Turkish Journal of Electrical Engineering & Computer Sciences

http://journals.tubitak.gov.tr/elektrik/

Research Article

Turk J Elec Eng & Comp Sci
(2017) 25: 4757 – 4767

doi:10.3906/elk-1612-279

Proposing a new clustering method to detect phishing websites

Morteza ARAB, Mohammad Karim SOHRABI*
Department of Computer Engineering, Semnan Branch, Islamic Azad University, Semnan, Iran

Received: 23.12.2016 • **Accepted/Published Online:** 20.06.2017 • **Final Version:** 03.12.2017

Abstract: Phishing websites are fake ones that are developed by ill-intentioned people to imitate real and legal websites. Most of these types of web pages have high visual similarities to hustle the victims. The victims of phishing websites may give their bank accounts, passwords, credit card numbers, and other important information to the designers and owners of phishing websites. The increasing number of phishing websites has become a great challenge in e-business in general and in electronic banking specifically. In the present study, a novel framework based on model-based clustering is introduced to fight against phishing websites. First, a model is developed out of those websites that already have been identified as phishing websites as well as real websites that belong to the original owners. Then each new website is compared with the model and categorized into one of the model clusters by a probability. The analyses reveal that the proposed algorithm has high accuracy.

Key words: Phishing, clustering, banking website, data mining, security

1. Introduction

Development of information and communication technologies has had positive effects on and outcomes in different areas of scientific, social, and economic domains of those societies that enjoy such technologies. Today, experts do not pay attention to specific cases in the analysis of development indices; they evaluate direct and indirect effects of each influential factor in combination with other relevant factors. The increasing influence of applicable and scientific tools on different social activities (especially economic structures) changed traditional approaches to business and markets radically. In the same vein, daily development of the Internet in different countries and the connection of a greater number of people around the world to the World Wide Web along with an increasing frequency of electronic communications among people and various organizations through the Internet and cyberspace provide desirable conditions for establishing economic and commercial interactions. Among these changes, the advent of e-banking is one of the favorable changes [1].

Certain phenomena like electronic business and trades, as efficient methods in increasing transactions and communicating with customers, and users' increasing usage of them have attracted the attention of many professionals to this technology so that in most of cases virtual offices are used besides physical ones to increase the market share and facilitate communication with customers. This has also led to the development of certain activities such as e-training, e-banking, and other electronic services. Regarding the convenience, speed, and security of this method, countries are thinking about using these brand-new technologies more than ever [2].

E-banking is a specific type of baking that utilizes an electronic environment (e.g., the Internet) for serving customers. In fact, all banking operations will be a kind of banking service. In this sort of banking,

*Correspondence: amir_sohraby@yahoo.com

all banking operations are done electronically and all of these operations are protected with a sufficient level of protection [3].

Because of ease of access to e-banking at any time, most users log into banking portals and websites to do logical and legal operations. However, some imposters log into portals to steal subscribers' accounts and achieve illegal objectives. This is called defrauding in the law but in computer science it is called phishing. Phishing is the criminal act of illegal access to secret and sensitive information such as username, password, and credit card details by showing an electronic connection as apparently trustworthy [4].

Phishing is normally done via email and SMS. The user is invited to enter information into a fake Internet site. The website is an exact copy of the graphic interface of a valid website, like online banking websites. The user is initially directed to this fake page by email or commercial ads of other sites. Users are then requested to enter certain sensitive data like information of their credit cards. If a user enters his/her data, phishers will access the private information. Certain websites such as PayPal, EBay, and online banks are common targets of phishing activities [5].

The phishing method was described in detail in 1987. The term was first used in 1995. It is an abbreviation for "password harvesting fishing", which means hunting for a user password by using bait. In this case, "phish" conveys the notion of cheating. The lowest presumable level of damage is an access to email and the highest risk is stealing from online financial accounts. This kind of information theft is prevailing increasingly and it might be due to the fact that naive people tend to disclose their personal information when facing an online thief. The main concern is that a thief might change the data received from various people and create fake accounts in victims' names to misuse their credibility and damage their reputations [6].

