An Image Steganalysis Method Based on Evidence-Theoretic KNN with a Small Sized Training Set

Weiping ZOU[1,2], Anne Sophie CAPELLE-LAIZÉ[1], Philippe CARRÉ[1], Nanrun ZHOU[2], Jianhua WU[2]

[1]XLIM-SIC Laboratory, UMR CNRS 7252, University of Poitiers
BP 30179, 86962 Futuroscope Chasseneuil CEDEX, France

[2]Department of Electronic Information Engineering, Nanchang University, Nanchang 330031, China
weiping.zou@univ-poitiers.fr, anne.sophie.capelle@univ-poitiers.fr,
philippe.carre@univ-poitiers.fr, nrzhou@ncu.edu.cn, jhwu@ncu.edu.cn

Résumé – L'apprentissage en profondeur (deep-learning) est aujourd'hui largement utilisé en stéganalyse. Cependant, cette approche nécessite deux contraintes importantes: un jeu de donnés de grande taille et des ressources informatiques puissantes. À l'inverse, la recherche en stéganalyse dans un environnement expérimental imparfait o, par exemple, le nombre des échantillons est limité, est jusqu'ici peu étudié. Nous proposons ici une approche alternative au deep-learning pour la stéganalyse capable de fournir des résultats performants dans un environnement contitué d'ensemble d'apprentissage de petite taille et nécessitant peu de puissance calculatoire. La méthode proposée combine la théorie des fonctions de croyance par le biais d'une version crédibiliste de l'algorithme des K plus proche voisins et l'apprentissage d'ensemble. Le premier concept transforme les distances entre l'image test et ses voisins en degrés d'évidence (masses). Au lieu d'utiliser directement cette masse pour la classification, l'originalité du travail réside en l'utilisation d'une stratégie de type *Ensemble Classifier*: l'algorithme des K-PPV évidentiel est appliqué dans différents sous-espaces de dimension réduite de l'espace de représentation. Pour chaque image testée, les masses issues de cet algorithme sont alors fusionnées pour décider de la nature stéganographiée ou non de l'image. Les résultats expérimentaux montrent que cette stratégie de type *Ensemble Classifier* permet d'optimiser les résultats de classification pas rapport à une approche standard et d'obtenir une précision correcte avec un ensemble d'apprentissage de petite taille.

Abstract – Deep learning is widely used in current steganalysis, which requires an important factor: a large sized dataset. The research of steganalysis with an imperfect experimental environment (such as limited samples) did not get enough attention. The paper proposes a steganalysis method, which is applied with a small sized dataset, and without much computing. The proposed method combines evidence theory, K-nearest neighbor algorithm (KNN) and ensemble learning. First a KNN classifier based on evidence theory is regarded as base learner, which transforms the distances between the test image and its neighbors into degrees of evidence (masses), and it combines the masses among the neighbors with Dempster's combination rule. Then the combining rule is also applied in the ensemble learning method. The two-step combining of evidence theory provides an advanced fusion strategy. In the end, the classification results are computed after transforming the masses into probabilities. Experimental results show that the proposed method is able to obtain good accuracy with a small size training set.

1 Introduction

Steganography is an art of hiding information into a digital medium in such a way that hidden information is imperceptible to the third party [1]. Due to the high capacity of hidden information and popularity on the Internet, digital images become the most commonly used carrier in steganography. Meanwhile, steganalysis is a method of detecting the presence of steganography, which aims to determine whether a suspicious medium contains hidden message. The image steganalysis has become an essential branch of modern steganalysis.

In recent years, the image steganalysis which combines deep learning is developing rapidly [3]. However, these steganalysis methods require a big-sized dataset to train proper models, e.g., in [2] the experiments used 100,000 images as training set. To "*bring the steganalysis out of the laboratory*" [4], the steganalysis with a non-ideal experimental environment such as small sized training set is worth to be explored. It is possible to propose a steganalysis with transfer learning, but actually the solution in the literature did not use limited training set. It is an open problem to apply transfer learning for image steganalysis with small database.

