УДК 004.93.12

ИССЛЕДОВАНИЕ МЕТОДОВ КЛАССИФИКАЦИИ В ЗАДАЧЕ СТЕГОАНАЛИЗА ИЗОБРАЖЕНИЙ

О.О. Шумская
Санкт-Петербургский институт информатики и автоматизации РАН
Россия, 199178, Санкт-Петербург, 14-я линия В.О., 39
E-mail: shumskaya.oo@gmail.com

А.Л. Ронжин
Санкт-Петербургский институт информатики и автоматизации РАН
Россия, 199178, Санкт-Петербург, 14-я линия В.О., 39
E-mail: ronzhin@iias.spb.su

Ключевые слова: классификатор, признаки, линейный дискриминант Фишера, наивный байесовский классификатор, нейронные сети, AutoMPL, опорные векторы, стегоанализ

Аннотация: Цифровые изображения встречаются ежедневно во всех сферах деятельности человека: графики, схемы, модели, чертежи, фотографии, логотипы и прочее. Ежедневно в сети Интернет миллионы людей обмениваются изображениями, не подозревая о возможном секретном содержимом, скрытом в файле от человеческого глаза. Стеганография – наука о способах передачи, хранения информации, обеспечивающих сокрытие наличия этой информации в некотором сигнале, предоставляет различные методы сокрытия данных в цифровых изображениях [1]. С целью обнаружения факта наличия секретных вложений в цифровых файлах применяются методы стегоанализа, представляющего собой науку о способах выявления фактов наличия скрытых сообщений в цифровых объектах. Ежегодно появляются новые методы встраивания информации, отличающиеся большей ёмкостью и незаметностью для человеческого глаза. Однако авторы нечасто приводят исследования по устойчивости метода к стегоанализу. В работах, где встречаются эксперименты по устойчивости к стегоанализу, преимущественно применяется один метод классификации, выбор которого не обоснован экспериментально. Исследование устойчивости перед различными методами стегоанализа и разными классификаторами позволит изучить метода с разных сторон и повысить устойчивость встраивания. В работе рассмотрены известные работы по стегоанализу с использованием методов машинного обучения. Приведены эксперименты с различными методами классификации и их вариациями с целью их сравнения и выявления подходящих классификаторов.

1. Введение

В общем случае стегоанализ цифровых объектов рассматривается как задача двухклассовой классификации, когда для каждого анализируемого объекта выбирается один из двух исходов: нет вложения или объект содержит скрытые данные. Так как стеганографическое встраивание секретной информации может осуществляться в пространственную (значения пикселей) и частотную области цифрового изображения, то и стегоа-

нализ может быть на основе признаков в частотной области, на основе пространственной области или с комбинированным набором признаков.

Полученные в ходе исследования цифрового объекта значения признаков объединяются в вектор, с которым уже работает классификатор.

При стегоанализе важно осуществлять классификацию, учитывая все признаки во взаимодействии, а не по отдельности, так как цифровые объекты, в том числе цифровые изображения, могут сильно разниться по яркости, насыщенности, контрасту, однородности и другим характеристикам. Однако существует большое множество методов классификации, которые применимы в области стегоанализа.

Все чаще встречается в работах классификатор на основе линейного дискриминанта Фишера (ЛДФ), например, в [2-5]. Классификатор отличается своей гибкостью относительно количества признаков в наборе, так как весь вектор признаков проецируется на прямую. Идея классификации заключается в поиске лучшего направления данной проекции, которое позволит однозначности отнести величину к определенному классу.

В работах [6, 7] в качестве классификатора применяется метод опорных векторов (support vector machine – SVM). В общем случае суть метода заключается в поиске такой прямой, которая позволяет наилучшим образом разделить на классы точки обучающей выборки, размещенные на плоскости. После определения такой прямой все последующие точки классифицируются следующим образом: точки выше прямой относятся к одному классу, ниже прямой – к другому.

