International Journal of Engineering

Journal Homepage: www.ije.ir

Detecting Fake Websites Using Swarm Intelligence Mechanism in Human Learning

F. Parandeh Motlagh*, A. Khatibi Bardsiri

[a] *Department of Computer Engineering, Kerman branch, Islamic Azad University, Kerman, Iran*
[b] *Department of Computer Engineering, Bardsir branch, Islamic Azad University, Bardsir, Iran*

PAPER INFO

Paper history:
Received 28 November 2017
Received in revised form 09 April 2018
Accepted 26 May 2018

Keywords:
Fake Websites
Phishing Attacks
Artificial Neural Network
Swarm Intelligence Algorithm
Learning- based Optimization Algorithm

ABSTRACT

The internet and its various services have made users to easily communicate with each other. Internet benefits including online business and e-commerce. E-commerce has boosted online sales and online auction types. Despite their many uses and benefits, the internet and their services have various challenges, such as information theft, which challenges the use of these services. Information theft or phishing attacks are internet attacks that are major approach to success it is social engineering that the phisher has used. In these types of attacks, the attacker deceives the users and steals their valuable information by using a fake website that looks like real websites. The damage caused by fake websites and phishing attacks is so high that researchers are trying to identify these types of websites in different ways. So far, various methods have been developed to identify phishing web sites which most of them based on data- mining and learning machine are trying to identify these malicious websites. Artificial neural network is a data-mining method for identifying phishing websites which is used in most studies; however the error rate of this can be significant in detecting these websites, so learning-based optimization algorithm is used as a Swarm intelligence algorithm to reduce its error. In the proposed method, the error rate of multi-layer artificial neural network in detecting phishing websites is considered as a target function which minimized by using learning-based optimization algorithm. In the proposed method, learning- based optimization algorithm selects weights and bias of multi-layer artificial neural network optimally to minimize the error of clssification as an objective function. The datasets used to evaluate the proposed method are Phishing Websites explaind by others. The results of evaluating phishing attack dataset indicate that the rate of error of fake website detection in the proposed method is constantly reduced by repetition. The results of our assessment also indicate that the average accuracy, sensitivity, specificity, precision of the proposed method are 93.42, 92.27, 93.19 and 92.78%, respectively. The decision tree and regression are more accurate in detecting fake websites than artificial neural network.

doi*: 10.5829/ije.2018.31.10a.05*

1. INTRODUCTION

The internet is a global network of interconnected computers that connects millions of people around the world. The internet has a variety of attractive services, such as social networks or sales websites, which make millions members join it. An important part of the internet is to serve e-commerce and online sales, and so far many web sites have been launched to provide these services [1]. Today, important contributions from financial transfers, online purchases, online auctions, ticket purchases, etc. are provided by the websites that provide this service. The emergence of smartphones has led a large volume of online users to refer to the services they seek from these websites. The store like Amazon and eBay, including websites, has millions of visitors and offers them services and sales [2]. The use of online services for various websites, such as banks, stores and sales, has reduced the number of people traveling to these places and receive the services online, and this has many benefits, for example, it is very effective in reducing traffic and air pollution. Although the online services of a variety of websites have many advantages, this phenomenon can also create security challenges for users. One of the security challenges associated to a variety of online is stealing websites or phishing attacks. In this type of attack, a phisher or

*Corresponding author Email: farnooshm865@gmail.com (F. Parandeh Motlagh)

Please cite this article as: F. Parandeh Motlagh, A. Khatibi Bardsiri, Detecting Fake Websites Using Swarm Intelligence Mechanism in Human Learning, International Journal of Engineering (IJE), IJE TRANSACTIONS A: Basics Vol. 31, No. 10, (October 2018) 1642-1650

online thief tries to attract users' comments and led them to enter a fake website instead of legitimate website. Phisher steals individuals' important information such as bank accounts and their user accounts by leading them to enter fake websites [3]. Phisher uses social engineering or malware techniques to deceive users and leads them to enter fake websites [4]. Phishing attack mechanism shows that there is a repetitive cycle and at first, phisher sends a fake link to users and wants them to click on it and enter to fake website, then phisher attempts to steal user's information. The ways phisher sends fake links includes E-mail, chat, video conference, social networks and SMS. Identifying fake websites from legal websites is one of the important security layers to increase the trust of users in online and e-commerce services [5]. One of the other challenges that fake websites create is their huge loss to information technology, which is estimated at billions of dollars, so identifying these websites is very important. Fake websites have properties legitimate websites do not have these features in common, so we can identify fake websites by identifying these properties. The number of fake websites as a google page-rank and domain is low and the number of citation to other websites is high [6]. One of the important ways to detect fake websites is using data-mining on Knowledge Discovery (KD) in these fake pages. An artificial neural network (ANN) is used as an important and efficient data mining tool for recognizing pattern and classifying information. We can consider multi-layer Perceptron (MLP) as one of the types of artificial neural network (ANN) for learning and recognizing pattern which has a simple structure that is efficient for learning. If selecting weights and thresholds used in multi-layer Perceptron (MLP) is optimum, it can decrease detection error of fake websites, so it is a NP-Hard optimization problem. In this paper, we use swarm intelligence based on human behavior in learning a new proposed method and suggest a new method to improve learning Multi-layer Perceptron (MLP) on detecting fake websites that the swarm intelligence decrease the error of classifying fake and legal websites.