The image steganalysis method proposed in this paper combines evidence-theoretic KNN with ensemble learning: several evidence-theoretic KNN base learners are computed using numerous feature sets, in which each feature set is a subspace of the original one's. The paper is organized as follows: Section 2 presents basic of evidence theory and evidential KNN algorithm. Section 3 details the interest of ensemble learning strategy in a reduced feature space and describes the proposed algorithm. Section 4 provides experimental results before discussion and conclusion.

Acknowledgements: This work is supported by the National Natural Science Foundation of China (Grant no. 61861029).

2 The steganalysis based on KNN with evidence theory

2.1 Evidence theory background

Evidence theory is a mathematical framework of describing quantified beliefs brought by various hypothesis of a question [7]. In evidence theory, a limited set of mutually exclusive hypothesis of a question is called a *frame of discernment*, denoted by Ω. A basic belief assignment (BBA) m is defined as $\sum_{A\subseteq\Omega} m(A) = 1$. The hypotheses of an undetermined question are represented by separated belief functions.

When several distinct and independent pieces of information exist, one can combine information using fusion operator. Let m_1 and m_2 be two normalized masses from distinct pieces of evidence, which are defined under the same frame of discernment Ω, and m is the mass combination result. According to the Dempster's combination rule [7], the mass $m = m_1 \oplus m_2$ is given by:

$$\begin{cases} m(A) = \eta \sum_{B\cap C=A} m_1(B).m_2(C), \forall A \subseteq \Theta, A \neq \emptyset, \\ m(\emptyset) = 0 \end{cases} \quad (1)$$

and

$$\eta^{-1} = 1 - \sum_{B\cap C=\emptyset} m_1(B).m_2(C). \quad (2)$$

where η is a normalization constant which represents the amount of conflict between two BBAs.

Evidence theory is often used in the classification problems. According to Smets [8], the decision about class membership have to be taken in favor of one single hypothesis ω_i of Ω using the Pignistic probability function derived from the mass m:

$$BetP(\omega_k) = \frac{1}{1 - m(\emptyset)} \sum_{\omega_k \subseteq A} \frac{m(A)}{|A|} \quad \forall \omega_k \in \Omega \quad (3)$$

where $|A|$ denotes the cardinality of A. The simplest decision rule selects the element of Ω with the highest pignistic probability.

2.2 ET-KNN for image steganalysis

The image steganalysis is to discriminate between innocent images and suspected images i.e., cover images and stego images. Based on evidence theory, the image steganalysis can be regarded as a question of distinguishing the class of the images, with the *framework of discernment* defined by $\Omega = \{cover, stego\}$.

Assume there exists a test image described by an image feature vector s and $x_1, \ldots, x_k$ are its k neighbors in a well-known training set. The image steganalysis based on evidence-theoretic KNN (ET-KNN) [6] considers that each neighbor of an unlabeled test image provides a piece of evidence: the shorter the distance with the neighbor, the more likely the test image belongs to the same class as the neighbor. Thus, considering a neighbor feature vector x_i, a basic belief mass is computed by:

$$\begin{cases} m_i(\omega_j) = \alpha \exp(\gamma.d(s, x_i)) \\ m_i(\Omega) = 1 - \alpha \exp(\gamma.d(s, x_i)) \end{cases} \quad (4)$$

where $\omega_j \in \Omega$ is the class membership associated with x_i, d is the distance between s and its neighbor, γ is the reciprocal of mean distance of each class, and $0 < \alpha < 1$ is a constant.

Based on [5], the tuning of γ has a meaningful influence on classification. An optimization of γ is carried out by minimizing an error function of training set, as described in [5]. The evidence-theoretic KNN with an optimization procedure is named as optimized evidence-theoretic KNN (OET-KNN). OET-KNN usually considers only k neighbors of the learning test image. Indeed, the more the distance between s and a neighbor x_i is high, the more the value $m_i(\omega_j)$ tends to zero and m_i becomes non-informative. Finally, the belief function that represents our knowledge about s is given by using Dempster's rule on the k masses m_i with $m = m_1 \oplus \cdots \oplus m_k$.

