set

Traffic type	Content
Email	Gmail (SMPT, ,IMAP, POP3), Email
VPN-Email	
Chat	Hangouts, ICQ, AIM, Facebook, Skype
VPN-Chat	
Streaming	Netflix, Vimeo, Spotify, Youtube
VPN-Streaming	
File transfer	Skype, FTPS, SFTP
VPN-File transfer	
VoIP	Voipbuster, Facebook, Hangouts, Skype
VPN-VoIP	
P2P	Bittorrent, uTorrent
VPN-P2P	

4.3 Initialization of parameters

To evaluate the quality of the proposed algorithm, the researchers determined the initial values for the ABC and OSELM parameters. Hence, for maximum iteration, we set the max_iter parameter and the population size to 20 and 200, respectively. The number of food parameters and the limit are the parameters used in ABC. The population size in the PSO algorithm is equal to 200. Table 2 shows the parameter settings. First, 0.8% of the data were included as training data and 0.2% of the data as test data. Table 3 shows the parameter settings in OSELM. Where N0 is the number of initial training data used in the initial OSLEM phase. BatchSize is the volume of the batch of information learned by OSELM at each step. NumEpoch is also equal to the number of epochs.

Table 2 How to set the parameter for the proposed method

Parameters	Initial values
max_iter	100
Population size	20
limit	100

Table 3 Initial values for parameters in OSELM

Parameters	Initial values
N0	1000
Percentage of data for training	0.8
Percentage of data for testing	0.2
numHiddenUnits	100
maxEpochs	10
BatchSize	100
ActivationFunction	'sin'

4.4Evaluation criteria

In this thesis, we use the criteria of accuracy, precision, recall, and F1 score for measurement. The choice of a criterion for evaluating the effectiveness of the method depends on the problem. Suppose a number of data samples are available. This data is fed to the model individually. For each data, one class is specified as output. The class predicted by the model and the actual class can be displayed in the form of a table. This table is called the confusion matrix.

Table 4 confusion matrix

	Predicated class label		
	Predicated / Real	Normal	Attack
Real class label	Normal	True Negative (TN)	False positive (FP)
	Attack	False Negative (FN)	True positive (TP)

- True Positive: Samples that have been correctly identified by the test as an attack.
- False positive: Samples that have been mistakenly identified as an attack by the test.
- True negative: Samples that have been correctly identified as normal by the test.
- False negatives: Samples that have been mistaken for normal by a test.

ISSN NO: 2770-4491
Date of Publication:11-05-2022

Real Positive Rate (TP + FN) / TP = TPR: Equivalent to the percentage of positive samples that are accurately classified in the positive class.
False Positive Rate (FP + TN) / FN = FPR: Equivalent to the percentage of negative samples that are incorrectly classified in the negative class.

4.4.1 Accuracy criteria

The ability of a test to correctly distinguish between normal and abnormal cases is called "accuracy."To calculate the accuracy of a test, the ratio of the sum of true positive and true negative samples to the total test items is calculated. Mathematically, this ratio can be calculated with Equation (5):

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \tag{5}$$

4.4.2Precision criteria

This criterion means the ratio of positively labeled samples to real positive samples (Equation (6)):

$$Precision = \frac{TP}{TP + FP} \tag{6}$$

4.4.3 Recall criteria

This criterionshows the efficiency of the classifier according to the number of class occurrence. In fact, recall is equal to the probability of correct prediction of the absence of the desired situation by the algorithms and is calculated according to formula (7).

$$Recall = \frac{TP}{TP + TN} \tag{7}$$

4.4.4 F1 scoreCriterion

The F1 criterion is a good measure of the accuracy of an experiment. This criterion considers accuracy and recall together (Equation (5)). The criterion F1 is equal to 1 at best and equal to 0 at worst.

$$F1\ score = \frac{2 \times Precision \times Recall}{Precision + Recall} \tag{8}$$

4.5 Evaluating the results

After implementing the proposed method in MATLAB 2020b environment, we compared our proposed method on the data set to other methods [6] (GA-WK-ELM). The comparison results are shown in Table 5. The results show the superiority of ABC-OSELM over other methods. It is quite clear that convergence in RBF kernel approaches is very rapid. Because the RBF kernel has only one parameter for optimization, methods based on the Wavelet kernel converge slowly because the Wavelet kernel has three parameters for optimization [6]. This problem is solved by genetic algorithm. In the proposed method, the OSELM algorithm, the best weights are obtained with the ABC optimization algorithm. ABC can use the entire search space to optimally select the parameter. OSELM also does not spend much time during training. Figure 2 shows the accuracy diagram of the proposed ABC-OSELM method.