2. PHISHING ATTACKS

Deceiving internet users and stealing their usernames and passwords is an example of internet crime and its background dates back to early 1990s. The first report about phishing attacks was about 1987 that this subject enters into internet security issues [7]. When the phishing challenge arose, it focuses on English users more than others, but today we see that these attacks are widespread throughout the world. In traditional phishing attacks, a phisher tries to communicate to users by using high public relations and humility and deceives them to open fake websites or take users' important information using different social engineering methods. In some cases, phisher uses social engineering to scare users; phisher threatens ordinary users; if they do not respond to emails, their service is disconnected or their accounts are blocked. In some cases, phisher also uses social engineering based on big promises, phisher introduces itself as a representative of monetary organizations and intends to reward users in exchange for their information. The root of phishing word is taken from "fishing"; because the phishers place their victims on fake websites and wait for them. The dimensions of phishing are wide, for example a phisher steals usernames and passwords and tries to participate in money laundering or illegal purchases. However, phishers focus on customers, they also feign the websites of some organizations and compromise their brands and reputations. There are several definitions of "phishing"; so users have a good understanding of these attacks. Some definitions suggest that phishing requires a precise sociology to identify Fisher's goals of its actions. For example, an anti-phishing team believes that phisher uses the mechanism of both social engineering techniques and technical problems to steal personal information and financial accounts [8]. Ming et al. [9] believed that phishing is a crime based on computer networks that are victims of losing fake and false information to a web site. Kirda et al. [10]asserts that phishing is creating an online fake. Zhang et al. [11] also believe that phishing and fake websites have at least two high visibility similarities with the legal website and the existence of a word information page. The above definitions are common in that the phishing website provides a fake version of the legal website and the use of high social skills to deceive users into sending their data to phisher. Until recently, phishers only sent fake emails to persuade their victims to enter fake sites or give them their information, but today, in addition to their email tools, they use social networks effectively to deceive users and try to expand fake links on these social networks. According to an estimate in a research institute, it has been made clear that a phisher sends an average about 3252 fake emails to their victims daily. According to Microsoft researches, it has been estimated that about 0.4% of emails were sent to users for phishing attacks. As social networks have a large number of users, so the phishers try the best to fake these social networks and studies show that the creation of fake pages from social networks has grown by about 12% in 2012. Obtaining information from a social network is important because the phisher can identify the friends or followers of a victim and send them fake email to deceive them. For example, a phisher can put itself in the position of one of our friends and pretend it loses its wallet and has an urgent need for money. Phishing websites are progressing rapidly now and it seems their growth is faster than a variety of phishing countermeasures [12]. The phishers use different methods for their attacks and they can be divided into the following categories [13]:

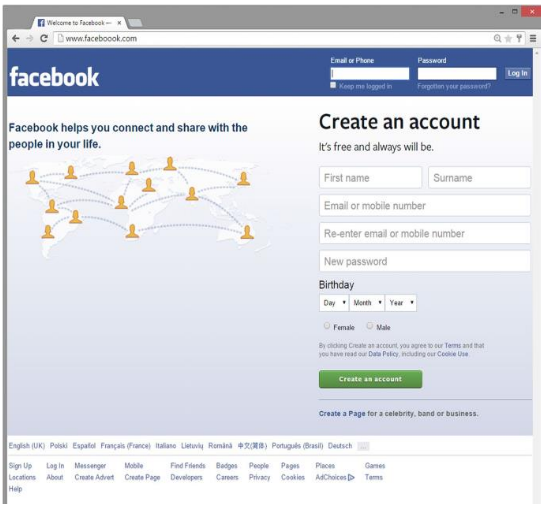
- Mimicking attack: In this attack, phishers send a fake email to their victims usually and ask them to click on the link and enter the fake site and use the encryption protocol for increasing users' trust.
- Forward attack: In this manner, the victim clicks on available link on email and enters a fake login page in the email.
- Pop- up attack: In this manner, as soon as user logins a fake website, a pop-up page appears and asks user login user accounts so this information send to attacker and thief.

Figure 1 shows a fake Facebook page which allows users log in [14]. In the this figure, appearance of a fake website is designed to be similar to the legal Facebook site, so users are less likely to doubt it.

The most of focus of fake websites which steal online include financial goals because phisher can act as money launder or earn income. The damage caused by phishing attacks is significant, because these damages focus on financial area more than others. Figure 2 shows that billions of dollars entail losses to e-commerce annually [15].

According to Figure 1 the amount of damage caused by fake websites in recent years, especially from 2013 to 2016, has been significant losses, thus identifying methods of phishing or fake websites is a major research. Identifying phishing attacks and fake websites requires that the process and mechanism of these attacks will be understand well. Figure 3 shows the mechanism of phishing attacks in a single cycle.

