ISSN 1814-1196
Научный вестник НГТУ
том 72, № 3, 2018, с. 121–134

http://journals.nstu.ru/vestnik
Science Bulletin of the NSTU
Vol. 72, No. 3, 2018, pp. 121–134

ИНФОРМАТИКА,
ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА
И УПРАВЛЕНИЕ

INFORMATICS,
COMPPUTER ENGINEERING
AND CONTROL

УДК 004.93.12

DOI: 10.17212/1814-1196-2018-3-121-134

Сравнительное исследование методов классификации в стегоанализе цифровых изображений*

О.О. ШУМСКАЯ[a], В.Ю. БУДКОВ[b]

199178, РФ, г. Санкт-Петербург, 14-я линия В.О., 39, Санкт-Петербургский институт информатики и автоматизации Российской академии наук

[a] *shumskaya.oo@gmail.com* [b] *visharmail@gmail.com*

Цифровые изображения – графики, схемы, модели, чертежи, фотографии, логотипы и прочее – встречаются ежедневно во всех сферах деятельности человека. 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 и 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 емкостью 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 метод 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.

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

* *Статья получена 13 июня 2018 г.*

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 (коэффициенты частотного преобразования, например, дискретного косинусного преобразования, дискретного преобразования Фурье, дискретного вейвлет-преобразования) области цифрового изображения, то и стегоанализ 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. Именно поэтому в стегоанализе нельзя применить классификаторы на основе деревьев: дерево решений предлагает ряд условий, по которым определить, является ли изображение с вложением или «чистым», можно с вероятностью 50 %, что недостаточно для классификации; дерево правил генерирует неоправданно большое множество неемких правил.

Однако существует большое множество методов классификации, которые применимы в области стегоанализа.

Всё чаще встречается в работах классификатор на основе линейного дискриминанта Фишера (ЛДФ) (см., например, [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 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.

С целью максимально возможной эффективности разрабатываемых методов стегоанализа исследователи строят достаточно большие признаковые пространства, включающие десятки и сотни тысяч признаков [7–9]. В пространствах высоких размерностей необходимо рассматривать гиперплоскости – пространства, размерность которых на единицу меньше, чем размерность исходного пространства, что влечет за собой большие объемы вычислений при большом размере набора признаков.

В работе [10] авторы для классификации применяют нейронные сети. Общую схему функционирования сети можно описать следующим образом: набор признаков через входной слой проходит два 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 [11–13]. Метод заключается в расчете апостериорной вероятности на основе известных из обучения классификатора априорных вероятностей. Так как стегоанализ представляет собой задачу двухклассовой классификации, то при анализе конкретного изображения необходимо рассчитать значения апостериорных вероятностей его принадлежности к каждому из двух рассматриваемых классов – «чистые» изображения и изображения с вложением. 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) [14, 15]. Этот 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.

1. ТЕСТОВАЯ ВЫБОРКА И НАБОР ПРИЗНАКОВ

В качестве тестовой выборки были отобраны 863 полутоновых изображения размером 256*256 из баз изображений UCID (Uncompressed Colour Image Database) [16] и USC-SIPI ID (University of Southern California Signal and Image Processing Institute Image Database) [17]: 411 без вложения (чистые, пустые) и 452 с вложением по одному из распространенных стеганографических методов (Jsteg, PM1, F5).

Отобранные изображения каждого вида были поделены на обучающую и тестовую выборки 65 % (561) и 35 % (302) соответственно.

Так, выбранные методы встраивания информации заключаются в сокрытии информации в частотную область цифрового изображения, и классификация должна быть основана на признаках в частотной области. Для проведения экспериментов подготовлена база значений набора из 14 признаков $\{F_1,\ldots, F_{14}\}$. Три признака на основе энергетических свойств изображения в частотной области [2]:

$$F_1 = \frac{E(f_0)}{E(f_{|\eta|=1})}, \tag{1}$$

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

$$F_2 = \frac{\sum_{|\eta|>1} E(f_\eta)}{E(f_{|\eta|=1})}, \tag{2}$$

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

$$F_3 = \frac{E_{n_{|\eta|>1}}}{E_{n_{|\eta|\le 1}}}, \tag{3}$$

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

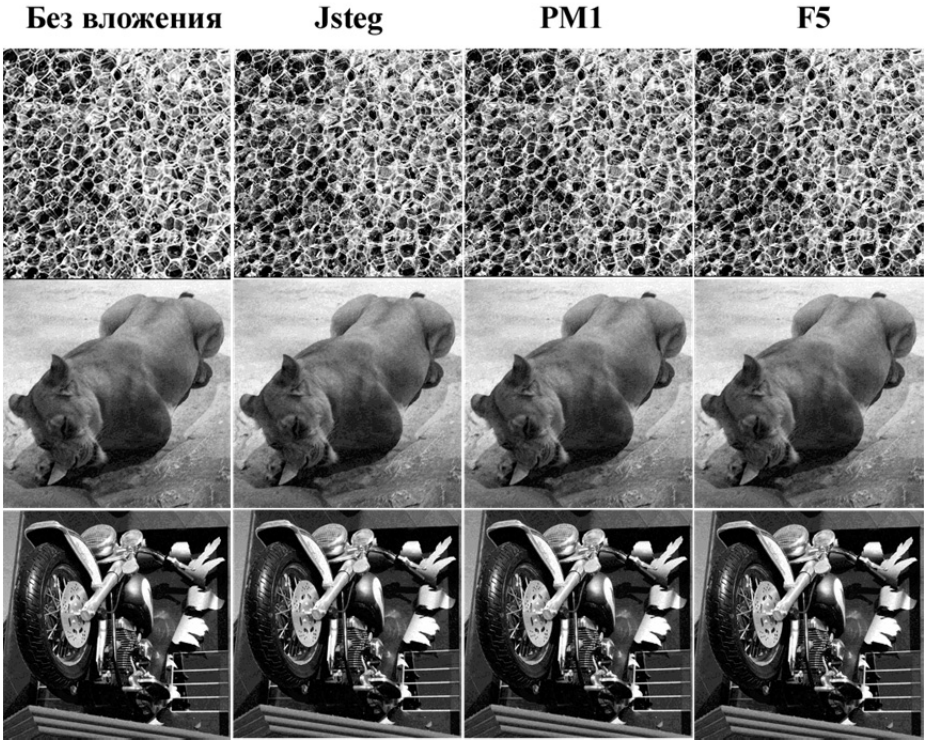

Рис. 1. Примеры изображений в четырех разных состояниях

Fig. 1. Examples of images in four various states

Отмечается, что именно признаки, основанные на соотношениях между энергией, собранной в отдельных частотных ДКП-коэффициентах, концентрируют в себе максимальную информацию о внутреннем содержании изображения.

Стеганографическое встраивание в частотную область осуществляется преимущественно в АС-коэффициенты со значениями по модулю, близкими к нулю. Поэтому двойные гистограммы для коэффициентов со значениями в диапазоне [–5, 5] концентрируют в себе максимальную информацию о вносимых искажениях при встраивании в частотную область.

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

$$f_4,\ldots,f_{14}=\frac{\sum_{k=1}^{B}\delta(d,\ d_k(i,j))}{\left\|\sum_{k=1}^{B}\delta(d,\ d_k(i,j))\right\|_{L_1}}, \tag{4}$$

где d – фиксированное значение коэффициента, $d \in [-5, 5]$; B – количество блоков в изображении; i, j – координаты положения коэффициента в блоке, $\delta(u,v)=\begin{cases}1, u=v,\\ 0,\ else,\end{cases}$ 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) (рис. 2). Подобное действие авторы объясняют следующим образом: при обрезании изображения слева и сверху разделение изображения на блоки сдвигается, коэффициенты дискретного косинусного преобразования освобождаются от влияния прошлой квантизации и содержат только статистические данные изображения, которые как раз важны при стегоанализе.

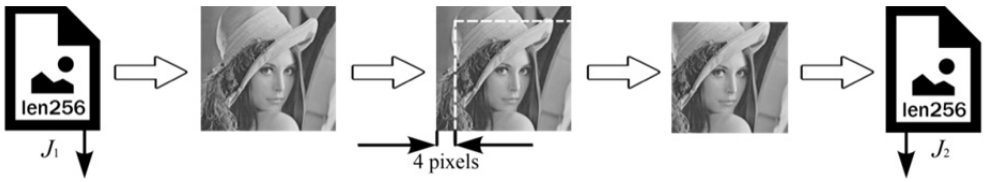

Рис. 2. Исследуемое изображение J_1 и изображение J_2, полученное путем обрезания изображения J_1

Fig. 2. The image under study J_1 and the image J_2 received by cropping the image J_1

