Avantages de la Sélection de Caractéristiques pour la Stéganalyse

Yoan Miche[1,2], Patrick Bas[1,2], Amaury Lendasse[1], Christian Jutten[2], Olli Simula[1]

[1]Helsinki University of Technology - Laboratory of Computer and Information Science
P.O. Box 5400, FI-02015 HUT, FINLAND

[2]Gipsa-Lab, Département Images et Signal
INPG, 46 avenue Félix Viallet, 38031 Grenoble cedex, FRANCE

{ymiche, lendasse, olli.simula}@cis.hut.fi
{patrick.bas, christian.jutten}@lis.inpg.fr

Résumé − La stéganalyse consiste à identifier la présence d'un message caché au sein d'un document. Cet article présente une méthodologie utilisant un ensemble de 193 caractéristiques d'images pour la stéganalyse. Deux buts sont visés: déterminer un nombre d'images suffisant pour un entraînement fiable d'un classifieur dans l'espace de grande dimension des données; puis utiliser la sélection de caractéristiques pour obtenir les plus pertinentes pour la classification. Cette réduction de la dimensionalité des données est effectuée avec l'algorithme forward et a permis de passer de 193 caractéristiques à 14 en conservant des performances très proches.

Abstract − Steganalysis consists in classifying documents as steganographied or genuine. This paper presents a methodology for steganalysis based on a set of 193 features, with two main goals: determine a sufficient number of images for effective training of a classifier in the obtained high-dimensional space, and use feature selection to select most relevant features for the desired classification. Dimensionality reduction is performed using a forward selection and reduces the original 193 features set by a factor of 13, with overall same performance.

Introduction

La stéganalyse consiste à détecter automatiquement la présence d'un message caché dans un contenu, message inséré par un algorithme de stéganographie. Ceci peut être envisagé comme un problème de classification, dans le sens où le but principal est de parvenir à créer une séparation optimale entre les espaces engendrés par les images originales et par les images stéganographiées. À cette fin, un certain nombre de caractéristiques extraites des images sont employées, nombre qui tend aujourd'hui à devenir particulièrement conséquent, voire trop important tant pour les classifieurs utilisés que pour l'interprétation des résultats.

Les premiers exemples de stéganalyse basée sur les LSB utilisaient un nombre de caractéristiques inférieur à la dizaine, avec une méthode de classification simple, et propre à chaque algorithme de stéganographie afin de détecter la possible présence d'une information dissimulée. En 1999, Westfeld propose un modèle d'attaque statistique basé sur un test du χ^2 sur les LSB des coefficients TCD des images analysées [2]. Plus récemment, en 2004, Fridrich dans [6] utilise un ensemble de 23 caractéristiques, tandis que Farid *et al.* proposaient déjà un ensemble de 72 caractéristiques en 2002 [12]. Depuis ces travaux, un nombre croissant de recherches fait usage de classifieurs à apprentissage supervisé dans des espaces de caractéristiques à grande dimension. On peut citer récemment Y. Q. Shi *et al.* [10], qui proposent un ensemble de 324 caractéristiques, basées sur des différences de blocs JPEG modélisées par des processus markoviens.

Bien que l'ajout de ces caractéristiques permette en effet d'améliorer les performances globales de classification, on ne peut s'empêcher de noter certains effets néfastes dus à cette augmentation. En effet, le nombre de caractéristiques ne revêt plus la même importance et la tendance est à en utiliser autant que possible tant que les performances de classification sont augmentées. Toutefois, il a été montré [4, 7] que la dimensionalité de l'espace dans lequel est entraîné le classifieur peut avoir une incidence cruciale sur ses performances, relativement au nombre d'images utilisées. La complexité (en termes de temps de calcul) du processus d'entrainement est un autre aspect important, puisque la plupart des classifieurs utilisés ont une dépendance au minimum linéaire avec le nombre de dimensions.

De la même façon, on peut s'interroger quant à l'interprétabilité des résultats et la possible analyse des faiblesses des algorithmes stéganographiques, lorsqu'un tel nombre de caractéristiques diverses est utilisé.