According to Figure 2, we consider a cycle of attacks in phishing attacks. In this cycle, phisher fakes a legal website at first and send the fake link to users as E-mail.



Welcome to Facebook —

www.faceboook.com

Facebook helps you connect and share with the people in your life.

Create an account

It's free and always will be.

By clicking Create an account, you agree to our Terms and that you have read our Data Policy, including our Cookie Use.

Create a Page for a celebrity, band or business.

facebook
Email or Phone
Password
Log In
Keep me logged in
Forgotten your password?
First name
Surname
Email or mobile number
Re-enter email or mobile number
New password
Birthday
Day
Month
Year
Female
Male
Create an account
English (UK) Polski Español Français (France) Italiano Lietuvių Română 中文(简体) Português (Brasil) Deutsch
Sign Up Log In Messenger Mobile Find Friends Badges People Pages Places Games
Locations About Create Advert Create Page Developers Careers Privacy Cookies AdChoices Terms
Help

Figure 1. The fake website of Facebook social network

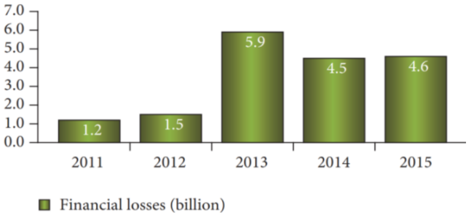
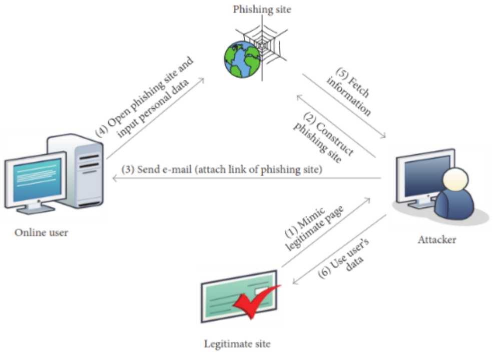


Figure 2. The losses from fake websites in recent years [15]



Phishing site
(4) Open phishing site and input personal data
(5) Fetch information
(2) Construct phishing site
(3) Send e-mail (attach link of phishing site)
Online user
Attacker
(1) Mimic legitimate page
(6) Use user's data
Legitimate site

Figure 3. The cycle of phishing attacks

So the users click on fake link in their Email and open fake websites, then the fake websites steal their information and send it to phisher and this mechanism is regularly implemented by phisher in the next cycle. Fake websites and the attacks based on them are done with various techniques to increase the chances of attacking and stealing information. Jain and Gupta [15] presented two following types of phishing attacks according to Figure 4:

- Social engineering
- Malware- based

In social engineering attacks, phisher tries to deceive users and steal valuable information through a wide range of social relationships, such as email, telephone, social networks, fake web sites, and so on. In malware-based attacks phisher tries to steal users' information or enter them to malicious websites unintentionally and steal valuable information using various types of malware such as viruses, worms, Keyloggers, Session Hijacking, Host File Poisoning, faking DNS Phishing, fake Content Injection, System Reconfiguration Attacks, fake Phishing search engines [15].

Different methods for identifying fake web sites are presented and introduced, which are generally divided into two categories of user awareness and software detection methods. User awareness method tries to increase users' knowledge for identifying fake websites. Software detection method tries to detect phishing web sites by various software techniques automatically. User awareness methods face many limitations including need for user training that is not widely welcomed.

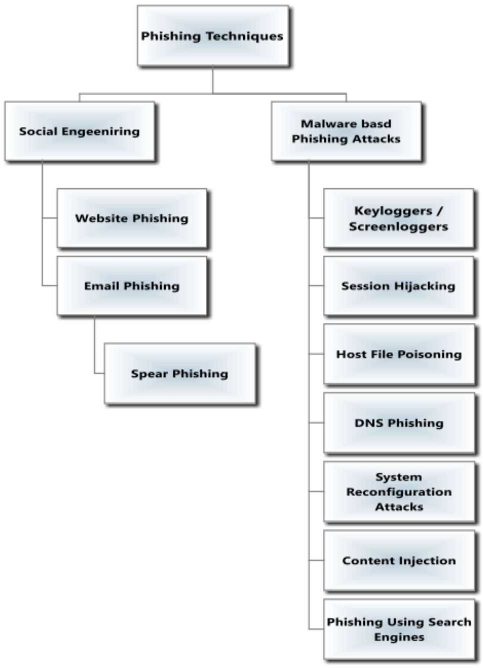


Figure 4. Classifying the ways of stealing information [15]

Software detection methods are more widely used than awareness-based methods, which are generally divided into four categories [16]:

- Methods based on backlist: In this method, a list of fake website addresses will be placed in a database to prevent users from visiting. However, the method is not very efficient, because the database needs to update regularly.
- Methods based on heuristics: In this method, a fake website is detected using exploratory methods. On the other hand, the efficiency of this method depends on the accuracy of the heuristics function.
- Methods based on visual similarity: In this method, a fake website is detected by the visual similarity available on website.
- Methods based on machine learning: In these methods, a fake website will be detected by using different techniques of discovery of knowledge and based on the properties used on web sites.