Various applications have been developed for phishing attacks, such as NoteCraft, which reports the location of hosting and the riskiness of the site against a phishing website that the user logs into. This service could be used for some web browsers. In addition, Google Safe tries to stop phishing attacks by presenting a special toolbar. Opera has arranged for some security measures by maintaining and enhancing the security of its web browsers. Microsoft Corporation resolved security gaps of version 9.5 and higher of its famous web browser Internet Explorer.

Through a short review of the problems, one can easily see that these invasions have spread so quickly that large software companies like Microsoft and regulations of countries have opted for measures that secure the safety of products. In the present study, phishing attacks against banking portals are detailed, types of phishing are discussed, and the problem is addressed in depth by developing certain algorithms.

In the remaining parts of the paper, a literature review is presented in Section 2. In Section 3, the proposed clustering method to detect phishing websites will be explained. Section 4 represents the experimental results and their evaluation, and, finally, the work will be concluded in Section 5.

2. Related works

Phishing can be described as an online threat. Alternatively, it is raised as intervention in an authentic website to obtain users' private data such as username, password, and social security numbers. Phishing websites are designed by imposters for copying authentic websites. These websites have high similarity with real websites to deceive the users. Due to the increasing number of such attacks, many researchers have tackled phishing. Due to the variety of data mining techniques [7], the increasing improvement of its methods [8–15], and the ability to apply its methods and techniques in pattern recognition and classification, data mining plays an important role in providing new and advanced methods to detect phishing attacks. In the following text, some of these studies will be discussed.

Ali et al. used a new expert system for detecting phishing attacks against e-banking systems [16]. They used a new expert system to detect phishing attacks. The proposed expert system uses distinct characteristics of an authentic website to distinguish it from a fake website; it is able to give necessary reasons for reporting the extent to which a website is doing phishing attacks. The main idea behind using the proposed expert system is adaptation of an artificial neural network to diminish the number of rules and increase the inference rate.

In a study by Mohammad et al., a machine learning method was adopted for modeling prediction and monitoring of learning algorithms such as multilayer perceptron, decision tree induction, and simple Bayesian classification so as to search among outputs [17]. The results showed that decision tree classification is more accurate in detecting phishing sites than other learning algorithms.

An effective image-based antiphishing plan was proposed in [18], which deals with differential key-point features in web pages. The plan used similar content descriptions and background contrast histograms to measure the degree of similarity between authentic pages and phishing pages. To determine the similarity between two pictures, a common approach is extracting a distinguished features vector from each image and measuring the distance between vectors. The results show the degree of visual difference between two images. The results suggest that the proposed plan has reached high accuracy and a low rate of error.

Ramesh et al. used distinguishing features of authentic websites to develop a rule-based model for detecting phishing attacks against Internet banking [19]. The proposed model was presented based on two new characteristic sets aiming to determine the relationship rate between page address and page content. Determining the extent of the aforementioned relation is done by using similarity rate algorithms. The resulting output is categorized by vector machine method. The characteristics presented in this study were independent of factors like search engines and list of websites visited by the user. The results of evaluating features through sensitive analysis point to the positive effect of these features on classification output. Rules were extracted by using the decision tree of the developed model. The results of evaluating the rule-based proposed model on a set of fake sites (phishing websites) and authentic ones showed the high accuracy of the model in detecting phishing attacks.

In another study, a multilayer classification model approach was adopted for filtering phishing emails [20]. The authors also adapted a creative method for extracting features of phishing emails based on weight of message content, header of message, and priority ranking. They also reviewed the effect of changing the time of the classification algorithm in a multilayer decision-making process to find desirable planning. The results suggested that the proposed algorithm reduces the frequency of logging into fake websites.