Cet article propose une méthodologie répondant à certains problèmes liés à la grande dimension posés par les nombreuses caractéristiques : la section suivante présente plus précisement ces effets du nombre important de caractéristiques. La section 2 décrit la méthodologie proposée, répondant à ces problèmes dans le cadre de la stégana-

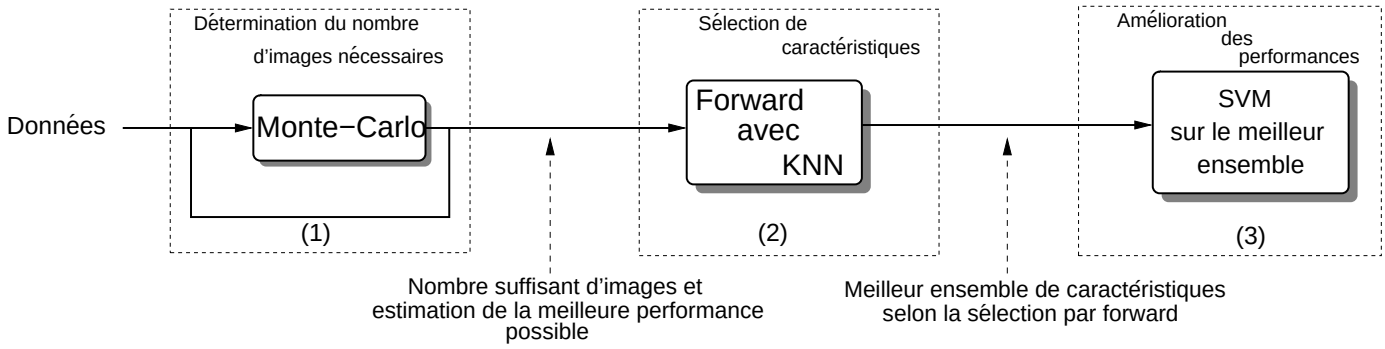

FIG. 1 – Méthodologie proposée : (1) Un nombre d'images approprié est déterminé par la méthode de Monte-Carlo ; (2) La sélection des caractéristiques par Forward est effectuée avec un KNN ; (3) Un "bon" ensemble de caractéristiques est déterminé et la performance est améliorée avec un SVM.

lyse, et l'application de cette méthode à l'algorithme de stéganographie Outguess [8], en utilisant un ensemble de 193 caractéristiques proposé par Fridrich [13] est illustré en section 3.

1 Effets de l'augmentation du nombre de caractéristiques

1.1 Dimensionnement des ensembles d'images pour la stéganalyse

Comme mentionné précédemment, un important nombre de caractéristiques utilisé pour la stéganalyse mène à un espace de grande dimensionalité dans lequel le classifieur est entraîné. Comme détaillé dans [4], un problème majeur dans de tels espaces est le phénomène des "points manquants" : si l'on prend le cas d'un hypercube de coté 1 en dimension d, la grille cartésienne de pas ϵ comprise dans le cube nécessite un nombre de points de l'ordre de $O((1/\epsilon)^d)$, afin d'être remplie. Dès lors, avec une dimensionalité aussi faible que 10 et une grille de pas 1/10, il faut 10^{10} points pour remplir chaque intersection de la grille.

La plupart des travaux en stéganalyse utilisent au minimum 10 à 20 dimensions, ce qui implique un nombre nécessaire d'images – pour remplir correctement l'espace –, impossible à atteindre. De ce fait, le classifieur utilisé doit extrapoler pour les images manquantes. Même s'il paraît clair qu'en pratique, la répartition des points n'a probablement pas besoin d'être aussi dense que dans l'exemple de la grille cartésienne, une estimation du nombre minimum d'images nécessaires à un entraînement correct du classifieur (ainsi qu'à l'obtention de performances stables), doit être déterminé. La première partie de la méthodologie présentée en section 2 propose une façon de résoudre ce problème.

