СРАВНЕНИЕ МЕТОДА ОПОРНЫХ ВЕКТОРОВ С МЕТОДОМ ГЕНЕРАТИВНО-СОСТЯЗАТЕЛЬНЫХ СЕТЕЙ В СТЕГАНОАНАЛИЗЕ

Барановский Г.В., Бекарев С.С., Гулис А.А, Шишов Е.П.

Белорусский государственный университет информатики и радиоэлектроники
г. Минск, Республика Беларусь

Стройникова Е.Д. – старший препод. каф. информатики

Рассмотрены понятия стеганографии и стеганоанализа. Внимание акцентировано на двух конкретных методах стеганоанализа: методе опорных векторов и методе генеративно-состязательных сетей, их преимуществах и недостатках.

Ключевые слова. Стеганография, стеганоанализ, стегоконтейнер, SVM, GNA, LSB, задача классификации, Machine Learning.

Введение

Стеганогра́фия — способ передачи и/или хранения информации с учётом сохранения в тайне самого факта такой передачи. В отличие от криптографии, которая скрывает содержимое тайного сообщения, стеганография скрывает сам факт его существования.

Понятие стеганоанализа

Стеганоанализ — раздел стеганографии изучающий способы выявления тайно передаваемой информации в анализируемом информационном объекте. Под тайно передаваемой информацией обычно подразумевается информация, скрытая теми или иными стеганографическими методами. Стеганоанализ имеет широкое применение в современном мире, например, в мультимедийных системах для обеспечения безопасности авторских прав и защиты цифрового контента.

Для обнаружения информации, скрытой с помощью стеганографических методов, разработан ряд методов стегоанализа. В данной работе будет рассмотрено два метода : метод опорных векторов (*Support Vector Machine*) и метод генеративно-состязательных сетей (*Generative Adversarial Network*).

Метод опорных векторов (SVM)

Метод опорных векторов (SVM) — это алгоритм машинного обучения, который может использоваться для классификации, регрессии и других задач машинного обучения. В основе SVM лежит поиск гиперплоскости в *n*-мерном пространстве (где *n* — количество признаков), которая наилучшим образом разделяет данные на разные классы.

Рассмотрим бинарную классификацию, когда мы пытаемся разделить данные на два класса, например зеленые точки и синие точки на плоскости. Метод опорных векторов ищет гиперплоскость таким образом, чтобы она была максимально удалена от ближайших точек каждого класса. Эти ближайшие точки называются опорными векторами.

В стеганоанализе SVM используют для построения модели, которая будет разделять стеганографически изменённые файлы от их оригинальных версий. Обычно для обучения модели используются пары “чистых” и "скрытых" файлов. Оригинальный файл считается "чистым", а изменённый файл (в котором скрыта дополнительная информация) считается "скрытым". Затем модель обучается на этих парах файлов с помощью SVM. После обучения модель может использоваться для определения, является ли новый файл стеганографически изменённым или нет. Для этого изображение (аудио или видео) преобразуется в набор признаков, которые используются для классификации моделью SVM. Если файл классифицируется как "скрытый", то это означает, что он содержит скрытую информацию.

Плюсы метода опорных векторов:

1. Метод опорных векторов обладает высокой точностью при обнаружении скрытой информации в изображениях.
2. Позволяет рассматривать различные виды нелинейности, изменяя ядра или их параметры.
3. Максимизирует разделяющую полосу, которая позволяет уменьшить количество ошибок классификации.

Минусы метода опорных векторов:

1. Требует большого количества вычислительных ресурсов и времени для обучения модели на больших объемах данных.
2. Может быть чувствителен к шуму и выбросам в данных, что может снизить точность обнаружения скрытой информации.
3. Не описаны общие методы построения ядер, наиболее подходящих для конкретной задачи в случае линейной неразделимости классов.

Метод генеративно-состязательных сетей (GAN)

В стеганографии GAN (Generative Adversarial Network) может использоваться для создания стеганографических изображений, которые могут скрыть информацию внутри изображения.

Генеративная модель GAN состоит из двух компонентов: генератора и дискриминатора. Генератор создаёт фальшивые изображения, а дискриминатор классифицирует, являются ли изображения реальными или фальшивыми. Оба компонента обучаются вместе, и цель состоит в том, чтобы генератор создавал фальшивые изображения, которые дискриминатор не может отличить от реальных.

В стеганографии генератор может использоваться для создания изображений, которые содержат скрытую информацию, например текст. Этот текст может быть добавлен в пиксели изображения, изменяя их значения так, чтобы скрытая информация не была видна невооруженным глазом. Дискриминатор может затем использоваться для проверки, является ли изображение, содержащее скрытую информацию, нормальным изображением или нет.

Плюсы метода GAN:

1. Высокая скрытность: GAN может создавать непредсказуемые паттерны и текстуры, которые могут быть использованы для скрытого хранения данных.
2. Устойчивость к атакам: при использовании GAN в качестве метода стеганографии обнаружение скрытой информации может быть очень трудным.
3. Большой объём данных: GAN может использоваться для создания большого количества скрытых данных и информации, что может быть полезно для хранения больших объёмов конфиденциальной информации.

Минусы метода GAN:

1. Сложность создания: создание хорошо работающей модели GAN для стеганографии может быть очень трудным и требовательным к ресурсам.
2. Сложность обнаружения: в случае обнаружения скрытой информации может быть очень сложно понять, как она была спрятана и как её можно извлечь.

Заключение

Сравнивая методы GAN и SVM, было установлено, что GAN эффективен в генерации новых данных с высокой степенью реалистичности, скрытности и устойчивости к атакам, но менее эффективен в обнаружении скрытых сообщений в существующих данных. SVM более эффективен в обнаружении скрытых сообщений в существующих данных, но требует большого количество вычислительных ресурсов для больших данных, чувствителен к шуму в данных.

Список использованных источников:

1. Christopher M. Bishop. Pattern recognition and machine learning, 2006. – с.402 [Электронный ресурс] URL: https://inlnk.ru/oe57QN.- Дата доступа(20 апреля 2023)

2 .А. В. Бычков, Алгоритмы синтеза изображений в больших разрешениях на основе генеративно-состязательных нейронных сетей, 2020, 53 с.[Электронный ресурс] URL:https://inlnk.ru/3ZMLKz.- Дата доступа(20 апреля 2023)

3. К.В.Воронцев, Лекции по методу опорных векторов, с.2-5, 12-13, 2019 [Электронный ресурс] URL:http://www.ccas.ru/voron/download/SVM.pdf.- Дата доступа(20 апреля 2023)

4.The A-Z guide to Support Vector Machine, 2021 [Электронный ресурс] URL : https://inlnk.ru/w4y6Y1.- Дата доступа(20 апреля 2023)

5.Стегоанализ как поиск скрытых сообщений [Электронный ресурс] URL : https://inlnk.ru/meL3xk .- Дата доступа(20 апреля 2023)