Table 5 comparing the efficiency of the proposed method with other methods

	Accuracy	**Precision**	**F-score**	**TPR**	**FPR**
GA-WK-ELM [6]	%81.9883	%20.836	%16.495	%28.3346	%10.272
ABC-OSELM	%88.8945	%56.4221	%55.0421	%55.5488	%6.3448

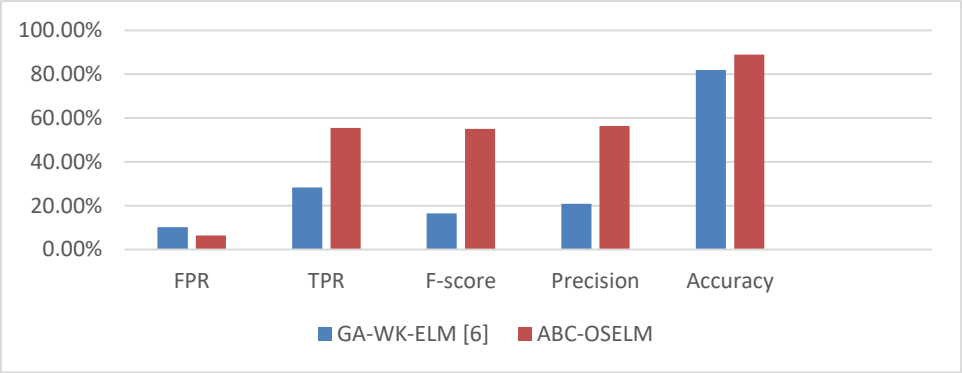

Figure 2 Comparison of the proposed method with the paper [6]

According to Figure 2, it is clear that the proposed method is more accurate than the basic paper [6]. Compared to the traditional extreme learning machine, the proposed method improves accuracy, false positive rate and false negative rate. ABC-OSELM is also more efficient than other algorithms in terms of batch input data. It was also observed that the network traffic classification findings with ABC-OSELM intelligent system are better than the findings obtained from traditional machine learning approaches. The proposed classification system has several advantages, including direct application to feature vectors, rapid training, and rapid testing. This infrastructure can be used for real-time work using these benefits. This approach consists of two steps, including classification and selecting the most appropriate weight for classification. The data set extracted from the network is given to the

ISSN NO: 2770-4491
Date of Publication:11-05-2022

machine learning classifier. Selecting the optimal weights for the ABC-OSELM classification is critical to achieving proper performance. In order to use the trial and error method, the ABC method was used to identify the best values. ABC is used in the selection of OSELM weights.

5. Conclusion

In this paper, we use the ELM method to classify Internet traffic. The OSELM approach has been used as one of the ELM approaches. In particular, certain software based on the artificial bee colony algorithm has been developed for selecting algorithm parameters (OSELM). The researchers also comprehensively compared these algorithms. Unlike modern work, test set accuracy, TPR, precision, F-score, FPR provided the conditions for accurate and comprehensive evaluation of the introduced methods. The proposed algorithm provides the best performance with 88.9% accuracy.