1.2 Augmentation de la complexité des calculs

Le problème des temps de calcul, particulièrement pour l'entraînement du classifieur, est un autre effet. On peut considérer que les classifieurs usuels, tels que le KNN [1] ou le SVM [9], ont une dépendance linéaire avec le nombre de caractéristiques et qu'ainsi, en réduisant sensiblement la dimensionalité de l'espace, les temps de calcul diminuent, permettant d'utiliser plus d'images et mieux palier à cet effet d'images manquantes. Le meilleur compromis entre fiabilité des résultats et performances, en termes de nombre de caractéristiques et d'images, doit être trouvé.

1.3 Manque d'interprétabilité des résultats

Même si les travaux utilisant des ensembles particulièrement exhaustifs ont de très bonnes performances, si l'on s'interroge sur les faiblesses d'un algorithme de stéganographie et les raisons pour lesquelles certaines caractéristiques réagissent particulièrement fortement, le travail d'analyse est rendu beaucoup plus difficile par la possible grande taille de l'ensemble de caractéristiques.

Le fait de réduire le nombre de caractéristiques nécessaires à une classification efficace, par une méthode de sélection de caractéristiques, permet de mieux comprendre l'algorithme sous-jacent du procédé stéganographique et évite de réduire la stéganalyse à une simple recherche de la meilleure performance de classification.

2 Méthodologie proposée et techniques utilisées

La figure 1 illustre la méthodologie proposée, en trois étapes.

2.1 Nombre suffisant d'images et classifieurs

Un nombre approprié d'images – dépendant du nombre de caractéristiques – doit d'abord être déterminé : un classifieur KNN [1] utilisé avec une méthode de Monte-Carlo [11] permet d'estimer le bruit et de donner un intervalle de confiance pour les résultats, avec une faible complexité. Des sous-ensembles des données sont tirés aléatoirement (sans répétitions) et utilisés avec le classifieur.

Deux classifieurs différents ont été utilisés pour les expérimentations : SVM et KNN. Le KNN est un classifieur à

apprentissage supervisé basé sur la distance, proposé par Devijver et Kittler dans [1]. Il est basé sur un vote de majorité des classes des k plus proches voisins du point considéré, pour la détermination de la classe de celui-ci.

Vapnik [9] en 1963 propose l'idée du SVM (développée plus récemment (1992,1995) par Boser, Guyon et Vapnik [5]). L'idée d'origine était de séparer les données grâce à un hyperplan dont la distance aux plus proches données est maximale, ce qui constituait un classifieur linéaire en grande dimension. Les développement récents introduisent une partie non-linéaire dans la méthode.

2.2 Sélection des caractéristiques par l'algorithme Forward et construction du classifieur final

La sélection des caractéristiques se fait par un algorithme "glouton" nommé Forward [14] : les caractéristiques sont sélectionnées une à une et celle donnant les meilleures performances avec l'ensemble de caractéristiques déjà retenu est ajoutée à cet ensemble. L'algorithme part d'un ensemble de caractéristiques vide et obtient donc un classement des caractéristiques selon leur ordre d'importance pour la maximisation de la performance de classification. Un SVM est ensuite appliqué sur le meilleur ensemble obtenu par le Forward afin d'augmenter les performances et les rendre comparables aux travaux actuels.

3 Résultats

La base d'images utilisée est constituée de 13000 images naturelles issues de 5 appareils numériques différents. Les images ont été redimensionnées à une taille de 800×600 pour éviter des possibles effets de blocs et autres artefacts de recompression JPEG sur une nouvelle grille. Dans le même temps, toutes les images sont passées en niveaux de gris (256 niveaux). Un découpage en 512×512 est ensuite effectué, puisque notre implémentation de l'extraction des 193 caractéristiques de Fridrich fonctionne sur des images de cette taille. Finalement, la moitié des images ainsi préparée est stéganographiée avec l'algorithme Outguess, à un taux d'insertion (ratio entre la quantité d'information embarquable dans le medium et la quantité embarquée) de 25%, tandis que l'autre moitié est conservée telle quelle.