Таким образом, конечным значением признака будет значение функционала

$$F_4\ldots F_{14}=\|f_4\ldots f_{14}(J_1)-f_4\ldots f_{14}(J_2)\|_{L_1}. \tag{5}$$

2. ИССЛЕДОВАНИЕ ВАРИАЦИЙ МЕТОДА ОПОРНЫХ ВЕКТОРОВ

Метод опорных векторов имеет множество настраиваемых параметров, однако регулирование осуществляется выбором типа функции ядра. В зависимости от того, как распределены классифицируемые объекты в пространстве, применяют метод опорных векторов с той или иной функцией ядра. Функция ядра определяет, каким образом классы будут условно разделяться в пространстве: линией, окружностями, кривой или с применением предварительной обработки данных, например, дисперсионного анализа.

Таблица 1

Table 1

Вариации метода опорных векторов

Variations of the support vector method

Функция ядра	Классификация	Подача		Точность класса
		с вложением	без вложения	
Точечная	с вложением	121	76	61,42 %
	без вложения	37	68	64,76 %
	отзыв класса	76,58 %	47,22 %	**62,58 %**
Радиальная	с вложением	122	66	64,89 %
	без вложения	36	78	68,42 %
	отзыв класса	77,22 %	54,17 %	**66,23 %**
Полиномиальная	с вложением	140	97	59,07 %
	без вложения	18	47	72,31 %
	отзыв класса	88,61 %	32,64 %	**61,92 %**
Нейронная	с вложением	106	75	58,56 %
	без вложения	52	69	57,02 %
	отзыв класса	67,09 %	47,92 %	**57,96 %**
На основе дисперсионного анализа	с вложением	122	64	65,59 %
	без вложения	36	80	68,97 %
	отзыв класса	77,22 %	55,56 %	**66,89 %**
Епачечникова	с вложением	113	69	62,09 %
	без вложения	45	75	62,50 %
	отзыв класса	71,52 %	52,08 %	**62,25 %**

Метод опорных векторов на основе дисперсионного анализа показал лучший результат среди вариаций – 66,89 %. Метод заключается в предварительной обработке классифицируемых данных с помощью дисперсионного анализа – выставление весов важности каждому объекту в выборке. Элементы с наибольшими весами принимаются за эталонные, и классификатор разделяет оставшиеся элементы выборки по классам на основе эталонных элементов [18].

3. ИССЛЕДОВАНИЕ НЕЙРОННЫХ СЕТЕЙ

У метода на основе нейронных сетей есть три настраиваемых параметра, которые влияют на работу классифкатора:

1) количество тренировочных кругов (стандартно 500);
2) изменение весов на каждом шаге (стандартно на 0,3);
3) импульс (стандартно 0,2).

Таблица 2

Table 2

Эксперименты для нейронных сетей с разными вариантами набора параметров

Experiments for neural networks with various sets of parameters

Параметры	Классификация	Подача		Точность класса
		с вложением	без вложения	
500 0,3 0,2	с вложением	128	70	64,65 %
	без вложения	30	74	71,15 %
	отзыв класса	81,01 %	51,39 %	**66,89 %**
500 0,4 0,2	с вложением	126	65	65,97 %
	без вложения	32	79	71,17 %
	отзыв класса	79,75 %	54,86 %	**67,88 %**
500 0,4-0,5 0,2	с вложением	126	67	65,28 %
	без вложения	32	77	70,64 %
	отзыв класса	79,75 %	53,47 %	**67,22 %**
500-600-700 0,5 0,3	с вложением	131	72	64,53 %
	без вложения	27	72	72,73 %
	отзыв класса	82,91 %	50,00 %	**67,22 %**
400 0,5 0,3	с вложением	129	64	66,84 %
	без вложения	29	80	73,39 %
	отзыв класса	81,65 %	55,56 %	**69,21 %**

Из приведенных результатов вычислительных экспериментов (табл. 2) видно, что лучшую точность классификации (69,21 %) показал набор признаков:

1) количество тренировочных кругов – 400;
2) изменение весов на каждом шаге – 0,5;
3) импульс – 0,3.

4. ИССЛЕДОВАНИЕ МЕТОДА AUTOMPL

У метода AutoMPL есть три параметра, обусловливающих процесс обучения классификатора:

1) S – количество тренировочных кругов (стандартно 10);
2) G – количество поколений для обучения (стандартно 10);
3) E – количество тренировочных ансамблей нейронных сетей, обучающихся параллельно (стандартно 4).