2.3 Limitations of OET-KNN for steganalysis

Applying OET-KNN directly on image training set with features for steganalysis is possible. However, it will lead to two drawbacks. Firstly, based on machine learning theory, as the classifier OET-KNN, when it is applied with a testing set, the variance caused by the peculiarities of one single training set may very high. Therefore, it usually uses different training sets to build a framework of ensemble learning to reduce the variances, in which OET-KNN is viewed as a base learner. Since the combination of base learners is capable of reducing the error rate, it is best to make the base learners as independent as possible. With the fact that the training set size is limited, how to maximize the diversity should be concerned. Secondly, the original dimension of the image vector in steganalysis is normally very high. A basic truth is that the computation burden shows a significant increase with the growth of the vector dimension. Therefore, how to reduce the feature set size should also be taken into consideration.

3 The ensemble learning with evidence theory fusion strategy

In traditional ensemble learning strategy of classification problem, each base learner provides binary results (e.g., 0 and 1), then it uses a fusion strategy e.g., majority voting to calculate the final results. As described in Section 2, the masses of each test image is a piece of evidence, which shows how much the current base learner supports to the test image. One can use Dempster's combination rule of evidence theory to combine the masses from different base learners. The proposed steganalysis method in this paper is called EN-OETKNN, which refers to the image steganalysis method that combines OET-KNN with ensemble learning.

3.1 Data pre-processing and distance calculation

Encountering with the situation of a small number of training samples, the proposed method adopts the bootstrapping

selection rule to choose a training set. With size m image vectors for training set, each time it selects size m image vectors with replacement as a training set for each base learner.

Based on the feature extraction algorithm used in the steganalysis, the initial feature dimension is usually high, some features are less meaningful and some feature effects are overlapped. Since there is no prior knowledge to distinguish which features are more 'valuable', in this paper it chooses features by using bootstrapping selection, the size of feature set is determined through the optimization experiments. The features of image vectors are correlated and in different scales. The small scale features may also express the difference between two class images, hence those features are as important as the big scale ones. The Mahalanobis distance is applied to calculate the distance between the test image and the training set.

3.2 The ensemble learning of steganalysis

Based on the ensemble learning and the multiple strategies described previously, a steganalysis method is proposed and illustrated in Fig. 1. After the processing of each base learner, the N base learners produce a mass matrix of size (N, n, 3), in which '3' means three different classes(cover, stego, and unknown). By evidence theory, the masses from different base learners are under the same *frame of discernment* Ω, then the masses of the same test image among different base learners are combinable. The combination can also be conducted by Dempster's combination rule.

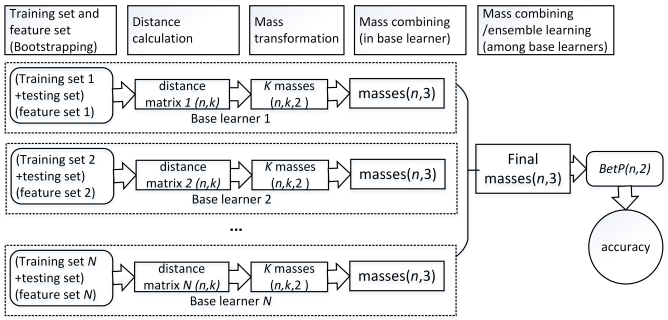


FIGURE 1 – The steganalysis of EN-OETKNN

For test image s, assuming the final masses from base learner T_i are $m_s^{T_i}$, then the one combines the masses with different base learners, e.g., the combination of s from T_1 to T_N is: $m_s^{T_1-T_N} = m_s^{T_1} \oplus m_s^{T_2} \oplus \cdots \oplus m_s^{T_N}$. According to TBM [8], the mass values of each test image are transformed into pignistic probability ($BetP$). For each test image, $BetP^{T_1-T_N}(stego)$ and $BetP^{T_1-T_N}(cover)$ are obtained. The final decision of each test image is determined by the pignistic probabilities of two classes (cover, stego) of the image. The steganalysis decision results are illustrated in Table 1. TN, TP represent the number of cover images and stego images which are classified correctly, and $accuracy = (TN + TP)/n$, in which n is the size of test set.