3.1 Recherche du nombre d'images nécessaire

La figure 2 illustre cette recherche du nombre d'images pour des ensembles de taille variant entre 100 et 4000 images (les temps de calculs au-delà deviennent irréalisables). Pour cette étude, on peut considérer (cf. Fig. 2) qu'un nombre suffisant d'images est 2000, étant donné l'augmentation de performances particulièrement lente au-delà. La suite de la méthodologie se fait sur 4000 images, les temps de calcul restant acceptables avec un tel ensemble dans notre cas.

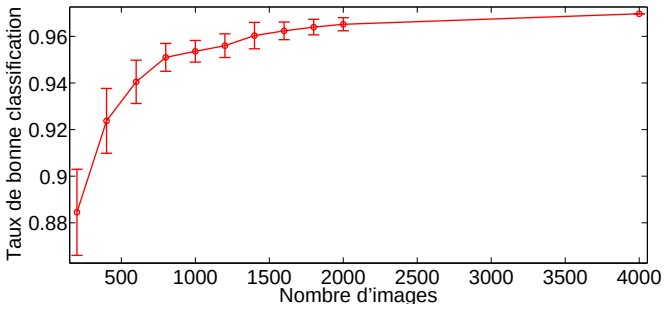

FIG. 2 – Monte-Carlo avec KNN sur des ensembles de 100 à 4000 images. Une valeur "suffisante" est ici 2000 images.

3.2 Sélection de caractéristiques et construction du classifieur

La figure 3 présente les résultats pour notre cas, où un SVM a été appliqué à chacun des ensembles obtenu par l'algorithme forward, pour comparaison.

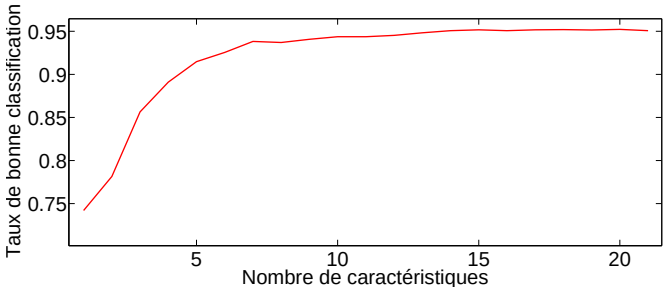

FIG. 3 – Forward avec KNN pour 1 à 22 caractéristiques. L'amélioration de performances par SVM est ici déjà effectuée.

Pour cette situation, les 14 premières caractéristiques obtenues par le Forward sont retenues. La figure 3 permet de voir que celles-ci constituent un bon compromis performance/complexité, dans le sens où les performances s'améliorent très lentement avec l'ajout d'autres caractéristiques, et que la réduction du nombre de caractéristiques est très sensible.

Cette sélection permet d'obtenir comme le montre la table 1 – pour le cadre de nos expérimentations utilisant Outguess et l'ensemble de 193 caractéristiques mentionné – un taux de bonne classification de 95.08% avec cet ensemble de 14 caractéristiques, soit seulement 1.9% de moins que lorsque l'on utilise l'ensemble des 193 caractéristiques.

TAB. 1 – Résultats des différents classifieurs en cross-validation (Leave-One-Out et 10-fold) et test.

	LOO/10-fold	Test	Temps Calc.
KNN 193	86.65%	85.89%	4.5min
KNN 193→14	93.20%	89.02%	60s
SVM 193	96.92%	96.76%	49h
SVM 193→14	95.08%	94.86%	4.5h

Quant aux temps de calcul, l'implémentation C du SVM utilisée [3] s'exécute en 49 heures sur l'ensemble des 193

caractéristiques (toujours avec 4000 images), tandis que l'on parvient à 4.5 heures avec cet ensemble réduit à 14, soit un ordre 10.9.

4 Analyse

Les résultats précédents peuvent être interprétés selon deux points de vue.

Du point de vue "Intelligence Artificielle" La réduction du nombre de dimensions nécessaires par 13, tout en conservant des performances très similaires dans l'intervalle de confiance, est un résultat important. En effet, la complexité des calculs et le temps d'exécution sont fortement réduits. D'autre part, de nouvelles analyses et expérimentations deviennent possibles du fait de cette plus faible dimensionalité des données.