Almomeni et al. proposed a novel method for addressing ambiguity in electronic banking evaluation, detecting phishing websites, and developing a smart, effective, flexible model for detecting phishing of e-banking websites [21]. The proposed model is based on fuzzy logic. It is used besides data-mining algorithms to describe factors affecting the phishing of e-banking websites through classification of types of phishing, determining six types of e-banking phishing and measures of attacking certain websites. The experimental results suggested that URL and different levels of penetration of phishing features play a significant role in the phishing of e-banking entities. Experimental results of this work revealed the importance of website standards in phishing of electronic banking. The URL (identification range) presented by the layer and phishing of various characteristics of influence in the final phishing of electronic banking have an important role.

Wei et al. developed a confronting vector for every transaction based on a sequence of the customer's historical behavior [22]. Through analysis of this behavior, one could detect fake websites. The results of large-scale tests of online banking data suggested that the system could attain high accuracy in dealing with large volumes of data.

Data mining techniques and clustering have been exploited to develop a new method for detecting phishing attacks against bank sites [23]. According to the results, 24 characteristics were analyzed and classified into 4 categories. The websites were compared with these 4 categories to be classified into one of them. Association classification was also used to develop a data mining approach and a rule discovery method was adopted to develop classification systems [24]. According to the results, there are rules for classifying a site into the group of phishing sites or normal ones. It can be done by 12 characteristics.

There also exist several literature surveys that classify different phishing detection methods and investigate their advantages and challenges [25,26].

3. The proposed method

Registered records of phishing attacks in 2012 have been used to analyze the proposed method. This dataset has been gathered from recorded data of Huddersfield University. Preprocessing has been carried out on the dataset and then the file format has been converted into Excel format for use in analysis and simulation. Standard datasets of phishing algorithms are stored as Arff files and they could be recovered and used by specific software. Because the objective is developing a program for testing and evaluation of the suggested method, modification of the structure of the dataset is needed to provide input for the software as a text file.

K-means is one of the highly applicable methods for data mining. The main reason for its applicability is the high flexibility of the method in dealing with records. This privilege could sometimes act against the proposed algorithms because initial centers select those records that take the problem away from attaining its objective. The k-means method can use different relations for its calculations. However, the most conventional relation is Euclidean distance. In our proposed method, we use an efficient method for this calculation, which will be detailed in this section. In the proposed method, a maximum distance matrix is created to measure the distances. In order to choose the best records, the introduced method has used weighting of records as shown in the example of this section. In the k-means algorithm cycle, when we are going to put a record in a cluster, we give it a weight. By presuming small coefficients, we try to diminish the distances between two primary phishing and not-phishing ideas.

For more illustration, we explain our method using an example. We use a selected part of our running dataset as our example's data. Table 1 shows a selected part of the dataset that we are using for describing our methodology. P1 and P2 are two random points assumed for sample processing. In the case of showing samples in a two-dimensional space, we will reach a figure similar to Figure 1. After implementing the algorithm, we have Figure 2. It should be noted that after each implementation, the selected cluster and its members could be different depending on initial centers.

After selection of clusters, the accuracy level of each cluster must be calculated. Accuracy level is calculated as follows:

$$\text{Suspected phishing in cluster} = \frac{\text{Count of phishing sites in cluster}}{\text{Count of all records in cluster}} \times 100$$

In this relation, the probability that a cluster is a phishing one is equal to the total number of phishing records (or sites) in a specific cluster divided by all records of that cluster. If a record is a phishing one in a cluster, it is specified in the "result" field.

To calculate the centers of each cluster, the proposed algorithm uses 31 features (attributes) of sites. These attributes are listed in Table 2. The last attribute is "result", which determines whether the corresponding website of the record is a phishing one or not. The coordinate of the centers of a cluster is calculated as the

Table 1. A random selected part of the dataset for describing the task method.