3. SWARM INTELLIGENCE BASED ON HUMAN'S BEHAVIOR

The optimization algorithm is based on education and learning a swarm intelligence algorithm includes learning and teaching approach in humans and it is of limited algorithms which are modeled on human behavior. In this algorithm, each problem is identified as a person who is in a training course. In this algorithm, anyone who can get the best score undertakes the role of teacher and the rest undertake the role of learners. In this algorithm, members of the population are trained and educated to raise the level of class knowledge, and each member of the community, in a kind of way, helps other members to move towards the optimal answer, which here is the maximum average. The results of statistical studies of different grades show that their scores distributed normally around the mean value of the class. Figure 5 shows the scores of two different classes which are normally distributed.

According to Figure 5, a class that has an average score M2 has higher educational qualities because its average value is greater than the average of another class or M1. The overall mechanism of optimization algorithm based on learning and education can be summarized in Figure 6:

According to above figure, in the optimization algorithm based on the education and learning, each member of the population who has obtained the best score will be considered as a teacher in the next repetition so he/she guides and educates other members. In this figure, TA is considered as the best member of the population in the current repeat because it has gotten the best score, so he/she plays the role of teacher in the next repetition. It is observed that the scores of the class have been improved and the average scores have also increased so that in new repetition, TB has the best score and will be considered as a teacher.

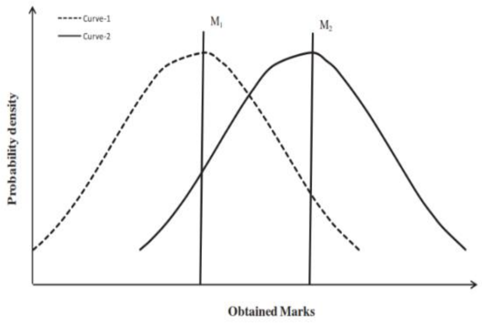
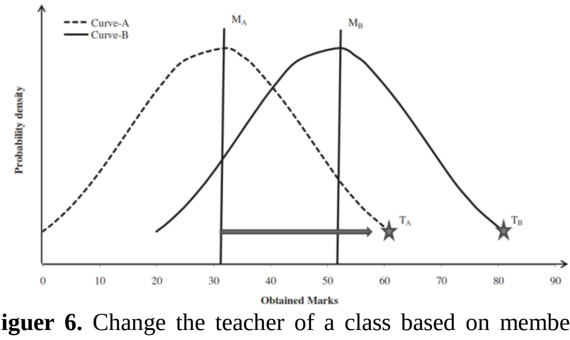


Figure 5. Distribution of the scores of two classes around the average [17]



Figuer 6. Change the teacher of a class based on members' average score [17]

In Equation (1), the teaching process modeling is represented by the best member of a population or a teacher in an education-learning optimization algorithm:

$$X_{new} = X_{old} + r(X_{teacher} - T_F M) \quad where \quad T_F \in \{1,2\} \tag{1}$$

where, X_{old} is the current position of a solution, X_{new} is the new position of a solution, X_teacher is the position of the best member of population or teacher, M is the average students score, TF is intensity of learning equaled 1 or 2, r is a random number at (0, 1). If X_{old} is a member of education-learning optimization algorithm, it can learn with other members' help, so two random members of population select to learn here such as xi and xj which have two modes by merit:

- Merit or score of xi is more than xj
- Merit or score of xj is more than xi

In the first case, xi has the role of learner more than xj and in the second, xj has the role of learner more than xi for x_{old} which modeled according Equations (2) and (3), respectively.

$$x_{new} = x_{old} + r(x_i - x_j) \tag{2}$$

$$x_{new} = x_{old} + r(x_j - x_i) \tag{3}$$

The results of various simulation shows that education-learning optimization algorithm solves the optimization problems with higher accuracy and convergence than ant Colony Optimization Algorithm (ACO), Genetic Algorithm (GA), Particle Swarm Optimization (PSO), Harmony Search (HS), Differential Evolution (DE) and in this regard, it is an efficient algorithm [17].

4. PROPOSE METHOD

Multi-layer Perceptron (MLP) is an example of a variety of neural networks modeled based on the structure of a brain and the relations between neurons. In multi-layer Perceptron (MLP) a set of neurons placed on hidden layers that their roles are influencing on input and creating a suitable output. Figure 7 shows the function of an artificial neuron in multi-layer Perceptron (MLP) for creating output and examples of input:

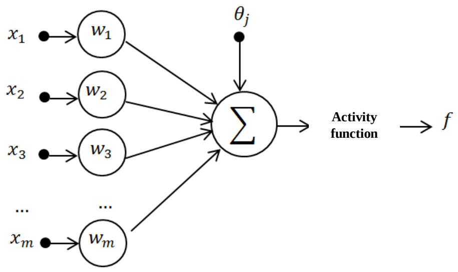


Figure 7. The effect of a neuron on inputs to weights and bais of neurons

According to Figure 7, each neuron multiplies its inputs in specified weights, and collects these weights with a bais or threshold value. Each neuron can affect its initial output under the influence of a mapping called the activity function to control its output variation. The output of a neuron can be considered as Equation (4) in form of total coefficients at the input of a neuron add a bais or threshold value:

$$f = \sum_{i=1}^{m} \boldsymbol{w_i}\, \boldsymbol{x_i} + \theta_j \tag{4}$$

Where wi is a weight, xi is an input of a neuron, m is the number of weights used in neuron input, θ_j is the number of threshold or bais of jth network neuron and f is output of a neuron. The output of neuron could influence by an activity function to limit its change domain. Equation (5) shows an example of this activity function that here is Gaussian.

$$F = \frac{1}{1+exp(-f)} \tag{5}$$

The desired activity function place the output on [0, 1]. Multi-layer Perceptron (MLP) could be used in classifying and detecting pattern and it could be trained based on the properties of internet pages and we can use it to detect fake websites. The training mechanism in training process of multi-layer Perceptron (MLP) could choice the amounts of weights and thresholds optimally, but these values are not necessarily optimal because they are optimized if the degree of detection or classification error for web-fake sites is minimal. Minimizing the classification's error of fake web sites actually indicates that the weight and bias of the artificial neural network are chosen appropriately and optimally. For example, if two neural networks named ANN1 and ANN2 are used to detect and classify fake websites and the ANN1 error rate is lower than ANN2, then it can be concluded that the weights and baises used in the ANN1 artificial neural network is more optimum than ANN2 artificial neural network. The optimal selection of weights and baises used in the artificial neural network can reduce the amount of error detected by fake Web sites and the error of detecting fake and legitimate websites reduce more by choosing baises and weights more optimally. We can use objective function in Equation (6) to detect how much multi-layer perceptron (MLP) qualified on identifying fake websites:

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 \tag{6}$$

Where n is the number of test samples, Yi is the real class number of an example which may be fake or legitimate, Y ̂_i is the estimated class number which estimated by using artificial neural network. Certainly, an appropriate and optimum neural network has the minimum error value or MSE and reducing this error means that the baises and weights select optimally. We can consider this problem by optimization perspective whose objective function is defined in Equation (6) and the end goal is minimizing the error of objective function. In the proposed method, each artificial neural network is considered as a solution to the optimization problem and it is attempted to identify an optimal artificial neural network for further reduction of detection errors of fake websites. In other words, much multi-layer Perceptron (MLP) considered as a list or array of weights and baises of artificial neurons and we try to select a list of optimal weights and baises for reducing the objective function. We can define the optimization problem in the proposed method as a more complete objective function of Equation (7):

$$op = \begin{cases} min \quad e = \frac{1}{n}\left(\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2\right) \\ s = \ll w_1, w_2, \dots., \theta_1, \theta_2, \dots. \gg \end{cases} \tag{7}$$

The above optimization problem to detect fake websites is a difficult problem and methods such as gradient can't find the optimal answer, but we can extract the optimal bais and weight of artificial neural network using fragmentary algorithms or swarm intelligence and then propose an optimal artificial neural network to detect phishing attacks or fake websites. Figure 8 shows a general framework for detecting fake websites by using much multi-layer Perceptron (MLP) and learning-education optimization algorithm.

According to the framework, we consider an artificial neural network with a given number of layers and neurons as a member of population in learning-education optimization network. Here, any artificial neural network or its equivalent is defined in the education- learning optimization algorithm as an array of weights and bais according to Equation (8):

$$s = \ll w_1, w_2, \dots., \theta_1, \theta_2, \dots. \gg \tag{8}$$

Multi-layer Perceptron (MLP) coded as a member of TLBO population and a set of weights and bais.

The initial population of TLBO is selected by using a set of randomly weights and bais.

Every artificial neural network or population member of the TLBO algorithm is evaluated by the objective function of the problem or the following objective function:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

We consider an artificial neural network or its equivalent in the TLBO population with a minimum detection error in identifying fake websites as a community lecturer.

The learning and education process in TLBO algorithm is implemented on population members or artificial neural networks and their weights and bias are updated.

Repeat the steps of the TLBO algorithm and update the weight and baises of artificial neural networks and evaluate them.

Selecting optimal artificial neural network or the best member of TLBO algorithm by minimum error of MSE for detecting fake websites.