В работе [8] авторы применяют для классификации нейронные сети. Общую схему функционирования сети можно описать следующим образом: набор признаков через входной слой проходит 2 слоя нейронов, на каждом из которых взвешивается согласно соответствующей слою матрице весов. Значения на выходе сравниваются с входным набором, выполняется проверка: «узнала» система образ или нет. Если «узнала», то изображение можно отнести к данному классу, если нет – сеть проверяет принадлежность изображения к другому классу, изменяя матрицы весов. Однако на каждом уровне необходимо столько нейронов, сколько признаков в наборе, это может привести к массивным вычислениям в случае больших наборов признаков.

Довольно часто встречается в работах в качестве классификатора наивный байесовский классификатор (НБК), например, в работах [9-11]. Метод заключается в расчете апостериорной вероятности на основе известных из обучения классификатора априорных вероятностей. Решение принимается на основании сравнения двух рассчитанных вероятностей: объект относится к тому классу, чья апостериорная вероятность больше.

В некоторых работах можно встретить метод стегоанализа с классификатором на основе автоматического многослойного персептрона (AutoMLP) [12, 13]. Это простой алгоритм, повышающий темп обучения и регулирующий размер нейронных сетей во время обучения; включает идеи генетических алгоритмов и стохастической оптимизации. Суть заключается в поддержании малого числа сетей, которые обучаются параллельно с различными уровнями и различными числами скрытых модулей. После малого постоянного числа временных шагов определяется коэффициент ошибок, и худшие экземпляры заменяются копиями лучших сетей, измененных подобно мутации в генетическом алгоритме.

2. Тестовая выборка и набор признаков

2.1. Тестовая выборка

В качестве тестовой выборки были отобраны 963 полутоновых изображения размером 256*256 из баз изображений UCID и USC-SIPI ID: 511 без вложения (чистые, пус-

тые), в т.ч. 100 из которых обработаны с помощью современного редактора изображений Prisma; 452 с вложением по одному из распространенных стеганографических методов (Jsteg, PM1, F5).

Отобранные изображения каждого вида были поделены на обучающую и тестовую выборки в соотношении 65% (626) и 35% (337) соответственно.

2.2. Набор информативных признаков

Ключевым этапом является выбор информативных признаков, анализ которых позволяет отделять изображения, содержащие встроенную информацию, от чистых изображений. Ниже представлены признаки, составляющие набор для стегоанализа в данной работе.

Соотношения между энергией, собранной в отдельных частотных коэффициентах ДКП-спектра. В исследовании [2] показано, что для аддитивного встраивания информации в квантованные ДКП-коэффициенты JPEG-изображения характерно увеличение значений $E(f_0)$, $\sum_{|\eta|>1} E(f_\eta)$, $En_{|\eta|>1}$ и уменьшение значений $E(f_{|\eta|=1})$, $E(f_{|\eta|=1})$, $En_{|\eta|\leq 1}$.

(1) $$F_1 = \frac{E(f_0)}{E(f_{|\eta|=1})}, F_2 = \frac{\sum_{|\eta|>1} E(f_\eta)}{E(f_{|\eta|=1})}, F_3 = \frac{En_{|\eta|>1}}{En_{|\eta|\leq 1}},$$

где $E(f_0)$ – среднее значение частот нулевых АС-коэффициентов изображения по блокам.

$E(f_{|\eta|}=1)$ – среднее значение частот тех АС-коэффициента изображения, абсолютная величина которых равна 1.

$En_{|\eta|}>1$ – энергия тех АС-коэффициентов изображения, абсолютная величина которых > 1.

Для изображений характерна межблочная корреляция. Во время встраивания вносятся изменения в блоки изображения, что может привести к нарушению связи между блоками [14].

(3) $$F_4 = \frac{\sum_{i=1}^{n}(x_i-\overline{x})(y_i-\overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\overline{x})^2 \sum_{i=1}^{n}(y_i-\overline{y})^2}}$$

где $\overline{x}$, $\overline{y}$ – средние значения АС-коэффициентов соседних блоков, x_i, y_i – i-й АС-коэффициент соседних блоков.

Двойная гистограмма, предложенная авторами работ [4, 5], представляет собой матрицу, которая отражает, на каком месте сколько раз суммарно по всем блокам встретился коэффициент с определенным значением.