P1	P2
0.068062	0.075331
0.054554	0.017303
0.274122	0.025457
0.005567	0.05535
0.232608	0.003156
0.264638	0.195716
0.047981	0.027541
0.246464	0.036917
0.236194	0.109573
0.147271	0.025124
0.234608	0.270914
0.661368	0.964927
0.503214	0.751899
0.866753	0.760694
0.717752	0.866821
0.693812	0.919137
0.799963	0.959074
0.578791	0.502346
0.565667	0.505857
0.529279	0.748018
0.719463	0.828892
0.531862	0.545671
0.797031	0.696933
0.661368	0.964927
0.503214	0.751899
0.866753	0.760694
0.717752	0.866821
0.693812	0.919137

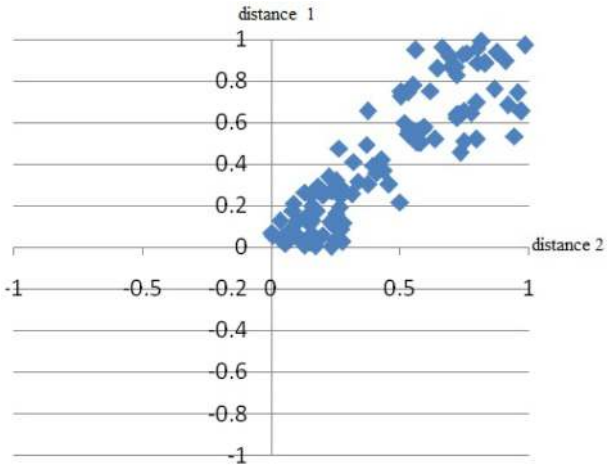

Figure 1. Experimental dots to describe the problem.

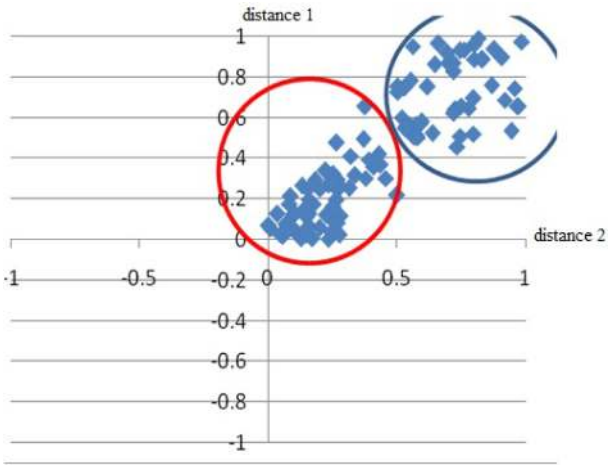

Figure 2. Cluster selected by the algorithm.

average of the values of all points of that cluster for each attribute:

$$\text{Mean of Cluster} = [\text{Average (Attribute\_0)}, \text{Average (Attribute\_1)}, ...,$$

$$\text{Average (Attribute\_30)}, \text{Average (Attribute\_31)}]$$

Finally, we reach a reliable model after all of the above steps are conducted. Now all the suspicious sites can be evaluated using the model and be assigned to the appropriate clusters. This will be explained in the following text.

Table 2. List of the used features of the websites to cluster them.

Attribute	Domain of values
@attribute having_IP_Address	{ –1,1 }
@attribute URL_Length	{ 1,0,–1 }
@attribute Shortining_Service	{ 1,–1 }
@attribute having_At_Symbol	{ 1,–1 }
@attribute double_slash_redirecting	{ 1,–1 }
@attribute Prefix_Suffix	{ 1,–1 }
@attribute having_Sub_Domain	{ 1,0,–1 }
@attribute SSLfinal_State	{ 1,0,–1 }
@attribute Domain_registeration_length	{ 1,–1 }
@attribute Favicon	{ 1,–1 }
@attribute port	{ 1,–1 }
@attribute HTTPS_token	{ 1,–1 }
@attribute Request_URL	{ 1,–1 }
@attribute URL_of_Anchor	{ 1,0,–1 }
@attribute Links_in_tags	{ 1,0,–1 }
@attribute SFH	{ 1,0,–1 }
@attribute Submitting_to_email	{ 1,–1 }
@attribute Abnormal_URL	{ 1,–1 }
@attribute Redirect	{ 1,0 }
@attribute on_mouseover	{ 1,–1 }
@attribute RightClick	{ 1,–1 }
@attribute popUpWidnow	{ 1,–1 }
@attribute Iframe	{ 1,–1 }
@attribute age_of_domain	{ 1,–1 }
@attribute DNSRecord	{ 1,–1 }
@attribute web_traffic	{ 1,0,–1 }
@attribute Page_Rank	{ 1,–1 }
@attribute Google_Index	{ 1,–1 }
@attribute Links_pointing_to_page	{ 1,0,–1 }
@attribute Statistical_report	{ 1,–1 }
@attribute Result	{ 1,–1 }

