ДЕТЕКТИРОВАНИЕ НАЛИЧИЯ СКРЫТОЙ ИНФОРМАЦИИ В ЦИФРОВЫХ ИЗОБРАЖЕНИЯХ С ИСПОЛЬЗОВАНИЕМ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ

Кадан А М., Сазановец И. А.
Кафедра системного программирования и компьютерной безопасности, Гродненский государственный университет им.Янки Купалы
ООО «ИнтексСофт»
Гродно, Республика Беларусь
E-mail: kadan@mf.grsu.by, sazanovec_ia_13@mf.grsu.by

Для детектирования наличия скрытой информации в графических файлах рассматриваются методы на основе технологий машинного обучения. Детектирование ведется с использованием «слепых» методов – при отсутствии данных об исходном алгоритме, использованном для внедрения скрытой информации. Для для формирования датасетов, используемых при обучении моделей детектирования, представлены методы на основе вейвлет-разложения. Приведены результаты испытания обученных моделей на тренировочных наборах данных.

Введение

Потребность скрыть информацию или поделиться ею, не привлекая внимания посторонних, существовала на протяжении всей истории общественных отношений. Научное направление, изучающее вопросы скрытой передачи информации, известно как стеганография. Наметившийся в последние годы интерес к стеганографическим методам связан в немалой степени с тем, что в отличие от криптографии, их использование практически не регулируется законодательством.

Современные методы стеганографии широко используют компьютерную технику для внедрения «стегоинформации» в другие цифровые данные, называемые «стегоконтейнерами», с качестве которых могут выступать цифровые изображения, аудио- или видео-данные, пакеты сетевого трафика и многое другое.

Одним из ключевых требований к стеганографическим алгоритмам является то, что внедрение информации не должно заметно изменять характеристики стегоконтейнера. Поэтому стеганографические алгоритмы часто эксплуатируют ограниченность биологических систем восприятия человека. Например, при сокрытии информации в изображениях, стегоалгоритмы изменяют интенсивность цветов так, чтобы, с одной стороны, этими изменениями внедрить стегоинформацию, а с другой, чтобы эти изменения не воспринялись органами зрения человека. В основе стегоалгоритмов по работе со звуком лежит тот же принцип – записываемая информация изменяет высокие частоты аудиосигнала, что вряд ли будет заметно при прослушивании [1].

I. Схема и средства сокрытия информации

Алгоритмически, стеганография состоит из двух фаз: одна для сокрытия информации, другая для извлечения. На случай, если всё-таки обнаружится факт наличия скрытого сообщения, в большинстве стеганографических программ перед внедрением сообщения его вначале зашифровывают. Базовая модель стеганографии представлена на рисунке 1.

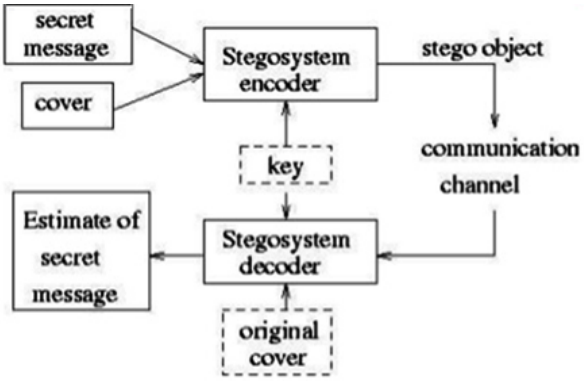

Рис. 1 – Базовая модель стегонографии

Актуальной задачей является детектирование наличия внедрённой информации при условии, что ничего не известно об исходном стегоалгоритме. В стегоанализе методы, решающие задачи такого типа, принято называть «слепыми».

II. Формирование изображений, содержащих скрытые данные

Алгоритмы стеганографии формируют разные типы искажений исходного контейнера. Поэтому написать детерминированный алгоритм для детектирования наличия стегоинформации становится едва ли возможным. Именно в таких ситуациях прибегают к использованию методов машинного обучения. Для обучения моделей машинного обучения были выбраны 750 графических изображений из открытых наборов данных, предварительно уменьшенные до размеров от 640x480 до 1147x768 пикселей. Стегоинформация была внедрена с помощью трех программ:

– Steganography Software F5 (алгоритм f5) [2];
– StegHide (стеганография на основе теории графов) [3];
– OpenStego (RandomLSB, модифицированный алгоритм наименьшего значащего бита) [4].

