An Improved Image Steganalysis Using a Novel Feature Selection Algorithm Based on Artificial Bee Colony

F. Ghareh Mohammadi 1 and M. Saniee Abadeh2

Faculty of Electrical and Computer Engineering, Tarbiat Modares University
Tehran, 14115-143, Iran;
1Email: f.garemohammadi@modares.ac.ir  2Email: saniee@modares.ac.ir

ABSTRACT: One of the most important phases of pre-processing is Feature selection, which can improve the predictive accuracy of steganalysis. In this study, we have presented a novel feature selection-based method to image steganalysis for detecting stego images from cover images based on artificial bee colony (ISBC). The experiments show that the proposed method is easy to be employed for steganalysis purposes. Moreover, its overall performance is better than recent ABC-based feature selection methods.

1 INTRODUCTION

One of the most significant current emerging branches in information forensics refers to the development of steganalysis performance. In this paper, we employed a new feature selection algorithm (ISBC) to develop it. Alternatively, we have to transfer the most significant information via a safe way somehow; Several methods are proposed to improve the steganalysis performance with growing feature space to be a good steganalysis approach that takes a long time to finish.(V. Holub et al., 2013)(J. Kodovsky & J. Fridrich, 2012). This paper proposes a new approach to lessen the time and improved detection precision of image steganalysis using bee colony based feature selection (ISBC).

Steganography is a dynamic tool with a widespread history and the capability to adjust to novel stages of technology. Steganography is the skill of embedding covert data into cover images. Except sender and the receiver, anybody does not know the existence of the covert messages, in result kept the important information from unauthorized or unexpected viewing (Geetha & Kamaraj, 2010). If there is a method that can be estimated sent images consists of a secret message with a high rate better than accidental guesstimate, With respect to, the steganographic system is considered as broken. For more perfect behavior of the concept of steganographic safety, the person who reads is referred to (Anderson, R. J. & Petitclos 1998).

Steganalysis includes noticing the utilization of steganography within a file with little or no information regarding the steganography methods or/ and the components. Steganographic methods occasionally leave a sign in the file that is encrypted. With this information, the presence of covert messages can be noticed. It is fair to describe steganalysis is both a knowledge and an art. The knowledge assists in forming the experiments covert messages, and the art of steganalysis has a key role in the collection of features to experiment themselves. (S. Geetha, Dr. N. Kamaraj 2010). The main purpose of steganalysis is to identify the existence of covert information in an object. Multimedia files such as images, videos and audios are perfect covers of object to steganography methods. So, the use of steganalysis is as what declared in multimedia. (Ghareh Mohammadi & Saniee Abadeh, 2012)

The rest of the paper is organized as follows: In section 2, we proposed the artificial bee colony-based feature selection algorithm (ISBC). Section 3 will discuss the experimental determined results. In the last section of the paper some conclusions have been discussed.
2 PROPOSED METHOD (ISBC)

This section presents a new method to resolve feature selection problems in steganalysis, which the paper proposes two subsections. Section 2.1 refers to a schematic method to find the best subset features. Section 2.2 briefly explains the structure of presenting improved artificial bee colony based feature selection.

The General structure of presenting a method contains three important steps. These steps are phases of ISBC respectively feature extractor, artificial bee colony as a feature selection, support vector machine. This structure illustrates how to employ ISBC, and first of all, it is needed to provide a feature vectors’ dataset extracted from a lot of images. For this issue we used a feature extractor (SPAM) to solve that issue, we will discuss it in the following section. After the feature extractor step, next we need to provide a method to select relevant features using ABC and compute fitness using SVM to evaluate the selected subset features and the outcome comes to ABC to validate the result, next if the condition is met or pre-determined number of cycles is reached, the process will finish.

2.1 Structure of proposed feature selection approach

Generally, a particular feature selection algorithm surrounds four factors: a subset generation, a fitness function, a stopping condition in ABC, and a validation process.

2.2 A novel artificial bee colony algorithm for feature selection (ISBC)

In the presented ABC-based feature selection method, ABC algorithm improves the procedure of feature selection and generates the ideal feature subset that improves the performance of the classifier. Figure 1 reveals the overall pseudo code of implementing feature selection using ABC. ABC is used as a feature selector and produces the feature subsets and a classifier is employed to evaluate every feature subset built by the onlookers; so, the presented approach is a kind of wrapper based systems that we are going to explain more as follows:

2.2.1 Initial population

In this paper, the ISBC is used to explore the new search space. Initial swarm is sometimes created randomly. We are going to set parameters, ABC parameters contain the number of food source, the number of colony size, lower bound, upper bound, limit, max cycle. The population of employed bees and onlooker bees are equal to the dimension of features. After that, we use one of the best classifier Support Vector Machine (SVM) to evaluate the predictive accuracy of selected feature to obtain the discriminating ability of every single feature in the dataset. The accuracy $(x_i)$ of per feature $i$ is computed by using 10-fold cross validation. After that, the objective $(F_i)$ is computed for per feature from its unknown and implicit relation.

2.2.2 Employed bees

In this process of artificial bee colony algorithm, each employed bee obtains a food source in the neighborhood of its recently selected food source and determines the accuracy of its nectar fitness. In other words, each feature in the food source is developed based on the process of keeping informed feasible solution by employing bees as demonstrated in equation (1).

$$J=\text{(int)}\left(\text{Rand. Next double}\right)\ast N$$

$$V_i = f_i + o\left(f_i \cdot f_j\right)$$

(1)

Where, $i=\{0,1,2,\ldots, N\}$ and $j=\{0,1,2,\ldots, N\}$ and $N$= upper bound of features and $f_i$ is the performance of the feature assigned to the employed bee and $f_j$ is the performance of the feature the onlooker has chosen. $o$ is a real random number in the range $[0, 1]$.