Assume that the algorithm receives a website as an input to check its situation. First of all, the value of the 30 mentioned attributes for the input website should be specified and the "result" attribute should then be added to the attributes' vector to determine whether the site is a phishing website or not. The Euclidean distances of 30 attributes from the mean of each cluster should then be calculated. Each record will then be assigned the closest cluster using the following relation:

Cluster of the new website= Min(Distance(website,cluster0),Distance(website,cluster1))

In this relation, the function Distance() calculates the Euclidean distance.

In accordance with the same cluster, the probability of phishing status of a website will be assigned to it. Figure 3 is a schematic of this process.

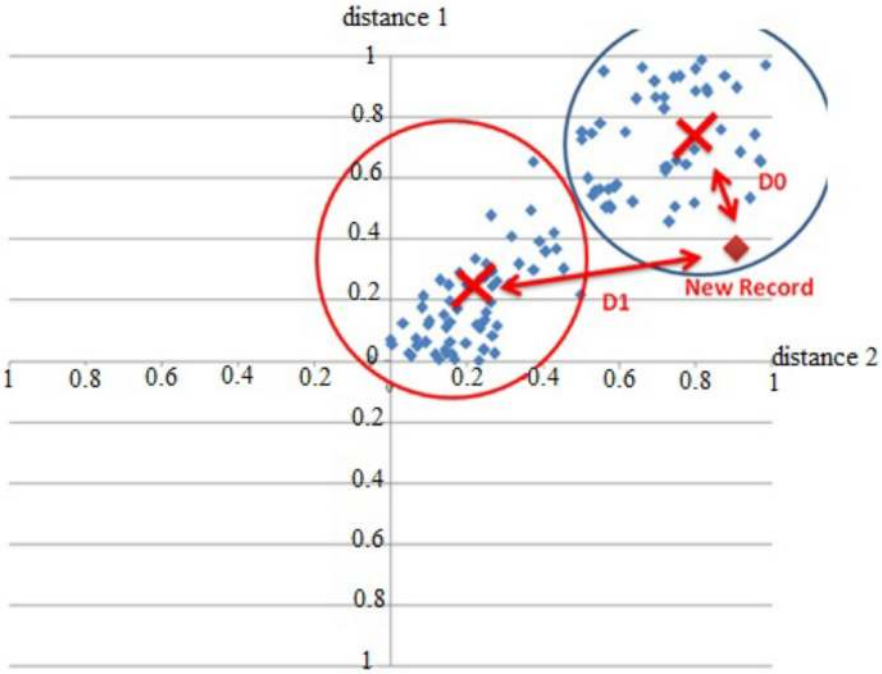

Figure 3. Comparing new website with designed model by the proposed method.

As shown in Figure 3, after entering new record (which shows a website), its distance from the cluster center is calculated and then it is assigned to the closest cluster.

Figure 4 shows the algorithm of the proposed clustering method to identify phishing websites.

