Scope of Support Vector Machine in Steganography

Rohit Tanwar, Sona Malhotra
Ph.D Scholar, Computer Engg, UIET Kurukshethra University
Asst. Prof., Computer Engg, UIET Kurukshethra University

Abstract: Steganography is a technique used for secure transmission of data. Using audio as a cover file opens path for many extra features. In order to overcome the limitations of conventional LSB technique, various variants were proposed by different authors. In order to achieve robustness, use of various optimization techniques has been tradition. In this paper the focus is put on use of Genetic Algorithm and Particle Swarm Intelligence in steganography. To list detailed scope, merits and de-merits of the two optimization techniques is the main constituent of this paper. In spite of analyzing the two techniques, the motivation and applicability of machine learning algorithm in the problem statement is also discussed. This paper will guide the path in using Support Vector Machine for optimizing the data hiding.

Keywords: Machine Learning; Support Vector Machine; Robustness; GA;

1. INTRODUCTION
Steganography has proven its use in securing information. It can be classified on the basis of the cover file used for hiding secret message. The major literature is studied focusing on audio as cover file. Many researchers has combined cryptography along with steganography so as to provide double layer of security. Many other researchers worked on making the position of bit replacement random so as to make it more robust. Embedding secret messages in digital sound seems to be more difficult as compared to other binary files because of higher sensitivity of Human Auditory System (HAS) as compared to Human Visual System (HVS) [10].

Fig. 1 Types of Steganography
A steganography technique is said to be good one if it poses some basic properties which are represented as a magic triangle:

\[ \text{Transparency} \quad \text{Data Rate} \quad \text{Robustness} \]

Fig. 2: Magic Triangle

The basic requirements of steganography can be described as follows:

- **Transparency:**
  It is the property by which the stego file does not create any suspicion of presence of any embedded message while compared to original cover file i.e. cover object and stego object must be perceptually indiscernible [7].

- **Data Rate:**
  It is defined as the number of bits of secret message that can be hidden in a byte of cover image. It is required for an application to have high data rate of embedding.

- **Robustness:**
  A technique is said to be highly robust if it can withstand against various attacks. The attacks can be of type intentional and un-intentional[8].

### 2. MOTIVATION

Let us re-describe some basic properties of steganography so as to describe the workflow.

Imperceptibility is a measure of the distortion produced in stego audio due to modifications like message embedding. It is essential to keep the distortion at a level below than the threshold which can be estimated on the basis of HAS/HVS and the cover file[1]. Since it is not possible practically to get a technique having high capacity and high robustness simultaneous so the attention moves towards balancing imperceptibility and robustness.
Various researchers have done their work in achieving robustness by maintaining imperceptibility. In order to minimize the distortion induced as a result of embedding process, the researchers put their attempts further to optimize the embedding process by using different optimization techniques.

In this paper the focus is to summarize the applications and aftereffects of genetic algorithm and particle swarm optimization. Both the techniques will then be compared with the achievements of machine learning so as to justify the scope of machine learning. Due to its uniqueness in giving solutions and less time consumption, Support Vector Machine is now being used as a preferred choice of machine learning.

### 3. Literature Review

Majdak in his Ph.D work explored the scope of balancing robustness through imperceptibility. He discovered the various limitations of traditional substitution method and provided a path for overcoming them. His solution was based on GA. His work can be concluded as follows[4]:

- Audio file chosen in WAV format.
- Increased capacity while maintain imperceptibility.
- To maintain imperceptibility, GA was used as an optimization tool.
- Quality was measured and compared in PSNR metric.

Krishna Bowal et.al. combined cryptography, steganography and Genetic Algorithm to develop a 3-tier protection tool[1]. The secret message is encrypted using RSA and then embedded in cover audio using GA based technique. Their work can be concluded as follows:

- Along with increasing depth of embedding layer, choice of random layer was also done randomly.
- 2-way robustness.
- Multi-objective GA was used to optimize position and distortion.
- Data Rate is comparatively low.
- Imperceptibility measured by subjective audio tests.
- Steganalysis is supposed to be more challenging.

