International Journal of Digital Crime and Forensics
Volume 11 • Issue 1 • January-March 2019

Image Steganalysis in High-Dimensional Feature Spaces with Proximal Support Vector Machine

Ping Zhong, College of Science, China Agricultural University, Beijing, China

Mengdi Li, College of Information and Electrical Engineering, China Agricultural University, Beijing, China

Kai Mu, College of Information and Electrical Engineering, China Agricultural University, Beijing, China

Juan Wen, College of Information and Electrical Engineering, China Agricultural University, Beijing, China

Yiming Xue, College of Information and Electrical Engineering, China Agricultural University, Beijing, China

ABSTRACT

This article presents the linear Proximal Support Vector Machine (PSVM) to the image steganalysis, and further generates a very efficient method called PSVM-LSMR through implementing PSVM by the state-of-the-art optimization method Least Square Minimum-Residual (LSMR). Also, motivated by extreme learning machine (ELM), a nonlinear algorithm PSVM-ELM is proposed for the image steganalysis. It is shown by the experiments with the wide stego schemes and rich steganalysis feature sets in both the spatial and JPEG domains that the PSVM can achieve comparable performance with Fisher Linear Discriminant (FLD) and ridge regression, and its computational time is far more less than that of them on large feature sets. The PSVM-LSMR is comparable to Ridge Regression implemented by LSMR (RR-LSMR), and both of them require the least computational time among all the competitions when dealing with medium or large feature sets. The nonlinear PSVM-ELM performs comparably or even better than FLD and ridge regression for the spatial domain steganographic schemes, and its computational time is apparently less than that of them on large feature sets.

KEYWORDS

Extreme Learning Machine Kernel Matrix, Steganalysis, Steganography, Proximal Support Vector Machine

1. INTRODUCTION

Image steganography is an important convert communication technology that conceals secret messages in images by the means of slight changes in pixel values or DCT coefficients. Currently, the most secure steganographic algorithms are content-adaptive ones, such as HUGO (Pevný et al., 2010), UNIWARD (Holub & Fridrich, 2014), WOW (Holub & Fridrich, 2012) and so on. They tend to hide the secret data in the complicated texture regions and show the excellence anti-detection ability.

DOI: 10.4018/IJDCF.2019010106

With the rapid development of image steganography, steganalysis techniques that are related to detecting the existence of the hidden messages in images also have made great progress. The popular methods consist of extracting the relevant features that help to detect the presence of hidden message, and then designing suitable classifier to separate the classes of cover and stego images. In order to improve the detection performance, the feature dimensions are ever-increasing. As is well known, the state-of-the-art steganalysts are the Spatial Rich Model (SRM) (Fridrich & Kodovský, 2012) and its variants (Holub & Fridrich, 2013; Denemark et al., 2014), which may contain more than 30,000 features. For the large scale and high dimensional training sets, it has been shown that ensemble classifiers, such as FLD ensemble classifier, (Kodovský et al., 2012) are successful. Later, Fridrich et al. (Cogranne et al., 2015) demonstrate that a simple well regularized FLD or a ridge regression can achieve the comparable performance with the ensemble classifiers. In addition, they show that ridge regression implemented by LSMR achieves almost the same detection accuracy as an ensemble classifier for a computational time up to 10 times smaller.

However, for the widely popular linear Support Vector Machine (SVM) and Gaussian SVM, they are difficult to be trained after the presence of rich media models (Kodovský et al., 2012). Compared with the FLD and ridge regression, the major reason for the difficulty in training these standard SVMs is that they require considerably long computational time to solve a linear or a quadratic program involved. In addition, except the regularization parameter which these machine learning algorithms have, Gaussian SVM also needs to search for another optimal kernel parameter in the training process, and it is time consuming.

Quite different from the standard SVMs, the linear Proximal Support Vector Machine (PSVM) (Fung & Mangasarian, 2001) which has been proposed based on the much more generic regularization networks (Evgeniou et al., 2000) can be fast implemented without of extensive computation. The linear PSVM separates two classes of data points through proximal hyperplanes with the maximum margin. The strong convexity of the formulation leads to the simple proximal code, which is not always the case in the standard SVMs. Motivated by the PSVM, a simplified nonlinear method referred to Extreme Learning Machine (ELM) (Huang et al., 2012) has been presented for learning single hidden layer feed forward neural networks. ELM has the ability of dealing with the nonlinear feature construction by ELM kernel matrix without the selection of parameters (Huang et al., 2015).

All the methods including the regularized FLD, ridge regression, and the linear PSVM can be interpreted as the least square estimations with l_2 regularization (the regularized FLD corresponds to the ridge regression when the features have zero mean (Cogranne et al., 2015)). The fact that a simple FLD classifier or a ridge regression can achieve good detection accuracy (Cogranne et al., 2015) motivates this paper. It is reasonable to study other regularization methods and make a comparison among their performance. In particular, the regularized FLD and ridge regression get their solutions by inverting a matrix of size $d \times d$ (where d is the number of features), while the linear PSVM can get the solution by inverting an equal or even smaller size of matrix. The potential benefit is reducing training complexity and improving the efficiency. Furthermore, the linear PSVM can be implemented by the fast optimization method LSMR (Fong & Saunders, 2011) to substantially improve the computational efficiency. In addition, we propose a nonlinear PSVM, called PSVM-ELM, based on the ELM kernel matrix by combining the merits of the linear PSVM and ELM. The experiments show that the detection accuracy of the linear PSVM is comparable to that of the FLD and ridge regression, and its computational time is far less than that of the FLD and ridge regression when dealing with the large feature sets. The PSVM-LSMR can further improve the PSVM on both the detection accuracy and computational time, and the performance of it is comparable to that of the RR-LSMR. Both of them require the least computational time among all the competitions when dealing with medium or large feature sets. In addition, it is shown that the detection accuracy of the nonlinear classifier PSVM-ELM is rather good for the spatial domain steganographic schemes. Also, the PSVM-ELM is trained much more efficiently than the FLD and ridge regression on the large feature sets.