Du point de vue de la stéganalyse Les résultats sont légèrement inférieurs aux meilleurs obtenus pour l'algorithme Outguess, en utilisant l'ensemble des caractéristiques proposées par Fridrich dans [13]. Toutefois, le gain en interprétabilité des résultats et l'importante diminution du temps de calcul sont des avantages non négligeables. Enfin, l'analyse des caractéristiques sélectionnées permet de mettre en lumière les faiblesses de l'algorithme de stéganographie utilisé : les caractéristiques sélectionnées traduisent une répartition anormale des coefficients TCD de valeurs égales à -1 ou -2. Nos futurs travaux consisteront à utiliser cette analyse afin d'améliorer la sécurité des algorithmes de stéganographie.

5 Conclusion

La méthodologie proposée permet donc de respecter certains principes et restrictions dûs à l'utilisation de classifieurs à apprentissage supervisé, par une estimation empirique du nombre d'images nécessaire à un apprentissage fiable et à variance faible, statistiquement parlant. D'autre part, par l'utilisation d'une méthode de sélection de caractéristiques, il est possible de conserver des performances proches de celles obtenues avec un ensemble particulièrement exhaustif de caractéristiques, en réduisant celui-ci par un facteur 13, dans le cas de l'expérimentation présentée.

Références

[1] Devijver P. A. and Kittler J. *Pattern recognition : a statistical approach*. Prentice Hall, New York, 1982.

[2] Westfeld A. and Pfitzmann A. Attacks on steganographic systems. In *IH '99 : Proceedings of the Third International Workshop on Information Hiding*, pages 61–76, London, UK, 2000. Springer-Verlag.

[3] Chang C. and Lin C. *LIBSVM : a library for support vector machines*, 2001. http ://www.csie.ntu.edu.tw/ cjlin/libsvm.

[4] François D. *High-dimensional data analysis : optimal metrics and feature selection*. PhD thesis, Université catholique de Louvain, September 2006.

[5] Boser B. E., Guyon I. M., and Vapnik V. N. A training algorithm for optimal margin classifiers. In *Fifth Annual Workshop on Computational Learning Theory*, pages 144–152, 27-29 Juillet 1992.

[6] Fridrich J. Feature-based steganalysis for jpeg images and its implications for future design of steganographic schemes. In *Information Hiding : 6th International Workshop*, volume 3200 of *Lecture Notes in Computer Science*, pages 67–81, May 23-25 2004.

[7] Verleysen M. and François D. The curse of dimensionality in data mining and time series prediction. In *IWANN'05 : 8th International Work-Conference on Artificial Neural Network*, volume 3512 of *Lecture Notes in Computer Science*, pages 758–770, 8-10 Juin 2005.

[8] Provos N. Defending against statistical steganalysis. In *10th USENIX Security Symposium*, pages 323–335, 13-17 April 2001.

[9] Vapnik V. N. *Statistical Learning Theory*. Wiley-Interscience, 1998.

[10] Shi Y. Q., Chen C., and Chen W. A markov process based approach to effective attacking jpeg steganography. In *8th Information Hiding Workshop*, Old Town Alexandria, Virginia, USA, 10–12 July 2006.

[11] Christian P. R. and Casella G. *Monte Carlo statistical methods*. Springer, 1999. ISBN :038798707X.

[12] Lyu S. and Farid H. Detecting hidden messages using higher-order statistics and support vector machines. In *5th International Workshop on Information Hiding*, Noordwijkerhout, The Netherlands, 2002.

[13] Pevny T. and Fridrich J. Merging markov and dct features for multi-class jpeg steganalysis. In *IS&T/SPIE EI 2007*, volume 6505 of *Lecture Notes in Computer Science*, January 29th - February 1st 2007.

[14] A. W. Whitney. A direct method of nonparametric measurement selection. In *IEEE Transactions on Computers*, volume C-20, pages 1100–1103, September 1971.