Figure 8. The framework of proposed method

By encoding any artificial neural network in the form of Equation (8), we create some of them randomly in the form of Equation (9) in the problem space:

$$s_i = l + (u - l).rand(0,1) \tag{9}$$

In this equation, we use MLP and select it randomly to determine the initial population with its weight and bais between two upper and lower ranges are shown l and n, respectively. In this equation, s_i is artificial neural network or i-th member of population in proposed algorithm. Here, each Here, each of these solutions is considered to be a member of the population of education-learning optimization algorithm that is according to Equation (10) a set of them contains the initial population:

$$s = \{s_1, s_2, \dots, s_N\} \tag{10}$$

In this equation, S is the members of population of education-learning optimization algorithm and N is the number of initial population. Each of these members is evaluated by the objective function of the problem defined in (6) and the minimum of them selects for the objective function as the artificial neural network and according to Equation (11) this teacher guides the members of population that here means select new weights and baises for any artificial neural network:

$$s_{new} = s_{old} + r(s_{teacher} - T_F M) \quad where \;\; T_F \in \{1,2\} \tag{11}$$

Here, S_{old} is a MLP, S_{new} is an artificial neural network trained by teacher and its weights and baises updated, $S_{teacher}$ is the optimum artificial neural network in role of teacher in current repetition, the optimization algorithm is based on learning and teaching. The perpose of this equation is influencing weights and baises of artificial neural network belongs to the most optimum artificial neural network in role of teacher. According to the steps of learning-education optimization algorithm, each member of population - here are neural networks – can influence and teach each other. In order to learn the artificial neural networks from each other, which is a kind of weight changing process and based on other artificial neural networks, we can use Equations (12) and (13) based on competency:

$$s_{new} = s_{old} + r(s_i - s_j) \tag{12}$$

$$s_{new} = s_{old} + r(s_j - s_i) \tag{13}$$

In these equations, we consider the competency or detection error of fake websites in s_i is more than s_j or detection error of fake websites in s_j is more than s_i, respectively. In above equations, the artificial neural network such as s_{old} uses the weights and baises of two neural network named s_i and s_j to update its weight and bais and create new neural network s_{new} which has optimum weight and bais. Repeating the steps of learning-training optimization algorithm on each artificial neural network, their weights and baises update in each repetition and based on their new weight and bias in detecting website resulted in that the minimum ones are used in the final repetition for the final detection of fake websites.

5. ANALYSIS

We used the dataset associated to phishing attacks available and free on online database UCI [18]. This data set includes 11055 samples and 30 input properties and one output for any record valued 1 and -1 that represent real and fake websites, respectively. Each sample or record represents the properties of a website and link. Dataset characteristics are categorized into four different categories based on different items, such as link attributes or user referrals, source code contents, properties associated with the web server according to Table 1.

Considering that proposed method in this research is a data-based methodology, it is necessary to classify the information in order to assess classification criteria such as accuracy, sensitivity, specificity and precision.

In Equations (14), (15), (16) and (17), the criteria for the assessment of accuracy, specificity, sensitivity and precision are presented respectively:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{14}$$

$$Specificity = \frac{TN}{TN+FP} \tag{15}$$

$$Sensitivity = \frac{TP}{TP+FN} \tag{16}$$

$$Precision = \frac{TP}{TP+FP} \tag{17}$$

TABLE 1. Four different category of dataset characteristics

Description	Feature Type
These features describe a link or an address such as the number of address points, address length and so on	Features of website addresses
These features are related to web pages and the feature of the website, such as the site rank on Google and so on	Features unrelated to pages and links
These features are related to the source code of the website, javascript code, and so on that one example of this is deactivation right- click by users	Features related to client or server side
These features are related to credibility, life span, and other domain indexes, for example the hosting of fake websites' life is not high	Features related to domain

In these equations, we use evaluation criteria TP, FP, TN, FN which are True positive (TP), False positive (FP), True negative (TN), Flase negative (FN), respectively.

One of the important properties of learning- based optimization algorithm is non- use of any parameters in the algorithm which is a unique swarm intelligence algorithm. In implementation of the proposed method, we use an artificial MLP with hidden two layers and 5 artificial neurons placed on each layer of it. An example of output of the proposed method is displayed for 10 members population and 50 repeats in Matlab Software (see Figure 9).

According to the above output, the degree of classification error and the identification of fake websites in proposed method show a downward trend and this sentence confirms that the selection of weights and bays of the artificial neural network based on repetition of the algorithm is constantly convergent to optimality. In fact, learning-based optimization algorithms make the selection of weights and bays in the artificial neural network move in terms of repetition to optimality. This is an important factor in reducing the detection error of fake website. In other words, learning-based optimization algorithm makes the learning of the artificial neural network deeper; because the detection error of these websites shows a down trend as it repeats. For evaluating proposed algorithm according to criteria such as accuracy, sensitivity, specificity, precision, we consider the number of initial population 20 and the number of repetitions of algorithm 30, then we calculate the values of evaluation criteria TP, FP, TN, FN and calculate the average of them in 40 different experiments and we use indicators such as accuracy, sensitivity, specificity and precision. The results of evaluation are summarized in Table 2.

According to various experiments, we can conclude that the average value of accuracy, sensitivity, specificity, precision is 93.42, 92.27, 93.19 and 92.78%, respectively. As mentioned above, the proposed method is a data-mining method because it uses the artificial neural network as a knowledge discovery tool whose learning is also enhanced by learning-based optimization algorithm.

