УДК 004.08

ПРОТИВОДЕЙСТВИЕ УТЕЧКАМ ИНФОРМАЦИИ С ПОМОЩЬЮ МЕТОДОВ ЛИНГВИСТИЧЕСКОЙ ИДЕНТИФИКАЦИИ

Хазагаров А.А. (Университет ИТМО), Воробьева А.А. (Университет ИТМО)

В данной работе рассматривается проблема утечки информации. Для решения этой задачи применяется подход к определению авторства сообщений источника утечки на основе глубоких нейронных сетей. Также рассматривается проблема представления текста и выделения лингвистических и стилистических признаков сообщений.

Утечки информации – проблема для множества различных компаний и учреждений. Она вызвана повсеместной цифровизацией, где практически вся информация хранится, обрабатывается и передается через электронные устройства. Анонимность интернет-коммуникаций только увеличивает актуальность этой проблемы, делая возможным анонимное распространение конфиденциальной информаций за пределы организации, что может нанести серьезный ущерб. Специалистам информационной безопасности необходимо своевременно реагировать на утечки, проводить исследование подобных инцидентов и определять источник утечки информации, что может являться трудноразрешимой задачей. Применение методов лингвистической идентификации может быть полезно для дианонимизации нарушителя и раскрытия киберпреступлений.

Целью данной работы является разработка метода противодействия утечкам информации с помощью методов лингвистической идентификации на основе лингвистических и стилистических особенностей сообщений. В работе будет рассматриваться применение глубоких нейронных сетей для решения поставленной проблемы, а также предобработка текста и выделение информативных признаков.

Для проведения эксперимента было написано программное обеспечение для сбора статей из веб-ресурсов habr.com, vk.com и использовались готовые наборы данных Эхо Москвы, LiveJournal, и Twitter. Рассмотрены способы выделения признаков, а также сравнение нескольких архитектур нейронных сетей.

Результаты работы нейронных сетей разделяются на две категории по выходному слою: определение того, что сообщение принадлежит конкретному пользователю и к какому конкретному пользователю принадлежит сообщение. В первом эксперименте в качестве инициализирующего слоя подавалось количество найденных признаков в тексте. Всего выделено 445 признаков, и они подразделяются они на лексический уровень символов, лексический уровень слов и синтаксический уровень. Далее идут 5 полносвязных LTSM слоя и выходной слой с сигмоидальной функцией активации. При использовании корпуса текста из 30 авторов были получены результаты: вероятность корректной идентификации 84% при выходном слое, определяющего, принадлежит ли сообщение пользователю; 40% – при определении авторства сообщений. При использовании векторизации One hot encoding и токинизации на уровне слов для того же набора данных были достигнуты результаты: 89% и 73% соответственно.

В работе было рассмотрено применение LTSM сетей для определения авторства сообщения, а также способы извлечение признаков. Были рассмотрены два варианта сетей, отличающихся выходной функцией. В дальнейшем планируется рассматривать только сети с логической выходной функцией, так как такие сети значительно легче по архитектуре, у них выше точность, скорость функционирования и не требует переобучения при добавлении пользователя.