A feature vector is a binary bit string of dimension equal to the amount of features is given to each employed bee to act for its feature selection. Inside it, bit ‘1’ figures out the feature is a selected feature and bit ‘0’ figures out the feature is not selected.

Then, every employed bee is assigned to each solution and evaluates the fitness of each solu-
tion by using equation (2). The novel candidate solution is then evaluated by means of a support vector machine (SVM) classifier. If the novel fitness is better than the current one, the implied bees will substitute their solution with this novel candidate solution; if the novel fitness is worse than the current one, the novel candidate will not consider and will be ignored.

$$\text{Fitness} (R) = \frac{1}{1+ f_r}$$

(2)

2.2.3 Onlooker bees process

After finishing employed bees share information about their solution, the onlooker bees start to get information from the employed bees and choose a good food source to visit according to the probability of each food source using equation (3) and make the process of updating feasible solution based on equation (4).

$$P(R) = \frac{\text{fitness}(R)}{\sum_{s=1}^{m} \text{fitness}(s)}$$

(3)

After that the onlooker computes the new solution $V_i$ using the accuracies of the feature the employed bee is trading and the feature the onlooker bee has chosen. If the new solution $V_i$ is lesser than $f_i$, the employed bee will be toward feature subset including the feature which was earlier pointed to and the newly chosen feature. If the $V_i$ is greater than $V_i$, the employed bees feature will be kept and the newly selected feature would not be considered. The novel solution $V_i$ is calculated by using equation (4).

This approach, every time the employed bee is allocated a novel feature subset, the onlooker bee exploits and attempts to produce a new feature subset arrangement. After many feasible features having been exploited for starting the feature subset, the amount of nectar gets accumulated points to better feature subset arrangement. If none of employed bees have improved, after that the employed bee convert to a scout bee. The scout bee is allocated to a new binary feature subset consisting in the equation (4). The late solution in the onlooker bee’s memory will be placed in the novel candidate solution if the new solution’s fitness is better than the current one.

$$FS = X_{min} + \omega (X_{max} - X_{min})$$

(4)

Where $X_{max}$ and $X_{min}$ represent the upper and lower bounds of the number of population and $\omega$ is a real random number in the range [-1,1]. The bees keep countinuing the identical procedure till the best feature subset is made.

2.2.4 Scout bees process

In this phase, if the fitness value of the current food source has not been improved by predefined number of iterations, named the “limit”, the food source will be left. Then, the scout bees will randomly produce a new food source position in all feature dimensions. This process is applied to avoid choosing the sub-optimal solution.

3 EXPERIMENTAL RESULTS

To reveal the performance of the proposed feature selection algorithm and to compare it with other well-known feature selection approaches, one set of experiments was carried out research in steganalysis. For experimental studies, we have employed the breaking out steganography system (BOSS) version 1.01 gray scale image databases that the rate of hidden text embedding is 0.4 per pixel. This database contains 1250 cover images and 1250 stego images. So, the number of classes in our experiments is binary.
3.1 The spam features

In this paper, we applied the Subtractive Pixel Adjacency Model (SPAM) method (Pevny et al., 2010) that is used to produce the features for steganalysis. The length of SPAM’s feature vector has 686 attributes and one class attribute.

3.2 Set parameters

Artificial bee colony (ABC) has a less parameter than other algorithms like Ant colony optimization (ACO) which gives it a competitive edge over the other swarm intelligence algorithm. The parameters initializations of ISBC are set as follows: the number of food source = Number of feature in the data set, the number of colony size = 2* Number of features in the dataset, lower bound = 1, upper bound = N, limit = 80, max cycle = 20.

3.3 Evaluation of proposed method

The results of this study are given in Table 1. It shows the best average detection accuracy of each approach. It is obvious that the high precision of image steganalysis identity is related to the proposed ISBC method.
Table 1. Detection accuracy comparison of other methods versus proposed method ISBC

<table>
<thead>
<tr>
<th>Factors/approaches</th>
<th>ABC-j48</th>
<th>(ABC-j48graft)</th>
<th>Proposed (ABC-SVM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>No of selected feature</td>
<td>80</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>Selected feature accuracy</td>
<td>51.02%</td>
<td>51.15%</td>
<td>53.32%</td>
</tr>
<tr>
<td>Primary accuracy</td>
<td>46.08%</td>
<td>46.04%</td>
<td>52.32%</td>
</tr>
<tr>
<td>No of primary feature</td>
<td>686</td>
<td>686</td>
<td>686</td>
</tr>
<tr>
<td>%Reduction in feature</td>
<td>%88.33%</td>
<td>%88.33%</td>
<td>%88.33</td>
</tr>
<tr>
<td>No. Of Instance</td>
<td></td>
<td></td>
<td>2500 images (BOSS-Hugo)</td>
</tr>
<tr>
<td>%increase in accuracy</td>
<td>4.94% increase</td>
<td>5.11% increase</td>
<td>1% increase</td>
</tr>
</tbody>
</table>

Figure 2. Comparison of the classification accuracy of the difference steganalysis method.

4 CONCLUSION

In this paper, an improved Image steganalysis feature selection algorithm is presented based on artificial bee colony (ISBC). It is very obvious that the presented method performs frequently superior to the other ABC-based and other well-known feature selection approaches with the equal number of feature subsets. Performance evaluation of suggested methods also shows that the presented method is more efficient than the other methods. This study makes known that the proposed ABC-based feature selection (ISBC) approach improves the classification accuracy of image steganalysis with lesser time.
REFERENCES


Ghareh Mohammadi, F, Saniee Abadeh, M, 2012. A Survey of Data Mining Techniques for Steganalysis. RECENT ADVANCES IN STEGANOGRAPHY.