(4) $$f_5, \ldots, f_{11} = \frac{\sum_{k=1}^{B} \delta(d, d_k(i,j))}{\left\|\sum_{k=1}^{B} \delta(d, d_k(i,j))\right\|_{L_1}},$$

где d – фиксированное значение коэффициента, d ϵ [-5, -2, -1, 0, 1, 2, 5], B – количество блоков в изображении, i, j – координаты положения коэффициента в блоке,

$$\delta(d, d_k(i,j)) = \begin{cases} 1, d_k(i,j) = d \\ 0, else \end{cases},$$

L_1 норма – максимальная из сумм элементов по столбцам.

Каждую характеристику авторы рассчитывают дважды: для исследуемого изображения (J_1) и для изображения, которое получают путем обрезания исследуемого изображения сверху и слева на 4 пикселя (J_2). Таким образом, конечным значением признака будет значение функционала:

(5) $$F_5 \ldots F_{11} = \|f_5 \ldots f_{11}(J_1) - f_5 \ldots f_{11}(J_2)\|_{L_1}.$$

Для решения задачи стегоанализа можно использовать текстурные признаки, применяемые в различных задачах распознавания образов. В набор признаков были вклю-

чены некоторые признаки, приведенные в [15], такие как энергия, однородность, среднее по *i*, среднее по *j*, дисперсия по *i*, дисперсия по *j*.

Согласно закону Бенфорда вероятность появления цифры на первом месте в числе тем выше, чем меньше эта цифра. Авторы работы [16] предложили частный случай закона Бенфорда для квантованных ДКП-коэффициентов.

(6) $$F_{18}, F_{19} = N\log_{10}(1 + \frac{1}{s+x^q}),$$

где $x \in [8, 9]$,

N, s, q – параметры, зависящие от качества JPEG-сжатия.

Так же в набор были включены признаки, совмещающие в себе закон Бенфорда для чисел от 2 до 9 и идею смещения изображения [7, 8].

3. Вычислительные эксперименты

Эксперименты были проведены с 5 классификаторами, рассмотренными ранее: AutoMLP, НБК, ЛДФ и SVM. Для полного анализа была рассчитана общая точность методов. Результаты применения методов классификации к подготовленному набору изображений представлены ниже.

Таблица 1. Результаты классификации.

Классификация	Подача		Точность класса	Точность метода
	с вложением	без вложения		
AutoMLP				
с вложением	131	48	73,18%	73,62%
без вложения	41	117	74,05%	
НБК				
с вложением	116	63	64,80%	65,00%
без вложения	55	103	65,19%	
Нейронная сеть (НС)				
с вложением	129	50	72,07%	74,33%
без вложения	37	121	76,58%	
ЛДФ				
с вложением	124	55	69,27%	71,03%
без вложения	43	115	72,78%	
SVM				
с вложением	118	61	65,92%	67,77%
без вложения	48	110	69,62%	

4. Заключение

Вычислительные эксперименты показали, что в общем случае точность рассматриваемых классификаторов разнится максимум на 9,4% (между НС и НБК). НБК является наиболее простым алгоритмом для реализации и минимальным по объемам вычислений. Однако, помимо того, что у него минимальная общая точность среди всех методов, так и точность классификации изображений с вложением наименьшая среди методов.

Общая точность методов показывает, что для рассматриваемого случая наиболее подходящие методы – это ЛДФ, AutoMLP и НС (разница между общими точностями <3,5%). По точности обнаружения изображений без вложения НС имеет преимущество

перед AutoMLP на 2,53% и перед ЛДФ на 3,8%. По точности обнаружения изображений с вложением нейронные сети имеют преимущество перед AutoMLP на 1,11% и перед ЛДФ на 3,91%.

Среди рассмотренных вариантов можно было бы выбрать средний вариант – AutoMLP, однако с поправкой на стремление уменьшить ошибку первого рода, то есть увеличить вероятность обнаружения изображений с вложением, наиболее подходящим методом оказывается метод на основе нейронных сетей.

Также замечено, что при увеличении базы записей, точность методов повышается, так как учитывается больше различных вариантов изображений (контрастность, структура, количество мелких деталей, однородность, область и объем встраивания).