References

1. Tahaei, H., Afifi, F., Asemi, A., Zaki, F., & Anuar, N. B. (2020). The rise of traffic classification in IoT networks: A survey. *Journal of Network and Computer Applications*, *154*, 102538.
2. Abualigah, L., Diabat, A., & Geem, Z. W. (2020). A comprehensive survey of the harmony search algorithm in clustering applications. *Applied Sciences*, *10*(11), 3827.
3. Singh, K., & Agrawal, S. (2011, April). Comparative analysis of five machine learning algorithms for IP traffic classification. In *2011 International conference on emerging trends in networks and computer communications (ETNCC)* (pp. 33-38). IEEE.
4. Chen, Y., Xie, X., Zhang, T., Bai, J., & Hou, M. (2020). A deep residual compensation extreme learning machine and applications. *Journal of Forecasting*, *39*(6), 986-999.
5. Li, Y., Qiu, R., & Jing, S. (2018). Intrusion detection system using Online Sequence Extreme Learning Machine (OS-ELM) in advanced metering infrastructure of smart grid. *PloS one*, *13*(2), e0192216.
6. Ertam, F., & Avcı, E. (2017). A new approach for internet traffic classification: GA-WK-ELM. *Measurement*, *95*, 135-142.
7. Ertam, F., & Avcı, E. (2016). Network traffic classification via kernel based extreme learning machine. *International Journal of Intelligent Systems and Applications in Engineering*, 109-113.
8. Gautam, C., Tiwari, A., & Leng, Q. (2017). On the construction of extreme learning machine for online and offline one-class classification—An expanded toolbox. *Neurocomputing*, *261*, 126-143.
9. Bakhshi, T., & Ghita, B. (2016). On internet traffic classification: A two-phased machine learning approach. *Journal of Computer Networks and Communications*, *2016*.
10. Ghafari, J., Herbert, E., Senecal, S., Migault, D., Francfort, S., & Liu, T. (2014). Extreme learning machines for Internet traffic classification. In *ESANN*.
11. Zeng, Y., Gu, H., Wei, W., & Guo, Y. (2019). $ Deep-full-range $: A deep learning based network encrypted traffic classification and intrusion detection framework. IEEE Access, 7, 45182-45190.
12. Zhou, H., Wang, Y., Lei, X., & Liu, Y. (2017, December). A method of improved CNN traffic classification. In 2017 13th International Conference on Computational Intelligence and Security (CIS) (pp. 177-181). IEEE.
13. Lim, H. K., Kim, J. B., Heo, J. S., Kim, K., Hong, Y. G., & Han, Y. H. (2019, February). Packet-based network traffic classification using deep learning. In 2019 International Conference on Artificial Intelligence in Information and Communication (ICAIIC) (pp. 046-051). IEEE.
14. Hasibi, R., Shokri, M., & Dehghan, M. (2019). Augmentation scheme for dealing with imbalanced network traffic classification using deep learning. arXiv preprint arXiv:1901.00204.
15. Wang, P., Ye, F., Chen, X., & Qian, Y. (2018). Datanet: Deep learning based encrypted network traffic classification in sdn home gateway. IEEE Access, 6, 55380-55391.
16. Wang, W., Zhu, M., Wang, J., Zeng, X., & Yang, Z. (2017, July). End-to-end encrypted traffic classification with one-dimensional convolution neural networks. In 2017 IEEE International Conference on Intelligence and Security Informatics (ISI) (pp. 43-48). IEEE.
17. Aceto, G., Ciuonzo, D., Montieri, A., & Pescapé, A. (2018, June). Mobile encrypted traffic classification using deep learning. In 2018 Network traffic measurement and analysis conference (TMA) (pp. 1-8). IEEE.
18. Zou, Z., Ge, J., Zheng, H., Wu, Y., Han, C., & Yao, Z. (2018, June). Encrypted traffic classification with a convolutional long short-term memory neural network. In 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS) (pp. 329-334). IEEE.
19. Lotfollahi, M., Siavoshani, M. J., Zade, R. S. H., & Saberian, M. (2020). Deep packet: A novel approach for encrypted traffic classification using deep learning. Soft Computing, 24(3), 1999-2012.
20. Chen, Z., He, K., Li, J., & Geng, Y. (2017, December). Seq2img: A sequence-to-image based approach towards ip traffic classification using convolutional neural networks. In 2017 IEEE International conference on big data (big data) (pp. 1271-1276). IEEE.
21. Chen, C. N., Hoang, T. T., & Cho, M. Y. (2019). Parameter optimisation of support vector machine using mutant particle swarm optimisation for diagnosis of metal-oxide surge arrester conditions. Journal of Experimental & Theoretical Artificial Intelligence, 31(1), 163-175.
22. Blondin, J., & Saad, A. (2010, August). Metaheuristic techniques for support vector machine model selection. In 2010 10th International Conference on Hybrid Intelligent Systems (pp. 197-200). IEEE.
23. Draper-Gil, G., Lashkari, A. H., Mamun, M. S. I., & Ghorbani, A. A. (2016, February). Characterization of encrypted and vpn traffic using time-related. In Proceedings of the 2nd international conference on information systems security and privacy (ICISSP) (pp. 407-414).
24. Wang, W., Zhu, M., Wang, J., Zeng, X., & Yang, Z. (2017, July). End-to-end encrypted traffic classification with one-dimensional convolution neural networks. In 2017 IEEE International Conference on Intelligence and Security Informatics (ISI) (pp. 43-48). IEEE.