Таблица 3

Table 3

Эксперименты для AutoMPL с разными вариантами набора параметров

Experiments for AutoMPL with various sets of parameters

Параметры	Классификация	Подача		Точность класса
		с вложением	без вложения	
10 10 4	с вложением	132	72	64,71 %
	без вложения	26	72	73,47 %
	отзыв класса	83,54 %	50,00 %	**67,55 %**
20 20 4	с вложением	119	52	69,59 %
	без вложения	39	92	70,23 %
	отзыв класса	75,32 %	63,89 %	**69,87 %**
15 15 4	с вложением	151	99	60,40 %
	без вложения	7	45	86,54 %
	отзыв класса	95,57 %	31,25 %	**64,90 %**
10 10 7	с вложением	113	63	64,20 %
	без вложения	45	81	64,29 %
	отзыв класса	71,52 %	56,25 %	**64,24 %**
20 20 7	с вложением	144	90	61,54 %
	без вложения	14	54	79,41 %
	отзыв класса	91,14 %	37,50 %	**65,56 %**
15 15 7	с вложением	149	99	60,08 %
	без вложения	9	45	83,33 %
	отзыв класса	94,30 %	31,25 %	**64,24 %**

По результатам экспериментов (табл. 3) наибольшую точность классификации (69,87 %) показал набор признаков:

1) количество тренировочных кругов – 20;

2) количество поколений для обучения – 20;

3) количество тренировочных ансамблей нейронных сетей, обучающихся параллельно, – 4.

Однако стоит заметить, что AutoMLP показывает высокую точность обнаружения изображений с вложением (95,57 % при наборе значений параметров {15, 15, 4}), но из-за низкой точности обнаружения чистых изображений общая точность неконкурентоспособна.

5. СРАВНИТЕЛЬНЫЕ ЭКСПЕРИМЕНТЫ

Сравнительные эксперименты были проведены с каждым методом с соответствующими выбранными параметрами (табл. 4). Для полного анализа была рассчитана общая точность методов (табл. 5). Результаты применения методов классификации к подготовленному набору изображений представлены ниже.

Таблица 4

Table 4

Результаты классификации

Classification results

Классификация	Подача		Точность класса
	с вложением	без вложения	
AutoMLP {20; 20; 4}			
С вложением	119	52	69,59 %
Без вложения	39	92	70,23 %
Отзыв класса	75,32 %	63,89 %	
НБК			
С вложением	113	57	66,47 %
Без вложения	45	87	65,91 %
Отзыв класса	71,52 %	60,42 %	
Нейронные сети {400; 0,5; 0,3}			
С вложением	129	64	66,84 %
Без вложения	29	80	73,39 %
Отзыв класса	81,65 %	55,56 %	
ЛДФ			
С вложением	126	58	68,48 %
Без вложения	32	86	72,88 %
Отзыв класса	79,75 %	59,72 %	
SVM – Anova			
С вложением	122	64	65,59 %
Без вложения	36	80	68,97 %
Отзыв класса	77,22 %	55,56 %	

Таблица 5

Table 5

Точность методов

Accuracy of the methods

AutoMLP	НБК	Нейронные сети	ЛДФ	SVM
69,87 %	66,23 %	69,21 %	70,20 %	66,89 %

Стоит сразу отметить, что в рассматриваемой области начальной гипотезой принимается то, что изображение содержит некоторое секретное вложение, поэтому его и проверяют. Вследствие этого ошибкой первого рода является классификация изображения с вложением в класс изображений без вложения, а ошибка второго рода – классификация «чистого» изображения в класс изображений с вложением. В стегоанализе ошибка первого рода (пропуск изображения с вложением как пустого) является опаснее, нежели перепроверка или попытка извлечения данных из пустого изображения. Это немаловажный фактор, влияющий на выбор метода классификации.

Вычислительные эксперименты показали, что в общем случае точность рассматриваемых классификаторов разнится максимум на 5 % (между ЛДФ и НБК). Стоит отметить, что 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 (см. табл. 5) показывает, что для рассматриваемого случая наиболее подходящие методы – это ЛДФ, AutoMLP и нейронные сети (разница между общими точностями < 1 %). По точности обнаружения изображений без вложения AutoMLP имеет преимущество перед нейронными сетями на 8,33 % и перед ЛДФ на 4,72 %. По точности обнаружения изображений с вложением нейронные сети имеют преимущество перед AutoMLP на 6,33 % и перед ЛДФ на 1,9 %.