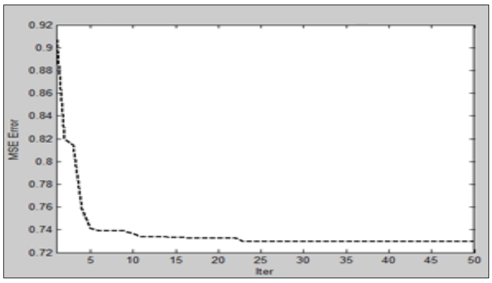


Figure 9. An output of implementing the proposed method in Matlab Software

TABLE 2. The average values of accuracy, sensitivity, specificity, precision

Average value, %	Index
93.42	Accuracy
92.27	Sensitivity
93.19	Specificity
92.78	Precision

Therefore, we can compare it by some data-mining methods in detecting fake websites for comparing efficiency. One of the important data-mining tools in pattern recognition and classification is WEKA Software which developed based on Java programming language. We use precision for comparing the proposed method with other data mining methods such as artificial neural network, decision tree and regression technique. The results are shown in Figure 10.

Comparing precision index in the proposed method as an important evaluation indicator with other data-mining methods in detecting fake websites indicates that the precision index in proposed method, regression method, decision tree, MLP is 92.78, 90.20, 91 and 89.40%, respectively. It shows that the value of precision in proposed method to detect fake websites is more than other methods. It means that the proposed method with precision 92.78% could detect how much a fake website is really fake, with its precision being significant.

6. CONCLUSION

Fake websites are considered as one of the major challenges of internet and e- commerce and cause a lot costs as losses to users, online institutes and internet infrastructures annually. In this type of security challenge, a fake website place itself instead of a real website and tries to attract users and steal information by using various social engineering techniques. We can also identify fake websites according to their properties.

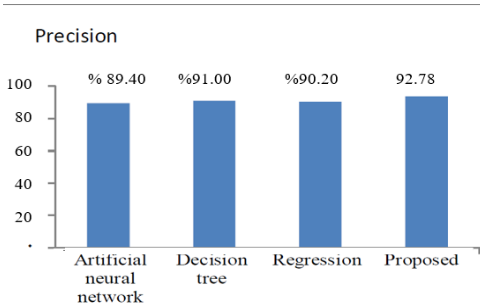


Figure 10. Comparison of the proposed method with other methods for detecting fake websites based on precision

In this paper, we tried to detect these attacks according to their properties using an artificial neural network. In our proposed method, we use learning- based optimization algorithm to increase learning ability of an artificial neural network to select the weight and bus selection mechanism with high accuracy and thus the amount of detection error of fake websites. Our evaluation results indicate that the proposed approach is more accurate than artificial neural network, decision tree and regression technique in identifying fake websites. One of the challenges of the proposed learning approach is based on the high profile of the data set that can delay detection of fake websites. Therefore, in the future, we will try to reduce the feature space associated with fake web sites and apply according to the important learning feature.

7. REFERENCES

1. Cotten, S.R., Ford, G., Ford, S. and Hale, T.M., "Internet use and depression among older adults", ***Computers in Human Behavior***, Vol. 28, No. 2, (2012), 496-499.
2. Tsang, S., Koh, Y.S., Dobbie, G. and Alam, S., "Detecting online auction shilling frauds using supervised learning", ***Expert Systems with Applications***, Vol. 41, No. 6, (2014), 3027-3040.
3. Derouet, E., "Fighting phishing and securing data with email authentication", ***Computer Fraud & Security***, Vol. 2016, No. 10, (2016), 5-8.
4. Krombholz, K., Hobel, H., Huber, M. and Weippl, E., "Advanced social engineering attacks", ***Journal of Information Security and applications***, Vol. 22, (2015), 113-122.
5. Chawla, M. and Chouhan, S.S., "A survey of phishing attack techniques", ***International Journal of Computer Applications***, Vol. 93, No. 3, (2014).
6. Basnet, R.B., Sung, A.H. and Liu, Q., "Feature selection for improved phishing detection", in International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems, Springer., (2012), 252-261.
7. James, L., "Phishing exposed, Elsevier, (2005).
8. May, O.A.D., "23, 2011 from the us patent and trademark office re", ***US Applied***, No. 11/980,690.
9. Qi, M. and Yang, C., "Research and design of phishing alarm system at client terminal", in Services Computing, 2006. APSCC'06. IEEE Asia-Pacific Conference on, IEEE., (2006), 597-600.
10. Kirda, E. and Kruegel, C., "Protecting users against phishing attacks with antiphish", in Computer Software and Applications Conference, 2005. COMPSAC 2005. 29th Annual International, IEEE. Vol. 1, (2005), 517-524.
11. Zhang, Y., Hong, J.I. and Cranor, L.F., "Cantina: A content-based approach to detecting phishing web sites", in Proceedings of the 16th international conference on World Wide Web, ACM. (2007), 639-648.
12. Intelligence, M., "Annual security report", *Symantec Corp*, (2010), http://www.symantec.com/connect/blogs/messagelabsintelligence-annual-security-report-2009-security-yearreview.
13. Mohammad, R.M., Thabtah, F. and McCluskey, L., "Tutorial and critical analysis of phishing websites methods", ***Computer Science Review***, Vol. 17, (2015), 1-24.

