Л. А. ВИТКОВА
Санкт-Петербургский Федеральный исследовательский центр Российской Академии Наук
Санкт-Петербург

АНАЛИЗ ПРОФИЛЕЙ В СОЦИАЛЬНЫХ СЕТЯХ В ЦЕЛЯХ ИНФОРМАЦИОННОЙ БЕЗОПАСНОСТИ

В современных информационных войнах почти всегда используются боты в социальных сетях. В процессе распространения информации могут быть задействованы копии оригинальных профилей, вновь созданные страницы. Злоумышленник умышленно создает искусственный профиль схожий с настоящим, с целью обмануть аудиторию, эксперта. В работе предложены дискретные признаки объектов профилей и алгоритм сбора, маркировки дат создания объектов для обнаружения ботов в социальной сети.

Введение. В 21 веке социальные сети постепенно вошли в жизнь общества и к 20-м годам стали неотъемлемой частью человека. Страница в социальной сети (профиль) зачастую является неотъемлемым атрибутом социального взаимодействия. Однако параллельно с цифровизацией межличностных коммуникаций появлялись и развивались информационные угрозы. Одной из таких угроз являются боты, в том случае, если они используются в различных информационных атаках. Сложность обнаружения ботов, то есть искусственно управляемых аккаунтов, велика и растет тогда, когда профиль в социальной сети соответствует экспертному представлению страницы живого человека [1, 2]. По сути, принятие экспертом правильного решения может быть невозможно тогда, когда профиль полностью копирует другой. Так, например злоумышленник может скопировать страницу в одной социальной сети и создать схожую в другой. Эксперт не сможет распознать копию, так как он не видел оригинал. При этом, существующие модели и подходы могут обнаружить бота при условии, если профиль пустой, например совсем не заполнен. В таком случае эксперт также распознает бота. Однако же эксперт без дополнительных признаков не в состоянии отличить поддельный профиль от настоящего в случае создании копии [3, 4]. Задача текущего исследования состоит в том, чтобы выработать и предложить дискретные признаки бота, которые позволят эксперту или модели машинного обучения с большой точностью обнаруживать ботов в социальных сетях.

Дискретные признаки. Возьмем в качестве примера профиль в социальной сети «Вконтакте». Профиль пользователя сети Вконтакте содержит множество характеристик, среди которых: (1) идентификатор профиля; (2) ФИО; (43) дата рождения, пол, город проживания, работа и прочее; (4) фотоальбом; (5) записи на стене профиля (пост, комментарий). Эксперт может видеть информацию, однако он не всегда имеет доступ к дискретным признакам. Рассмотрим таблицу.

Таблица

Дискретные признаки объектов профиля

Объект	Дискретный признак
Идентификатор профиля	Дата создания (доступна при сборе данных)
Фотография в фотоальбоме	Дата создания медиа файла в альбоме
Раздел в фотоальбоме	Дата создания раздела в фотоальбоме
Пост на стене	Дата создания
Комментарий	Дата создания

В таблице предложены объекты профиля в социальной сети, у которых сохраняется информация о дате создания. Эксперт может не иметь доступа к анализу дат регистрации идентификаторов, дат создания (загрузки фотографий, публикации постов). Однако такие признаки могут быть учтены в моделях машинного обучения. При этом, сами по себе даты создания так или иначе различны, важна близость между дискретными признаками объектов. Для ботов, например, характерно создание и заполнение профиля в короткий срок, тогда как для живого человека характерно длительное управление своим аккаунтом и постепенное заполнение страницы в социальной сети.

Алгоритм сбора и маркировки дат.

В ходе исследования был разработан алгоритм сбора и маркировки дат регистрации идентификаторов в социальной сети. Алгоритм состоит из двух частей.

На первом этапе производится сбор дат регистрации профилей на основе идентификатора аккаунта:

Шаг 1. Извлечение идентификатора аккаунта из набора данных.

Шаг 2. Вызов процедуры для извлечения дат по id профилей в список. Даты идентификаторов собираются с использованием запроса «vk_link», который обращается по адресу: «https://vk.com/foaf.php?id=(идентификатор аккаунта). А также с использованием запроса «parsed_xml», который извлекает со старицы следующие параметры: а) дата регистрации, б) время регистрации.

Шаг 3. Запись даты и времени регистрации в набор данных.

На втором этапе производится анализ профилей в социальных сетях по признаку «дата регистрации аккаунта» и добавляется дополнительный признак:

Шаг 4. Разметка дат по двум категориям:

0 – уникальные даты регистрации;

1 – повторяющиеся даты регистрации.

На выходе получается набор данных с заполненными датами регистрации идентификаторов аккаунтов с дополнительным признаком (уникальность даты в наборе).

Предложенный алгоритм маркировки дат регистрации применим также к другим схожим дискретным признакам, таким как: дата создания медиа файла, создание поста, комментария. Признак схожести даты создания объекта не является 100 % гарантией того, что аккаунт искусственно управляется злоумышленником и не может быть основой для утверждения «аккаунт-бот». Однако отклонение от статистической нормы может являться одним из основных признаков для обнаружения бота.

Заключение. Повсеместная цифровизация, погружение в социальные сети, рост количества информационных атак ставят перед специалистами в области информационной безопасности, перед учеными-исследователями новые задачи. Сегодня множество угроз реализуется с использованием ботов. В работе рассматриваются дискретные признаки, такие как дата создания объекта: пост, комментарий, медиа файл, аккаунт. И предлагается алгоритм сбор дат создания (на примере дат регистрации аккаунтов) и маркировки. В дальнейшем планируется провести экспериментальные исследования с использованием наборов данных из разных социальных сетей.

Работа выполнена при финансовой поддержке Гранта РНФ № 18-71-10094-П в СПб ФИЦ РАН.

ЛИТЕРАТУРА

1. **Kolomeets M., Chechulin A, Kotenko I.** Bot detection by friends graph in social networks // Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA). 2, 12. 2021. pp. 141-159. DOI: 10.22667/JOWUA.2021.06.30.141. URL: http://isyou.info/jowua/papers/jowua-v12n2-6.pdf.
2. **Виткова Л.А.** Анализ моделей данных популярных социальных сетей // В сборнике: Актуальные проблемы инфотелекоммуникаций в науке и образовании. сборник научных статей: в 4х томах. Санкт-Петербургский государственный университет телекоммуникаций им. проф. М.А. Бонч-Бруевича. Санкт-Петербург, 2021. С. 160-163.
3. **Gavra D., Namyatova K., Vitkova L.** Detection of induced activity in social networks: model and methodology. *Future Internet*. 2021. Т. 13. № 11.
4. **Логинова А.О.** Анализ существующих подходов к классификации и типологии ботов. *Инновационные технологии: теория, инструменты, практика*. 2020. Т.1. С. 462-467.

L.A.Vitkova (St. Petersburg Federal Research Center of the Russian Academy of Sciences, St. Petersburg)

Analysis of social media profiles for information security

In modern information wars, bots in social networks are almost always used. Copies of the original profiles and newly created pages may be used in the process of distributing information. The attacker deliberately creates an artificial profile similar to the real one, in order to deceive the audience, or expert. The paper proposes discrete features of profile objects and an algorithm for collecting, marking the dates of creation of objects for detecting bots in a social network