Среди рассмотренных вариантов можно было бы выбрать средний вариант – ЛДФ, однако с поправкой на стремление уменьшить ошибку первого рода, т. е. 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).

ЗАКЛЮЧЕНИЕ

Результаты проведенного исследования показали, что одностороннее рассмотрение стеганографического алгоритма приводит к снижению показателя устойчивости перед стегоаналитическими методами.

В работе были рассмотрены наиболее популярные методы классификации, такие как наивный байесовский классификатор, линейный дискриминант Фишера, метод опорных векторов, AutoMPL, нейронные сети. Проведены эксперименты с различными вариациями методов SVM, AutoMPL и нейронных сетей с целью выявления оптимальных параметров методов для лучшей точности классификации.

Проведены сравнительные вычислительные эксперименты со всеми методами для выявления наиболее действенной методики выявления вложений в стегоизображениях.

При рассмотрении вариаций метода AutoMLP отмечалось, что метод показывает очень высокую точность обнаружения изображений с вложением (более 95 %) при определенных значениях параметров. Таким образом, в спорных ситуациях или при наличии сомнения в решении можно воспользоваться следующим:

1) дополнительной проверкой с помощью метода AutoMLP, варьируя значения параметров в зависимости от рассматриваемой ситуации и конкретного вопроса;

2) комбинацией двух методов классификации, например, AutoMLP с его точностью обнаружения чистых изображений (бóльшая среди рассмотренных

методов) и метода на основе нейронных сетей с его высокой точностью обнаружения изображений с вложением, определив веса каждому методу.

Полученные результаты применимы не только в стегоанализе для выявления вложений, но также и в стеганографии для построения целевых функций, позволяющих осуществлять адаптивное встраивание секретной информации с минимизацией вносимых искажений в информативные признаки.

СПИСОК ЛИТЕРАТУРЫ

1. *Коханович Г.Ф., Пузыренко А.Ю.* Компьютерная стеганография: теория и практика. – Киев: МК-Пресс, 2006. – 288 с.

2. A steganalysis method in the DCT domain / M. Jia-Fa, N. XinXin, X. Gang, Sh. Wei-Guo, Zh. Na-Na // Multimedia Tools and Applications. – 2016. – N 75 (10). – P. 5999–6019.

3. *Шумская О.О.* Метод стегоанализа JPEG-изображений на основе энергетических признаков в частотной области // Научная сессия ТУСУР-2017: материалы международной научно-технической конференции студентов, аспирантов и молодых ученых. – Томск, 2017. – Ч. 6. – С. 41–44.

4. *Fridrich J.* Feature-based steganalysis for JPEG images and its Implications for future design of steganographic schemes // Information Hiding: 6th International Workshop, IH 2004. – Toronto, 2004. – P. 67–81. – (Lecture Notes in Computer Science; vol. 3200).

5. Alpha-trimmed image estimation for JPEG steganography detection / M.-C. Chen, S.S. Agaian, C.L.P. Chen, B.M. Rodriguez // 2009 IEEE International Conference on Systems, Man, and Cybernetics. – San Antonio, Texas, USA, 2009. – P. 4581–4585.

6. Steganalysis of LSB matching using differences between nonadjacent pixels / Zh. Xia, X. Wang, X. Sun, Q. Liu, N. Xiong // Multimedia Tools and Applications. – 2016. – Vol. 75 (4). – 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). – Beijing, China, 2015. – P. 63–71.

8. An improved approach to steganalysis of JPEG images / Q. Liu, A. Sung, M. Qiao, Z. Chen, B. Ribeiro // Information Sciences. – 2010. – Vol. 180 (9). – P. 1643–1655.

9 *Kodovsky J., Fridrich J.* Steganalysis of JPEG images using rich models // Proceedings of SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics XIV. – San Francisco, USA, 2012. – P. 7–20.

10. *Zong H., Liu X., Luo X.* Blind image steganalysis based on wavelet coefficient correlation // Digital Investigation. – 2012. – Vol. 9. – P. 58–68.

11. 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 Conference, August 12–14, 2003. – Acapulco, Mexico, 2003. – P. 51–56.

12. New results on generalization of roos-type biases and related keystream of RC4 / S. Maitra, G. Paul, S. Sarkar, M. Lehmann, W. Meier // Progress in Cryptology – AFRICACRYPT 2013: 6th International Conference on Cryptology in Africa: proceedings. – Cairo, Egypt, 2013. – P. 222–239. – (Lecture Notes in Computer Science; vol. 7918).

