

УДК 004.08

СРАВНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ЗАДАЧИ ВЫЯВЛЕНИЯ ТЕКСТОВОЙ СТЕГАНОГРАФИИ

Хазагаров А.А. (Университет ИТМО, Санкт-Петербург)

Научный руководитель – к.т.н., Воробьева А.А.

(Университет ИТМО, Санкт-Петербург)

В работе рассматривается сравнение методов машинного обучения, способы векторизации и извлечение признаков из текста для выявления применённых методов текстовой стеганографии.

Введение. Применение методов текстовой стеганографии может применяться для противоправных действий. Например, скрытые утечки информации. Повсеместная цифровизация, где практически вся информация хранится, обрабатывается и передается в цифровом виде, усугубляет проблему. Специалистам информационной безопасности необходимо своевременно реагировать на утечки, проводить исследование подобных инцидентов и определять источник утечки информации, что может являться трудноразрешимой задачей. Одной из актуальных проблем является невысокая точность выявления применения к тесту методов стеганографии.

Основная часть. Наибольшей популярностью для работы с текстом пользуются рекуррентные нейронные сети. Среди них выделяют RNN, LSTM и GRU. Они себя хорошо зарекомендовали в лингвистическом анализе, в том числе определение тональности текста, машинном переводе и многих других направлениях. Однако, классические методы машинного обучения в некоторых задачах сопоставимы с нейронными сетями при меньшем потреблении ресурсов. Также, важную часть при работе с текстом является векторизация и способ извлечения признаков. Во многом, от такого, какие из этих параметров будут выбраны, зависит точность и производительность алгоритма. В данной работе мы постараемся найти, оптимальные параметры для поставленной задачи.

Выводы. В качестве результатов можно выделить, что рекуррентная нейронная сеть типа LSTM и n-граммы показали наилучший результат по точности и потреблению ресурсов для выявления применения текстовой стеганографии.

Хазагаров А.А. (автор)

Подпись

Воробьева А.А. (научный руководитель)

Подпись