III. Формирование файлов признаков

Для использования моделей машинного обучения необходимо на основе исходных графических изображений сформировать датасет – файл признаков, который бы отражал характерные особенности «чистых» изображений и изображений, содержащих стегоинформацию. Датасет представлен в виде файла csv-формата. Каждая запись датасета содержит 84 признака, а также 85-й классификационный признак: «0» для «чистого» изображения и «1» в противном случае. В итоге был сформирован датасет из 3000 записей – 750 записей для «чистых» изображений, и по 750 записей для изображений со стегоинформацией, внедренной программами Steganography Software F5, StegHide и OpenStego соответственно.

Для построения по изображению вектора признаков, поскольку мы имеем дело с цифровыми изображениями, имеет смысл применять дискретные вейвлет-преобразования. Разложение будем производить по вейвлет-функциям Хаара, db2 и bior1.3.

Так как каждый цветовой канал в изображении представляется прямоугольной матрицей, то будем использовать дискретное вейвлет-преобразование, реализованное методом *wavedec*2 модуля *PyWavelets* языка Python.

Метод *wavedec*2 возвращает структуру вида [cAn, (cHn, cVn, cDn), ... (cH1, cV1, cD1)], где:

– cAn – аппроксимационный коэффициент разложения n-го уровня;
– cHn – горизонтальный коэффициент разложения n-го уровня;
– cVn – вертикальный коэффициент разложения n-го уровня;
– cDn – диагональный коэффициент разложения n-го уровня.

Признаки, характеризующие графическое изображение, будем формировать с использованием дискретных вейвлет-преобразований (с разложением не выше 3-го уровня) и статистических моментов 1-4 порядков.

IV. Применение методов машинного обучения

Задача слепого детектирования наличия стегоинформации в графическом изображении, рассматриваемая в данной работе, относится к задачам бинарной классификации и может быть решена в рамках технологии машинного обучения с учителем.

Был проведен эксперимент по применению для решения поставленной задачи следующих методов, с использованием их реализаций из библиотеки *scikit – learn* для языка Python:

– алгоритма K-ближайших соседей;
– наивного байесовского классификатора;
– дерева принятия решений;
– линейной регрессии;
– метода опорных векторов;
– нейронной сети прямого распространения.

Лучшие результаты удалось получить при использовании метода опорных векторов и нейронной сети прямого распространения с архитектурой «многослойный персептрон» с двумя скрытыми сломи. В случае использования вейвлет-функций *db*2 и *bior*1.3 результаты оказались хуже, чем при использовании вейвлет-функции Хаара.

Также они хуже при использовании *minmax*-нормализации вместо стандартной.

Наиболее существенно влияющим на результат было значение признаков, связанных с моментами 3-го и 4-го порядков. Признаки, связанные с моментами 1-го и 2-го порядка не оказывали существенного влияния на результат. Также существенно не повлияли на результат значения признаков, вычисленных на горизонтальных коэффициентах вейвлет-разложения.

Заключение

Можно утверждать, что достаточно эффективными методами для решения задачи «слепого дектирования» являются метод опорных векторов с параметрами C=1000, gamma=0.001 и kernel=rbf и многослойный персептрон с двумя скрытыми слоями (90 и 20 нейронов).

В качестве эффективных признаков могут быть использованы коэффициенты асимметрии и эксцесса, вычисленные в частотной плоскости для аппроксимационного, вертикальных и диагональных коэффициентов двумерного трёхуровневого вейвлет-преобразования с использованием вейвлет-функции Хаара. С помощью обученных моделей можно с вероятностью, близкой к 0.7, предсказать, содержит ли графическое изображение скрытое сообщение или нет.

Список литературы

1. David Wheeler. Audio Steganography Using High Frequency Noise Introduction [Электронный ресурс] / RIT Scholar Works. 2012. – Режим доступа: https://pdfs.semanticscholar.org/d547/3318c5c9171fe38abc550b89a15022d559cb.pdf. – Дата доступа: 28.09.2019.
2. F5-steganography [Электронный ресурс] / The world's leading software development platform GitHub. – Режим доступа: https://github.com/matthewgao/F5-steganography. – Дата доступа: 28.04.2019.
3. Steghide [Электронный ресурс] / Sourceforge. – Режим доступа: http://steghide.sourceforge.net/. – Дата доступа: 28.04.2019.
4. OpenStego [Электронный ресурс] / OpenStego. – Режим доступа: https://www.openstego.com/. – Дата доступа: 28.04.2019.