In this algorithm, m_1 and m_2 are the means of two clusters, S_1 and S_2, respectively. The first loop of the algorithm of the means of two clusters is such that:

$$(m_1, m_2) = (x_1, x_2) | \forall x_i, x_j \in \text{RecordSet:distance}(x_1, x_2) > \text{distance}(x_i, x_j))$$

Then all records of the RecordSet except m_2 are assigned to cluster S_1 and the number of records of cluster S_1 is calculated as k_1 simultaneously. It is clear that k_2, the number of records of cluster S_2, should be initially set to 1 in this step.

Now, in the main block of the algorithm (While-loop), the weighted distance of each record from the centers of clusters is calculated and the record is assigned to the proper cluster. Remove(x_p,S_i) eliminates record x_p from an inappropriate cluster and Add(x_p,S_i) assigns it to the appropriate cluster. Consequently, x_p is placed in a cluster that has the minimum distance from its center. This process can be formally explained with the following relation:

$$\forall S_i(t) = \{x_P : |x_P - m_i(t)|^2 < |x_P - m_j(t)|^2 \ j, 1 \le j \le 2\}$$

```
Procedure Clustering
Input
  RecordSet : Set of all corresponding records of websites;
Begin
  x1 = the first record of RecordSet;
  x2 = the second record of RecordSet;
  MaxDist = Distance(x1, x2);
  m1 = x1; m2 = x2;
  For each two records x1 and x2 of RecordSet do
  Begin
      If Distance(x1, x2)>MaxDist Then
      Begin
        MaxDist = Distance(x1, x2);
        m1 = x1; m2 = x2;
      End;
  End;
  k1=1;
  For each records x1 of RecordSet do
      If x1 <> m2 Then
      Begin
          Add(x1, S1); Inc(k1)
      End;
  Add(m2, S2); k2=1;
  Flag = True;
  While Flag = True do
  Begin
    Flag = False;
    For each record xp of S1 do
    Begin
      If Distance(xp, m1)>Distance(xp, m2) Then
      Begin
          Remove(xp, S1); Dec(k1);
          Add(xp, P2); Inc(k2);
          Flag = True;
      End;
    End;
    For each record xp of S2 do
    Begin
      If Distance(xp, m1)<Distance(xp, m2) Then
      Begin
          Remove(xp, S2); Dec(k2);
          Add(xp, P1); Inc(k1);
          Flag = True;
      End;
    End;
    S1 - S1∪ P1; S2 - S2∪ P2;
    m1 = (∑∀xi∈S1 xi)/k1
    m2 = (∑∀xi∈S2 xi)/k2
  End;
End;
```

Figure 4. Proposed clustering method to detect phishing websites.

Here, "m" stands for the mean of cluster "S". In the same vein, "t" signifies that the algorithm runs the tth round of the While-loop.

Finally, the mean of the cluster should be recalculated at the end of each round of the While-loop. We have:

$$m_i\,(t+1) = \frac{1}{\left|S_i^{(t)}\right|} \sum_{x_i \in S_i^{(t)}} x_j$$

Here, in the t+1-loop, the mean of the cluster is calculated from the beginning for the next round of the While-loop. It is repeated until nothing happens in the clusters through a complete round of the While-loop.

4. Experimental results

In this section, we explore some characteristics of the proposed system along with mentioned issues in this area:

- Time taken to build a model
- Correctly classified samples
- Incorrectly classified samples
- Prediction accuracy

Table 3 shows the results of comparing the proposed method with other classification and clustering algorithms based on the mentioned evaluation criteria.

Table 3. Comparing the proposed method with classification algorithms.

Evaluation criteria/algorithms	MLP (multilayer perceptron)	J48 (decision tree induction)	NB (Naïve Bayes)	Proposed algorithm
Production time of model (s)	0.87	0.03	0	0.5
Correctly classified instances	194	197	187	198
Incorrectly classified instances	6	3	13	2
Prediction accuracy	97%	98.5%	93.5%	99%

The most important standard for detecting the efficiency of a classification algorithm is "Accuracy". This criterion suggests that the designed classifier has correctly classified a certain percentage of all experimental records. Classification accuracy is calculated using the following relation [27]:

$$\text{Accuracy} = \frac{\text{The number of records that are classified correctly}}{\text{The total number of records}}$$