13. *Евсютин О.О., Мещеряков Р.В., Шумская О.О.* Стегоанализ цифровых изображений с использованием наивного байесовского классификатора // Материалы 10 Всероссийской мультиконференции по проблемам управления (МКПУ-2017). – Ростов н/Д.: ЮФУ, 2017. – С. 56–58.

14. Steganalysis and payload estimation of embedding in pixel differences using neural networks / V. Sabeti, Sh. Samavi, M. Mahdavi, Sh. Shirani // Pattern Recognition. – 2010. – Vol. 43 (1). – P. 405–415.

15. *Lubenko I., Ker A.D.* Steganalysis with mismatched covers: do simple classifiers help? // MM&Sec '12: proceedings of the 14th ACM Multimedia and Security Workshop, September 6–7, 2012. – New York, NY: ASM, 2012. – P. 11–18.

16. Image Databases [Electronic recourse] // ImageProcessingPlace.com. – URL: http://www.imageprocessingplace.com/root_files_V3/image_databases.htm (accessed: 14.09.2018).

17 The USC-SIPI Image Database [Electronic recourse] // USC University of Southern California. – Los Angeles, CA, 2017. – URL: http://sipi.usc.edu/database (accessed: 14.09.2018).

18. *Bharathi A., Natarajan A.M.* Cancer classification using support vector machines and relevance vector machine based on analysis of variance features // Journal of Computer Science. – 2011. – Vol. 7 (9). – P. 1393–1399.

Шумская Ольга Олеговна, младший научный сотрудник Санкт-Петербургского института информатики и автоматизации Российской академии наук. E-mail: shumskaya.oo@gmail.com

Будков Виктор Юрьевич, кандидат технических наук, старший научный сотрудник Санкт-Петербургского института информатики и автоматизации Российской академии наук. E-mail: shumskaya.oo@gmail.com

DOI: 10.17212/1814-1196-2018-3-121-134

Comparative study of classification methods in the stegoanalysis of digital images[*]

O.O. SHUMSKAYA[a], *V.Y. BUDKOV*[b]

St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences, 39, 14th Line V.O., St. Petersburg, 199178, Russian Federation

[a] *shumskaya.oo@gmail.com* [b] *visharmail@gmail.com*

Abstract

Everyone can see digital images – schedules, schemes, models, drawings, photos, logos and others – daily in all fields of human activity. Millions of people exchange images with each other in the Internet daily, without suspecting possible confidential contents hidden from a human eye in the file. Steganography is a science about the ways of transfer and storage of information which provide hiding the availability of this information in some signal. It provides various methods of concealment of data in digital images [1].

Methods of steganalysis which is a science about the ways of identification of the availability of hidden messages in digital objects are applied to reveal the availability of confidential information in digital files. Annually new methods of information embedding characterized by a bigger capacity and invisibility for a human eye are developed. However, the authors infrequently do research on the method tolerance to steganalysis. In papers where experiments on tolerance to steganalysis are described, one classification method is mainly applied, whose choice isn't validated experimentally. The research on the tpolerance as compared with various steganalysis methods and various qualifiers will allow studying a steganographic method from different perspectives and will help to increase embedding stability.

Well-known works on steganalysis based on machine learning methods are considered in the paper. Experiments on comparing various methods of classification and their variations to identify appropriate qualifiers are also described.

[*] *Received 13 June 2018.*

Keywords: qualifier, classification methods, informative features, Fisher linear discriminant, naive Bayesian qualifier, neural nets, AutoMPL, support vectors, steganalysis

REFERENCES

1. Kokhanovich G.F., Puzyrenko A.Yu. *Komp'yuternaya steganografiya: teoriya i praktika* [Computer steganography: theory and practice]. Kiev, MK-Press Publ., 2006. 288 p.

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 (10), pp. 5999–6019.

3. Shumskaya O.O. [Method of JPEG-images steganalysis on the basis of energy features in the frequency domain]. *Nauchnaya sessiya TUSUR-2017*: materialy mezhdunarodnoi nauchno-tekhnicheskoi konferentsii studentov, aspirantov i molodykh uchenykh [Proceedings of the international scientific-technical conference of students, graduate students and young scientists "Scientific session TSUCSR-2017"]. Tomsk, 2017, pt. 6, pp. 41–44. (In Russian).