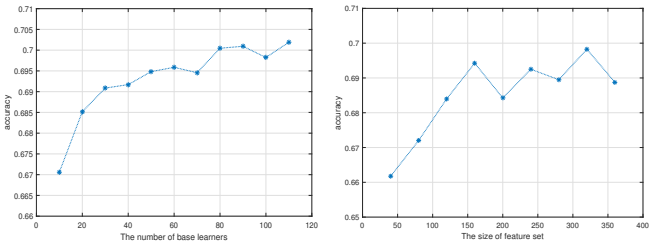
TABLE 1 – The classification result matrix

pignistic probability comparison	cover image	stego image
$BetP(cover) > BetP(stego)$	TP	FP
$BetP(cover) < BetP(stego)$	FN	TN

4 Experiments

The original images are from BOSS database (version: BOSSbase-1.01) [9]. The stego images are embedded with different payloads by using nsF5 [10] . The feature extraction algorithm used here is CC-PEV [11], which extracts a 548-dimensional vector from each image. The CC-PEV contains Cartesian Calibrated features and PEV features, which shows good ability to manifest the effects of nsF5.

In the proposed method, there are two parameters need to be optimized: the number of base learners and feature set size. The tuning results of the two parameters are shown in Fig. 2.



(a) Selection of the number of base learners (b) Selection of the feature set size

FIGURE 2 – The parameter selection of the proposed method

In Fig. 2, the x-axis is parameters to be tuned, and the y-axis is the accuracy. In the case of the same size training set, the experiments set different number of base learners and different size feature sets. The parameters are considered optimal or near optimal when the accuracy is getting stable. Therefore, the number of base learners is set to 91, and the feature set size of image vector is set to 300, which is smaller than the original image vector dimension. As the number of the neighbors - the k, it is found that k becomes less important after it exceeds a certain size, so here the k is set to 101.

To illustrate the improvement of the proposed method, it is compared with Kodovský's classic steganalysis [14]. The training set size is limited to 1000 intentionally to fit the subject of small-sized training set. In order to avoid overfitting and selection-bias, a 10-folds cross-validation is applied to obtain the final accuracy. The embedding payload used are 0.1 bpAC (bits per non-zero AC coefficient) and 0.2 bpAC. For each payload, there are five different training sets. The accuracies of the two methods by using the same training set are shown in Fig. 3. From the figure, the proposed method is able to improve the steganalysis effect when training set size is small.

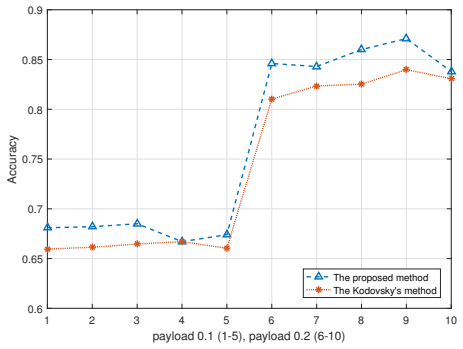


FIGURE 3 – The comparison between the proposed method with Kodovský's classic steganalysis method

Last but not least, we are proposing some discussions. The Rich Model with modern steganography is also applied with the proposed method. The experiments used S-UNIWARD [12] as steganography algorithm, and used the variant of Spatial Rich Model- maxSRMd2 [13] as the feature set. During the experiments, it was found that the tuning of parameters should be more careful. In the end, by using the same training sets, when the payload is 0.1 bpp (bits per pixel) and 0.2 bpp, the mean accuracies of the proposed method are 56.82% and 63.98 %, and accuracies by using Kodovský's classic steganalysis are 55.80% and 63.16%.