V.Santhi and Longeswari Govindaraju in their work of providing 3-tier security using Cryptography and GA along with steganography did following variations[12]:

- Negative bytes can be included in population generation.
- Negative audio byte processed as 32-bit audio sample while positive as 16-bit audio sample.
- Resulting tool ensured large capacity and drastic reduction in noise.

Rishidas.S et.al. in their work analyzed SVM as a classifier in steganalysis. However the object for classification was images and even images of different formats but still their work provided the measure of strength of SVM as classifier in data hiding or extracting[3].
75 images in each class and 60% of this is used for training SVM.

- SVM trained using different kernels
- GIF format images are easily identifiable as compared to bitmap images.

Feno Heriniaina and Jung Sang in their work used Particle Swarm Optimization in substitution technique[5].

- Cover file first converted to frequency domain.
- Secret message converted into partitions and optimal matrix substitution was done as a part of embedding process.
- PSO was used to find optimal matrix for substitution

Initial work was done on image steganography where text message was hidden into a digital image. Later on the effort to hide image in image and audio into image also succeeded. But the drawback of using image as cover file was due to higher sensitivity of human visual system. It was easy to catch a little distortion introduced in an image by naked eyes and this would arouse suspicion violating the objective of steganography.

The most famous technique to implement steganography was substitution technique in which some of the bits of cover file were substituted with the message bits. The LSB’s were chosen for substitution as they contribute least in computation of statistical value of the cover file.

In order to increase the data hiding capacity of cover files, multiple LSB’s got substituted. However that type of method resulted in more distortion produced. Since substitution with LSB’s is very susceptible to intentional as well as unintentional attacks thus solution was proposed to hide data in deeper layers. Hiding data in deeper layers of cover led to more distortion in cover file.

Some algorithms were proposed then which not only hide data in deeper layers but also do adjustments by willingly modifying certain other bits of cover file in order to minimize the distortion thus occurred.

Due to large availability in practical and its inherent masking effect, audio started to begin as a choice for cover file. More over Human Auditory System suppressed low frequency audio with higher frequency audio; called masking effect. Even after the use of audio as cover the inherent problems of substitution technique remained as it is. Thus lots of variants of conventional LSB method came into practice with an attempt to be more robust.

Another layer of security was introduced by combining cryptography with steganography. Thus message to be hidden in cover was encrypted first and then hidden in cover audio. It was the time then that researchers started combining soft computing techniques with steganography. However the combination of a particular soft computing technique was there to achieve some specific goal.
In order to develop optimized technique, GA became the first choice to be combined with steganography. However the different techniques differ in their choice of fitness function and the logic of applying GA. GA is good optimization tool but it is less effective when generalization is required. SVM is proven to be a good classifier algorithm in such cases.

4. DIFFERENT OPTIMIZATION TECHNIQUES

4.1 Genetic Algorithm:
Advantages:
- Concept easy to understand
- Intrinsically parallel
- Always an answer; answer gets better with time
- Modular-separate from application
- Chances of getting optimal solution are more

Disadvantages:
- No guarantee of finding global maxima
- Time Taken for convergence
- Calculation of parameters is trial-error based
- unguided mutation

4.2 Particle Swarm Optimization
Advantages:
- Can be applied into both scientific and engineering research
- Only most optimist particle transmit information, thus fast
- Parameter calculation is simple
- Lower computational cost.