14. Iuga, C., Nurse, J.R. and Erola, A., "Baiting the hook: Factors impacting susceptibility to phishing attacks", ***Human-centric Computing and Information Sciences***, Vol. 6, No. 1, (2016), 8.

15. Jain, A.K. and Gupta, B.B., "Phishing detection: Analysis of visual similarity based approaches", ***Security and Communication Networks***, Vol. 2017, No., (2017).

16. Khonji, M., Iraqi, Y. and Jones, A., "Phishing detection: A literature survey", ***IEEE Communications Surveys & Tutorials***, Vol. 15, No. 4, (2013), 2091-2121.

17. Rao, R.V., Savsani, V.J. and Vakharia, D., "Teaching–learning-based optimization: An optimization method for continuous non-linear large scale problems", ***Information Sciences***, Vol. 183, No. 1, (2012), 1-15.

18. Mohammad, R., Thabtah, F.A. and McCluskey, T., "Phishing websites dataset", (2015).

Detecting Fake Websites Using Swarm Intelligence Mechanism in Human Learning

F. Parandeh Motlagh, A. Khatibi Bardsiri

[a] *Department of Computer Engineering, Kerman branch, Islamic Azad University, Kerman, Iran*
[b] *Department of Computer Engineering, Bardsir branch, Islamic Azad University, Bardsir, Iran*

PAPER INFO

Paper history:
Received 28 November 2017
Received in revised form 09 April 2018
Accepted 26 May 2018

Keywords:
Fake Websites
Phishing Attacks
Artificial Neural Network
Swarm Intelligence Algorithm
Learning- based Optimization Algorithm

چکیده

اینترنت و خدمات متنوع آن باعث شده است تا کاربران به سادگی بتوانند با یکدیگر ارتباط برقرار نمایند. از مزایای اینترنت می توان به کسب کار آنلاین و تجارت الکترونیک اشاره نمود. تجارت الکترونیک باعث رونق فروش آنلاین و انواع حراج آنلاین در اینترنت شده است. اینترنت و سرویسهای آن علیرغم کاربرد و مزایای زیاد دارای چالشهای مختلفی نظیر سرقت اطلاعات نیز می باشند که استفاده از این خدمات را با چالش مواجه می سازد. سرقت اطلاعات یا حملات فیشینگ نوعی حمله اینترنتی است که رویکرد اصلی در موفقیت آن مهندسی اجتماعی است که سارق بکار می برد. در این نوع حملات مهاجم با استفاده از یک وب سایت جعلی که شبیه وب سایت های واقعی است کاربران را فریب داده و اطلاعات باارزش آنها را مورد سرقت قرار می دهد. زیان ناشی از وب سایتهای جعلی و حملات فیشینگ به قدری زیاد است که پژوهشگران سعی میکنند با روشهای مختلف این نوع وب سایتها را شناسایی نمایند. تاکنون برای شناساسایی وب سایتهای فیشینگ روش های مختلفی ارایه شده که بیشتر آنها بر اساس روش های داده کاوی و یادگیری ماشین سعی می کنند این وب سایتهای مخرب را شناسایی نمایند. شبکه عصبی مصنوعی یکی از روش های داده کاوی در تشخیص وب سایتهای فیشینگ است که در بسیاری از پژوهش ها استفاده شده است با این وجود میزان خطای آن در تشخیص این وب سایتها می تواند قابل توجه باشد از این جهت برای کاهش خطای آن ،الگوریتم بهینه سازی آموزش و یادگیری، به عنوان یک الگوریتم هوش دسته جمعی انسانی استفاده شده است. در روش پیشنهادی میزان خطای شبکه عصبی مصنوعی چند لایه در تشخیص وب سایتهای فیشینگ به عنوان یک تابع هدف در نظر گرفته شده است و با استفاده از الگوریتم بهینه سازی آموزش و یادگیری مقدار این تابع کمینه می شود. در روش پیشنهادی الگوریتم بهینه سازی آموزش و یادگیری اوزان و بایاس شبکه عصبی مصنوعی چند لایه را به گونه ای بهینه انتخاب می نماید تا میزان خطای طبقه بندی به عنوان تابع هدف کمینه شود. مجموعه داده بکار رفته جهت ارزیابی روش پیشنهادی Phishing Websites است که توسط رامی و همکاران ارایه شده است. نتایج آزمایشات نشان می دهد که میزان خطای تشخیص وب سایتهای جعلی در روش پیشنهادی بر حسب تکرار کاهش می یابد. همچنین نتایج ارزیابی نشان می دهد متوسط دقت، حساسیت، تشخیص و صحت روش پیشنهادی به ترتیب برابر ۹۳/۴۲٪ ، ۹۲/۲۷٪، ۹۳/۱۹٪ و ۹۲/۷۸٪ می باشد و نسبت به شبکه عصبی مصنوعی چند لایه، درخت تصمیم گیری و رگرسیون دقت بیشتری در تشخیص وب سایتهای جعلی دارد.

***doi**: 10.5829/ije.2018.31.10a.05*