To compared with an example of deep learning approach, the same training set(cover and stego images) is applied in the SR-net [15]. Unsurprisingly, the lack of the training samples gives a bad accuracy. Recently the research by using transfer learning on the small training set is getting attention. The transfer learning is also applied in the steganalysis [16], but the authors did not apply the method with limited size training set. To compare with the proposed method, we would like to make an exploration in the future work.

5 Conclusion

The proposed steganalysis method combines the OET-KNN with ensemble learning. Using combining rule of evidence to combine the masses of neighbors, as well as the masses of base learners, it shows more advantage in the ensemble learning framework. The experiments of comparison show that the proposed method is performing well with small sized dataset. Deep learning is a powerful tool in multiple fields, including steganalysis. But the requirements of a large dataset is becoming its obstacle. The 'no-deep learning' steganalysis is still worth being explored.

Références

[1] Johnson, N. F., et Jajodia, S. *Exploring steganography : Seeing the unseen*. Computer. 1998. 31(2) :26-34.

[2] Qian, Y., Dong, J., Wang, W., et Tan, T. *Deep learning for steganalysis via convolutional neural networks*. Media Watermarking, Security, and Forensics. 2015. Vol. 9409, p.94090J.

[3] Chaumont, Marc. *Deep Learning in steganography and steganalysis from 2015 to 2018*. 2019. Xiv preprint arXiv :1904.0144.

[4] Ker, A. D., Bas, P., Bhme, R., Cogranne, R., Craver, S., Filler, T., et Pevn, T. *Moving steganography and steganalysis from the laboratory into the real world*. In Proceedings of the first ACM workshop on Information hiding and multimedia security. ACM. 2013. 45-58.

[5] Zouhal, L. M., et Denoeux, T. *Evidence-theoretic k-NN rule with parameter optimization*. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews). 2008. 28(2), 263-271.

[6] Denoeux, T. *K-nearest neighbor classification rule based on Dempster-Shafer theory*. IEEE transactions on systems, man, and cybernetics. 1995. 25(5), 804-813.

[7] Shafer, G. *A mathematical theory of evidence*. Princeton, N.J : Princeton university press. 1976.

[8] Smets, P., Kennes, R. *The transferable belief model*. Artificial intelligence. 1994. 66(2), 191-234.

[9] Bas, P., Filler, T., et Pevn, T. *"Break Our Steganographic System" : The ins and outs of organizing BOSS*. International workshop on information hiding. Springer, Berlin, Heidelberg, 2011. 59-70.

[10] J. Fridrich, T. Pevn, et J. Kodovsk. *Statistically undetectable JPEG steganography : Dead ends, challenges, and opportunities*. Proceedings of the 9th ACM Multimedia et Security Workshop, Dallas, TX. 2007. 314.

[11] J. Kodovsky, J. Fridrich *Calibration revisited*. Proceedings of the 11th ACM Multimedia and Security Workshop, Princeton, NJ, September 78, 2009. 6373.

[12] V. Holub, J. Fridrich. *Universal distortion design for steganography in an arbitrary domain*. Journal on Information Security, Special Issue on Revised Selected Papers of the 1st ACM IH and MMS Workshop, vol. 2014(1), 113.

[13] T. Denemark, V. Sedighi, V. Holub, R. Cogranne, et J. Fridrich *Selection-channel-aware rich model for Steganalysis of digital images*. IEEE International Workshop on Information Forensics and Security, 2015.48–53.

[14] Kodovsky, J., Fridrich, J., et Holub, V. *Ensemble classifiers for steganalysis of digital media*. IEEE Transactions Information Forensics and Security, 2012. 7(2), 432-444.

[15] Boroumand, M., Chen, M., et Fridrich, J. *Deep residual network for steganalysis of digital images*. IEEE Transactions on Information Forensics and Security, 2018. 14(5), 1181-1193.

[16] Ozcan, S., Mustacoglu, A. F. *Transfer Learning Effects on Image Steganalysis with Pre-Trained Deep Residual Neural Network Model*. IEEE International Conference on Big Data (Big Data).2018. 2280-2287