М. В. КОЛОМЕЕЦ
Санкт-Петербургский Федеральный исследовательский центр Российской академии наук,
Санкт-Петербург

МЕТОДИКА РАЗМЕТКИ МЕТРИК БОТОВ СОЦИАЛЬНЫХ МЕДИА

В данной работе предлагается методика аннотации наборов ботов на основе закупки, которая в отличие от аналогов позволяет не только получить достоверные (ground-truth) лейблы ботов, но и получить ряд характеристик наборов ботов, подходящих для задачи обучения моделей характеризации атак в социальных сетях. В работе также приводится экспериментальная оценка предлагаемой методики на основе социальной сети ВКонтакте, а также корреляционный анализ характеристик ботов, собранных и аннотированных с использованием предлагаемой методики.

Введение. Вредоносные боты являются одним из основных инструментов злоумышленников для проведения атак в пространстве социальных медиа. По этой причине, тематика обнаружения ботов, расчет рисков, а также противодействия атакам в социальных сетях стала одной из актуальных областей информационной безопасности. Для этого исследователи разрабатывают различные методы анализа ботов, большинство которых [1] основаны на методах обучения с учителем. Ключевой проблемой данных методов является проблема формирования корректных наборов данных для обучения [1], так как для аннотации наборов чаще всего привлекаются эксперты [2, 3], способности которых обнаружить и корректно охарактеризовать бота не однозначны, и являются предметом дискуссии [1]. Целью настоящего исследования было разработать методику получения таких наборов, которые не только содержали бы ground-truth лейблы бот/не бот, но и ряд характеристик ботов, с использованием которых можно было бы оценить атаку и выбрать правильную стратегию противодействия. Для этого предлагается формировать наборы методом закупки у продавцов [4] ботов в искусственно созданные сообщества в контролируемых исследователем условиях.

Методика разметки на основе закупки. Разработанная методика основана на создании фейкового сообщества, в которое, в контролируемых исследователем условиях, закупаются боты и исключаются реальные пользователи. Данная методика состоит из 5 шагов:

1. Создание списка продавцов ботов и их услуг с указанием описания набора ботов от продавца. Как правило, один продавец предлагает на выбор ботов нескольких разных качеств. Список включает:
    a. название продавца;
    b. качество ботов, определяемое экспертом на основании описания продавца;
    c. тип продавца: магазин ботов или биржа ботов; магазин, это площадка, где покупатель приобретает активность у одного единственного продавца, который имеет некоторые наборы ботов различных характеристик; биржа, это площадка, где покупатель размещает заказ, а продавец ботов за комиссию от сделки публикует заказ на своей бирже, где любой человек может его выполнить за деньги – такие боты чаще используются для генерации сложного текстового контента;
    d. вид активности – лайк, комментарий и т. д. (зависит от вида социального медиа и предложения продавца);
    e. цена за единицу активности – в рублях на день покупки.
2. Создание фейкового сообщества и наполнения его контентом. Сообщество должно удовлетворять двум условиям:
    a. сообщество должно выглядеть настоящим, чтобы не вызвать у продавца ботов подозрений. Для этого сообщество наполняется контентом (фотографии, посты, другие фейковые пользователи). Некоторые продавцы ботов отказывают в услугах, если в сообществе состоит мало людей, либо у сообщества низкая активность.

   b. сообщество не должно быть привлекательным для реального пользователя, чтобы исключить вероятность того, что реальные пользователи проявят активность в сообществе. Для этого общество должно обладать абсурдной / непривлекательной тематикой (например, осуществление перевозок между несуществующими городами и т. п.).
3. Закупка в сообщество партии ботов (один продавец, одно качество) и сохранение id аккаунтов, осуществивших активность, с сохранением метки продавца и качества для данной партии. При закупке продавцу дается задание, например – поставить N лайков посту X. По завершении закупки также размечается скорость ботов, как разница во времени между моментом оплаты и завершением задачи. Боты, управляемые программными средствами, могут совершить атаку моментально (завершить задание, данное продавцу ботов в течение нескольких секунд). А боты, имитирующие естественное поведение или управляемые людьми, совершают атаку в течение часа или суток. Скорость размечается как категориальная мера:
   a. задание завершено моментально – менее минуты;
   b. задание завершено за несколько часов;
   c. задание завершено за сутки и более.
4. Удаление активности в сообществе.
5. Повторение шага 3 для других продавцов ботов и предоставляемых ими категорий качеств.

В результате формируется набор данных, состоящий из ботов, который имеет следующие метки:

1. список id аккаунтов ботов – позволяет собрать данные для формирования признаков для обучения;
2. название магазина – для идентификации продавца и возможного дальнейшего уточнения отдельных характеристик ботов;
3. качество ботов, определенное экспертом – на основании описания продавца по шкале [НИЗКОЕ, СРЕДНЕЕ, ВЫСОКОЕ];
4. тип продавца – выражает стратегию управления ботов и включает 2 стратегии: магазин или биржа.
5. вид активности – какая задача ставилась продавцу (лайк, комментарий и т. д.);
6. цена, уплаченная за бот, в рублях – выражает стоимость атаки;
7. скорость – задание завершено моментально / за несколько часов / около суток и более. Выражает скорость атаки, которая зависит от сложности поведения ботов.