The second criterion used for classification efficiency is "Recall". This criterion is separately calculated for values of each class. The Recall criterion signifies the ratio of record numbers that are correctly classified to all the records of the same category that are wrongly placed in other categories. The following equation shows the way to calculate this criterion [27]:

$$\text{Recall}^{x} = \frac{\text{The number of records that were correctly classified as x}}{\text{The number of records of category x that are wrongly categorized}}$$

The third criterion that is used in evaluation of classification algorithms is "Precision". Similar to Recall, this criterion is calculated for values of each class separately. The Precision criterion shows the accuracy of the considered class according to all cases that have been proposed for this class. This criterion is calculated according to the following relation [27]:

$$\text{Precision}^{x} = \frac{\text{The number of records that were predicted correctly for category x}}{\text{The total number of records that were predicted for category x}}$$

Table 4. Comparing the proposed method with some of the best phishing detection algorithms.

		Class accuracy	F-measure	Recall^x	Precision^x
Byzantine method	Phishing	93.5%	92%	84.3%	93.5%
	Not phishing		95%	76.1%	81.2%
MLP method	Phishing	97%	97%	90%	97%
	Not phishing		92%	74%	86.3%
Fuzzy method	Phishing	81%	97%	77%	96%
	Not phishing		95%	79%	97%
Proposed method	Phishing	99%	97%	98%	99%
	Not phishing		96%	98%	98%

According to the evaluation criteria mentioned in this section, we propose the results of three common methods, namely fuzzy, byzantine, and MLP, and the proposed method in Table 4.

The numbers generated in this step belong to relevant papers reviewed. After reviewing the results in the simulation section of our proposed method, one could claim that it can detect phishing attacks with 99% accuracy.

In simpler terms, we can describe the difference between precision and accuracy in the following manner. Precision only depends on the distribution of random errors and it has no association with real values or specified amounts, while accuracy is described in a diagonal way. It is the total systematic error that might be composed of one or more systematic error items. It is noteworthy that diagonal largeness shows a high difference from the accepted reference value. Precision (diagonal) does not include random error.

5. Conclusion

In this paper, we have proposed a new detection method for phishing websites using a clustering approach. A weighted version of Euclidean distance has been presented to improve the performance of clustering. Correcting the membership of records in the clusters using weights has led to very good results that are comparable to the results of classification approaches. The proposed method has used 30 important features of websites to determine whether they are phishing ones or not. Experiments have been carried out on the dataset of Huddersfield University. The results of implementation of the work have been evaluated and have been compared with other supervised classification methods such as the decision tree and artificial neural networks. Experimental results showed that the proposed method is better than other classification and clustering algorithms in terms of accuracy.

References

[1] Shang S, Holbrook M, Kumara P, Carnot LF, Downs J. Who falls for phish?: A demographic analysis of phishing susceptibility and effectiveness of interventions. In: Proceedings of the 28th International Conference on Human Factors in Computing Systems; 2010. New York, NY, USA: ACM. pp. 373-382.

[2] Alarifi A, Alsaleh M, Alomar M. A model for evaluating the security and usability of e-banking platforms. Computing 2017; 99: 519-535.

[3] Safeena R, Kammani A, Date H. Assessment of Internet banking adoption: an empirical analysis. Arab J Sci Eng 2014; 39: 837-849.

[4] Yu WD, Nargundkar S, Tiruthani N. A phishing vulnerability analysis of web based systems. In: Proceedings of the 13th IEEE Symposium on Computers and Communications; 2008; Morocco. New York, NY, USA: IEEE. pp. 326-331.

[5] Moreno-Fernandez MM, Blanco F, Garaizar P, Matute H. Fishing for phishers. Improving Internet users' sensitivity to visual deception cues to prevent electronic fraud. Comput Hum Behav 2017; 69: 421-436.