![Optimization Techniques](image)

Fig.4. Optimization Techniques[9]
Disadvantages:

- Less exact at the regulation of speed and direction
- Cannot work out on the problems of scattering
- Cannot work out on the problems of non-coordinate system such as solution to the energy field
- Not exactly repeatable in terms of computational cost, making comparison hard

Voratas Kachitvichyanukul did comparison of following evolutionary algorithms and listed as follows:

TABLE 1: GA and PSO Comparison

| Parameters for Comparison                          | Genetic Algorithm | Particle Swarm Optimization |
|-----------------------------------------------------|-------------------|-----------------------------|
| Require ranking of solutions                        | Yes               | No                          |
| Influence of Population size on solution time       | Exponential       | Linear                      |
| Influence of best solution on population            | Medium            | Most                        |
| Ability to reach good solution without local search | Less              | More                        |
| Homogeneous sub grouping improve convergence        | Yes               | Yes                         |
| Tendency for premature convergence                  | Medium            | High                        |

5. Support Vector Machine

SVM is a regulated machine learning calculation which can be utilized for characterization or relapse issues. It utilizes a procedure called the bit trap to change your information and afterward taking into account these changes it finds an ideal limit between the conceivable yields. Basically, it does some to a great degree complex information changes, then makes sense of how to isolate your information taking into account the names or yields you’ve characterized.

5.1 Advantages of using SVM:

- Suitable for linearly separable as well as non separable data sets.
- Gives unique solution in less time
- No human expertise required in kernel transformation
- More suitable when irregular distributed data.
- Absence of local minima
- Memory efficient while training
6. CONCLUSION

Vector Machines may be used to provide faster and unique results. The technique can be used in overcoming the limitations in traditional substitution technique. The paper may guide so as to develop a technique which is highly reliable, robust and having good data hiding capacity by the inclusion of Support Vector Machine in substitution method.

The work can be extended further towards embedding in higher layer LSB’s. The distortion induced may be adjusted by designing and training appropriate SVM. The future work may be focused on combining various evolutionary algorithms with SVM so as to improve the classification task.

References

[1] Krishna Bhowal, Debnath Bhattacharya, Anindya Jyoti Pal and Tai-Hoon Kim, “A GA Based audio steganography with enhanced security,” Telecommunication Systems, Springer, ISSN: 2197-2204, July 2011.
[2] V.Santhy and Longeswari Govindaraju, “Stego Audio using Genetic Algorithm Approach,” in Research Journal of Applied Sciences Engineering and Technology, pp. 4806-4812, ISSN:2040-7459, Feb 2014.
[3] Rishidas.S, Gayathri Krishnan L and Sujith Kumar T P, “A Comparative Study of steganalysis using Support Vector Machines on Different Image Formats,” in International Journal of Engineering Research and Technology, pp. 762-764, ISSN:2278-0181, March 2015.
[4] Mazdak, Z., A.M. Azizah, B.A. Rabiah and M.Z.Akram and M.Zeki, “A genetic algorithm based approach for audio steganography,” World Academic Science. Eng. Technol., pp. 360-363, 2009.
[5] Feno Heriniaina and Jung Sang, “Using PSO Algoirthm for Simple LSB Substitution Based Steganography Scheme in DCT Transformation domain,” Springer Berlin Heidelberg, ISSN: 0302-9743, June 2011.
[6] Parisa Gerami, Subaria Ibrahim and Morteza, ”Least Significant Image steganography using Particle Swarm Optimization and Optical Pixel Adjustment”, International Journal of Computer Applications, ISSN:0975-8887, October 2012.

[7] Petitcolas, F.A.P, Anderson, R.J., & Kuhn, M.G. , ”Information Hiding-A Survey”, Proceedings of IEEE, ISSN: 1062-1078

[8] Lee,Y.K., & Chen, L.H.(2000), ”High Capacity Image Steganographic model”, IEEE Proceedings vision, image and signal processing, pp: 288-194, 2000.

[9] Voratas Kachitvichyanukul, ”Comparison of Three Evolutionary Algorithms: GA,PSO and DE”, Industrial Engineering and Management Systems, pp.215-223, ISSN:1598-7248, Sep 2012.