Экспериментальная оценка методики и корреляционный анализ. С использованием данной методики были собраны 70 размеченных наборов ботов от 30 компаний в социальной сети ВКонтакте. Экспертное качество было размечено тремя экспертами. В совокупности были собрано 22325 аккаунтов ботов, из которых 18444 аккаунтов являются уникальными. Разница между числом собранных ботов и числом уникальных аккаунтов возникает из-за того, что некоторые продавцы ботов не создают своих собственных ботов, а используют другие магазины как источники ботов, либо собирают наборы ботов для выполнения задания сразу из нескольких источников. Таким образом, был сформирован набор данных, который содержит следующие характеристики, представленные в таблице: цена, скорость, экспертное качество, тип продавца. Также, на основе сканирования аккаунтов ботов была извлечена характеристика заполненности профиля (не является частью методики, и представлена для корреляционного анализа).

Таблица

Характеристики ботов, полученные методом закупки

Характеристика	Категориальная мера	Диапазон значений
Цена	-	$[0, \infty]$
Тип продавца	{магазин, биржа}	{0, 1}
Скорость	{моментально, час, сутки}	{0, 1, 2}
Экспертное качество	{низкое, среднее, высокое}	{0, 1, 2}
Заполненность	-	[0, 1]

На основе разметки собранных наборов данных, включая разметки от 3ех экспертов, была рассчитана корреляция Спирмена между метриками наборов. Целью анализа является определение возможной мультколинеарности между характеристиками ботов, наличие которой укажет на взаимозаменяемость метрик. Данный корреляционный анализ актуален для социальной сети ВКонтакте на 2022 год. Его результаты представлены на рисунке в виде матрицы и позволяют сделать ряд выводов:

1) Корреляция между качеством эксперта и прочими характеристиками выражает то, на что акцентировал внимание эксперт в ходе разметки. Например, эксперт 3 имеет весьма высокую корреляцию со скоростью, поэтому его ответы не информативны.
2) Наборы данных, в которых много незаполненных аккаунтов (пустых/заблокированных), чаще приобретены в магазинах, а не на биржах, а боты в них обладают высокой скоростью. Заполненность умеренно коррелирует с прочими характеристиками (по шкале Чеддока).
3) Цена, скорость и тип продавца имеют слабую либо умеренную взаимную корреляцию, что говорит об отсутствии мультиколинерарности данных характеристик.

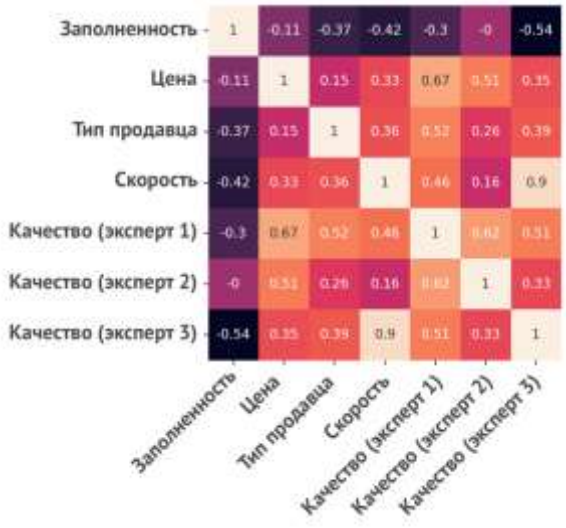

Рисунок. Корреляция метрик ботов

Заключение. Была разработана методика закупки ботов, которая позволяет получить размеченные наборы данных с ground-truth лейблом бот / не бот и рядом характеристик наборов ботов – экспертное качество бота, тип продавца, скорость, цена. С использование данной методики были собраны и размечены 70 наборов ботов от 30 компаний (22325 аккаунтов ботов, из которых 18444 аккаунтов являются уникальными). В результате эксперимента получилось построить метрики характеристик для наборов ботов и провести корреляционный анализ, который показал, что метрики обладают достаточно высокой информативностью, что позволяет на основе наборов данных с полученными метриками обучать модели, которые могут прогнозировать цену атаки, скорость, тип продавца ботов и качество используемых в атаке ботов.

Работа проводилась при поддержке гранта РНФ 18-71-10094-П.

ЛИТЕРАТУРА

1. **Orabi M., Mouheb D., Al Aghbari Z., Kamel I**. Detection of Bots in Social Media: A Systematic Review. *Information Processing & Management*. 2020. Vol. 57. № 4.
2. **Subrahmanian V.S., et al**. The DARPA Twitter bot challenge. *Computer*. 2016. Vol. 49. № 6. P. 38-46.
3. **Igawa R.A., Barbon Jr S., Paulo K.C.S., Kido G.S., Guido R.C., Júnior M.L.P., da Silva I.N**. Account classification in online social networks with LBCA and wavelets. *Information Sciences*. 2016. Vol. 332. P. 72-83.

4. **Kolomeets M., Chechulin A**. Analysis of the Malicious Bots Market. *29th Conference of Open Innovations Association (FRUCT)*. 2021. Vol. 29. P. 199-205.

M.V.Kolomeets (St. Petersburg federal research center of the Russian academy of science, St. Petersburg)

Technique for labeling of social media bot metrics

In this paper, we propose a technique for labeling bot datasets based on the purchase method. These method in comparison with analogs allows one to obtain ground-truth bot labels and a number of bot characteristics, that can be used in training of social media attacks characterization models. The paper also provides an experimental evaluation of the proposed technique based on the analysis of social network VKontakte, as well as a correlation analysis of the bot metrics collected and labeled with the proposed methodology.

Авторы готовы представить текст на английском языке для сборника материалов мультиконференции, который будет подан для индексирования в Scopus.