УДК 004.891.3

КЛАССИФИКАЦИЯ ТЕКСТОВЫХ ДОКУМЕНТОВ ДЛЯ СИСТЕМ ПОИСКА УТЕЧЕК ИНФОРМАЦИИ С ИСПОЛЬЗОВАНИЕМ ТЕХНОЛОГИИ ПЕРЕДАЧИ ОБУЧЕНИЯ

Зайцева Е. Г. (Университет ИТМО)
Научный руководитель – кандидат технических наук, доцент Спивак А. И.
(Университет ИТМО)

В докладе рассмотрен один из подходов к классификации текстовых документов на предмет их конфиденциальности, основанный на использовании предобученной нейронной сети.

Введение. На настоящий момент деятельность любой организации не обходится без работы с конфиденциальными данными. Исследования компании InfoWatch об утечках информации ограниченного доступа за первые 9 месяцев 2020 года показывают, что в России более 79% утечек произошли по вине внутренних нарушителей Любая утечка приносит компании как финансовый, так и материальный ущерб. Одним из самых эффективных методов защиты конфиденциальной информации от утечки в настоящее время являются DLP-системы. Настоящая работа посвящена разработке алгоритма классификации текста для DLP-системы, построенного на основе нейронной сети, с целью улучшить качество классификации, тем самым повысив эффективность системы поиска утечек информации.

Основная часть. BERT (Bidirectional Encoder Representations from Transformers) – это метод машинного обучения, предназначенный для предварительной подготовки к обработке естественного языка (NLP), разработанный компанией Google. Использование данной модели основано на технологии Transfer-learning: модель заранее предобучена на большом количестве неразмеченных данных, извлеченных из корпуса книг и Wikepedia, что позволяет нейронной сети получить «общее представление о языке». После этого модель необходимо обучить непосредственно под конкретную задачу. В настоящей работе основной задачей являлось распознавания конфиденциальных сообщений. Для этого на вход предобученной модели подавался размеченный датасет, содержащий конфиденциальную и неконфиденциальную информацию. В процессе дообучения модель учится распознавать конфиденциальную информацию с использованием знаний о языке, полученных на этапе предобучения.

Выводы. В настоящей работе представлен алгоритм для эффективного распознавания конфиденциальной информации в DLP-системах с использованием технологии передачи обучения. Использование данной технологии в нейронной сети BERT, предобученной на большом количестве текстов, позволяет достичь высокого значения метрики F1-score в задаче классификации конфиденциальной информации.

Зайцева Е. Г. (автор)

Спивак А. И. (научный руководитель)