[6] Ali MM, Rajamani L. Deceptive phishing detection system: from audio and text messages in instant messengers using data mining approach. In: Proceedings of the IEEE International Conference on Pattern Recognition, Informatics and Medical Engineering; 21–23 March 2012; Salem, India. New York, NY, USA: IEEE. pp. 458-463.

[7] Sohrabi MK, Akbari S. A comprehensive study on the effects of using data mining techniques to predict tie strength. Comput Hum Behav 2016; 60: 534-541.

[8] Sohrabi MK, Barforoush AA. Efficient colossal pattern mining in high dimensional datasets. Knowl-Based Syst 2012; 33: 41-52.

[9] Sohrabi MK, Barforoush AA. Parallel frequent itemset mining using systolic arrays. Knowl-Based Syst 2013; 37: 462-471.

[10] Sohrabi MK, Roshani R. Frequent pattern mining using cellular learning automata. Comput Hum Behav 2017; 68: 244-253.

[11] Sohrabi MK, Ghods V. Top-down vertical itemset mining. In: SPIE 2014 International Conference on Graphic and Image Processing; 24–26 October 2014; Beijing, China. Bellingham, WA, USA: SPIE. pp. 1-7.

[12] Sohrabi MK, Azgomi H. Parallel set similarity join on big data based on locality-sensitive hashing. Sci Comput Program 2017; 145: 1-12.

[13] Sohrabi MK, Marzooni HH. Association rule mining using new FP-linked list algorithm. Journal of Advances in Computer Research 2016; 7: 23-34.

[14] Sohrabi MK, Azgomi H. TSGV: A table-like structure based greedy method for materialized view selection in data warehouse. Turk J Electr Eng Co 2017; 25: 3175-3187.

[15] Sohrabi MK, Ghods V. Materialized view selection for a data warehouse using frequent itemset mining. Journal of Computers 2016; 11: 140-148.

[16] Ali MM, Rajmani L. APD: ARM deceptive phishing detector system phishing detection in instant messengers using data mining approach. In: Proceedings of 4th International Conference on Global Trends in Computing and Communication Systems; 2011. Berlin, Germany: Springer. pp. 490-502.

[17] Mohammad RM, Thabtah F, McCluskey L. Predicting phishing websites based on self-structuring neural network. Neural Comput Appl 2014; 25: 443-458.

[18] Rajalingam M, Alomari SA, Sumari P. Prevention of phishing attacks based on discriminative key point features of webpages. International Journal of Computer Science and Security 2012; 6: 324-332.

[19] Ramesh G, Krishnamurthi I, Kumar K. An efficacious method for detecting phishing web pages through target domain identification. Decis Support Syst 2014; 61: 12-22.

[20] Islam R, Abawajy J. A multi-tier phishing detection and filtering approach. J Netw Comput Appl 2013; 36: 324-335.

[21] Almomani A. Evolving fuzzy neural network for phishing emails detection. J Comput Sci 2012; 8: 1099-1107.

[22] Wei W, Li J, Cao L, Ou J, Chen J. Effective detection of sophisticated online banking fraud on extremely imbalanced data. World Wide Web 2013; 16: 449-475.

[23] Ajlouni M, Hadi W, Alwedyan J. Detecting phishing websites using associative classification. Journal of Information Engineering and Applications 2013; 5: 1899-1905.

[24] Abdelhamid N, Ayesh A, Thabtah F. Phishing detection based on associative classification data mining. Expert Syst Appl 2014; 41: 5948-5959.

[25] Khonji M, Iraqi Y. Phishing detection: a literature survey. IEEE Commun Surv Tut 2013; 15: 24-40.

[26] Khorshed T, Ali A, Wasimi SA. A survey on gaps, threat remediation challenges and some thoughts for proactive attack detection in cloud computing. Future Gener Comp Sy 2012; 28: 833-851.

[27] Va SL, Vijaya MS. Efficient prediction of phishing websites using supervised learning algorithms. Procedia Engineering 2012; 30: 798-805.