Список литературы

1. Коханович Г.Ф., Пузыренко А.Ю. Компьютерная стеганография. Теория и практика. К.: МК-Пресс, 2006. 288 с.
2. Jia-Fa M., XinXin N., Gang X., Wei-Guo Sh., Na-Na Zh. A steganalysis method in the DCT domain // Multimedia Tools and Applications. 2016. No, 75. P. 5999-6019.
3. Шумская О.О. Метод стегоанализа JPEG-изображений на основе энергетических признаков в частотной области // Материалы международной научно-технической конференции студентов, аспирантов и молодых ученых «Научная сессия ТУСУР-2017». Томск: В-Спектр, 2017. Ч. 6. С. 41-44.
4. Fridrich J. Feature-Based Steganalysis for JPEG Images and its Implications for Future Design of Steganographic Schemes // Proceedings of the Sixth International Workshop on Information Hiding, Lecture Notes in Computer Science. 2014. Vol. 3200. P. 67-81.
5. Chen M.-C. Alpha-trimmed Image Estimation for JPEG Steganography Detection // Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics. San Antonio, Texas, USA. 2009. P. 4581–4585.
6. Xia Zh., Wang X., Sun X., Liu Q., Xiong N. Steganalysis of LSB matching using differences between non-adjacent pixels // Multimedia Tools and Applications. 2016. P. 1947-1962.
7. Fusheng Y., Gao T. Novel Image Splicing Forensic Algorithm Based on Generalized DCT Coefficient-Pair Histogram // Proceedings of 10th Chinese Conference (IGTA 2015). China, Beijing, 2015. P. 63-71.
8. Zong H., Liu X., Luo X. Blind image steganalysis based on wavelet coefficient correlation // Digital Investigation. 2012. Vol. 9. P. 58-68.
9. Berg G. Searching For Hidden Messages: Automatic Detection of Steganography / G. Berg, I. Davidson, M.-Y. Duan, G. Paul // Proceedings of the 15th Innovative Applications of Artificial Intelligence (IAAI) Conference, August 12-14, 2003. Acapulco, Mexico. P. 51-56.
10. Maitra S., Paul G., Sarkar S.,Lehmann M., Meier W. New Results on Generalization of Roos-type Biases and Related Keystream of RC // Proceedings of the 6th International Conference on Cryptology in Africa (AFRICACRYPT). June 22-24, 2013. Cairo, Egypt. Vol. 7918. P. 222-239.
11. Евсютин О.О., Мещеряков Р.В., Шумская О.О. Стегоанализ цифровых изображений с использованием наивного байесовского классификатора // Материалы 10-й Всероссийской Мультиконференции по проблемам управления (МКПУ-2017). Ростов-на-Дону: Южный федеральный университет, 2017. С. 56-58.
12. Sabeti V., Samavi Sh., Mahdavi M., Shirani Sh. Steganalysis and payload estimation of embedding in pixel differences using neural networks // Pattern Recogn. 2010. № 43 (1). P. 405-415.
13. Lubenko I., Ker A.D. Steganalysis with mismatched covers: Do simple classifiers help? // Proceedings of the on Multimedia and Security, MM&Sec'12, September 6-7. 2012. New York, NY, USA. P. 11-18.
14. Евсютин О.О., Шумская О.О. Сравнение линейного дискриминанта Фишера и наивного байесовского классификатора в задаче стегоанализа JPEG-изображений // Материалы международной научно-практической конференции «Электронные средства и системы управления». Томск: В-Спектр, 2017. Ч. 2. С. 79-82.
15. Мицель А.А., Колодникова Н.В., Протасов К.Т. Непараметрический алгоритм текстурного анализа аэрокосмических снимков // Известия Томского политехнического университета. 2005. Т. 308 (1). С. 65-70.
16. Fu D., Shi Y.Q., Su W. A generalized Benford's law for JPEG coefficients and its applications in image forensics // Proceedings of SPIE 6505, Security, Steganography, and Watermarking of Multimedia Contents IX. USA, San Jose, 2007. P. 1L1–1L11.