4. *Fridrich J.* Feature-based steganalysis for JPEG images and its Implications for future design of steganographic schemes. *Information Hiding: 6th International Workshop, IH 2004*, Toronto, 2004, pp. 67–81.

5. Chen M.-C., Agaian S.S., Chen C.L.P., Rodriguez B.M. Alpha-trimmed image estimation for JPEG steganography detection. *2009 IEEE International Conference on Systems, Man, and Cybernetics*, San Antonio, Texas, USA, 2009, pp. 4581–4585.

6. Xia Zh., Wang X., Sun X., Liu Q., Xiong N. Steganalysis of LSB matching using differences between nonadjacent pixels. *Multimedia Tools and Applications*, 2016, vol. 75 (4), pp. 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)*, Beijing, China, 2015, pp. 63–71.

8. Liu Q., Sung A., Qiao M., Chen Z., Ribeiro B. An improved approach to steganalysis of JPEG images. *Information Sciences*, 2010, vol. 180 (9), pp. 1643–1655.

9 Kodovsky J., Fridrich J. Steganalysis of JPEG images using rich models. *Proceedings of SPIE, Electronic Imaging, Media Watermarking, Security, and Forensics XIV*, San Francisco, USA, 2012, pp. 7–20.

10. Zong H., Liu X., Luo X. Blind image steganalysis based on wavelet coefficient correlation. *Digital Investigation*, 2012, vol. 9, pp. 58–68.

11. Berg G., Davidson I., Duan M.-Y., Paul G. Searching for hidden messages: automatic detection of steganography. *Proceedings of the 15th Innovative Applications of Artificial Intelligence Conference*, August 12– 14, 2003, Acapulco, Mexico, pp. 51–56.

12. Maitra S., Paul G., Sarkar S.,Lehmann M., Meier W. New results on generalization of roos-type biases and related keystream of RC4. *Progress in Cryptology – AFRICACRYPT 2013: 6th International Conference on Cryptology in Africa: proceedings*, Cairo, Egypt, 2013, pp. 222–239.

13. Evsyutin O.O., Meshcheryakov R.V., Shumskaya O.O. [Steganalysis of digital images with use of the naive Bayesian qualifier]. *Materialy 10-i Vserossiiskoi mul'tikonferentsii po problemam upravleniya (MKPU-2017)* [Proceedings of the 10-th All-Russian multiconference on problems of management (MCPM-2017)]. Rostov-on-Don, Southern Federal University Publ., 2017, pp. 56–58. (In Russian).

14. Sabeti V., Samavi Sh., Mahdavi M., Shirani Sh. Steganalysis and payload estimation of embedding in pixel differences using neural networks. *Pattern Recognition*, 2010, vol. 43 (1), pp. 405–415.

15. Lubenko I., Ker A.D. Steganalysis with mismatched covers: do simple classifiers help? *MM&Sec '12: proceedings of the 14th ACM Multimedia and Security Workshop*, September 6–7, 2012. New York, NY, ASM, 2012, pp. 11–18.

16. Image Databases. *ImageProcessingPlace.com*. Available at: http://www.imageprocessingplace.com/root_files_V3/image_databases.htm (accessed 14.09.2018).

17 The USC-SIPI Image Database. *USC University of Southern California*. Los Angeles, CA, 2017. Available at: http://sipi.usc.edu/database (accessed 14.09.2018).

18. Bharathi A., Natarajan A.M. Cancer classification using support vector machines and relevance vector machine based on analysis of variance features. *Journal of Computer Science*, 2011, vol. 7 (9), pp. 1393–1399.

Для цитирования:

Шумская О.О., Будков В.Ю. Сравнительное исследование методов классификации в стегоанализе цифровых изображений // Научный вестник НГТУ. – 2018. – № 3 (72). – С. 121–134. – doi: 10.17212/1814-1196-2018-3-121-134.

For citation:

Shumskaya O.O., Budkov V.Yu. Sravnitel'noe issledovanie metodov klassifikatsii v stegoanalize tsifrovykh izobrazhenii [Comparative study of classification methods in the stegoanalysis of digital images]. *Nauchnyi vestnik Novosibirskogo gosudarstvennogo tekhnicheskogo universiteta – Science bulletin of the Novosibirsk state technical university*, 2018, no. 3 (72), pp. 121–134. doi: 10.17212/1814-1196-2018-3-121-134.