DOI 10.37882/2223-2966.2025.03.27

ОБНАРУЖЕНИЕ АКТОВ ИНТЕРНЕТ-АГРЕССИИ В СОЦИАЛЬНЫХ СЕТЯХ С ПОМОЩЬЮ АЛГОРИТМОВ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ

INTERNET AGGRESSION ACTS DETECTION IN SOCIAL NETWORKS USING DATA MINING ALGORITHMS

I. Krepak

Summary. IT development, in addition to an enormous benefit to humanity in solving every day and entertainment problems, has brought many unsolved problems due to their incorrect use. Most of these challenges are beyond the law and are cybercrimes. One of the types of cybercrime, which is currently one of the most active, is Internet bullying. Acts of cyberbullying have undergone many changes and now not only children, but also teenagers and adults become victims of Internet aggression. Social networks are both mass media and platforms where computer aggression often occurs and is almost uncontrolled (difficult to suppress). This article is devoted to the analysis of the frequency of background acts of Internet bullying in the network segment of the CIS. Collection and processing of the results took place in several stages, such as the formation of a knowledge base, preliminary processing, statistical analysis of the importance of keywords in the context of text files, data verification, and their classification. The results showed a high prevalence of the problem and a high frequency of background acts of computer aggression.

Keywords: internet crime, cyberbullying, automation, social network, telecommunications, data mining, machine learning, API, database, web page.

Крепак Иван
аспирант, Финансовый университет при Правительстве РФ, г. Москва
krepak.2311@yandex.ru

Аннотация. Развитие информационных технологий кроме колоссальной пользы человечеству в решении бытовых и развлекательных задач ввиду их некорректного применения принесли большое количество до сих пор нерешённых проблем. Большинство таких вызовов находятся за гранью закона и являются киберпреступлениями. Одним из видов киберпреступности, который в настоящее время одним из самых активных, является Интернет травля. Акты кибербуллинга перетерпели большое количество изменений и теперь жертвами Интернет агрессией становятся не только дети, но и подростки с взрослыми. Социальные сети — это одновременно средства массовой информации и платформы, где компьютерная агрессия часто происходит и почти не контролируется (сложно подавить). Данная статья посвящена анализу частоте фоновых актов Интернет травли в сетевом сегменте СНГ. Сбор и обработка результатов происходили в несколько этапов, такие как формирование базы знаний, предварительная обработка, статистический анализ важности ключевых слов в контексте текстовых файлов, проверка данных и их классификация. Результаты показали большую распространённость проблемы и большую частоту фоновых актов компьютерной агрессии.

Ключевые слова: интернет-преступность, интернет-травля, автоматизация, социальная сеть, телекоммуникации, интеллектуальный анализ данных, машинное обучение, API, база данных, веб-страница.

Введение

Интернет травля — один из основных видов киберпреступности, не требующих высокой ИТ квалификации или фундаментальных технических знаний. Кибербуллинг представляет собой последовательность злонамеренных действий агрессора с помощью информационных технологий, направленных на дестабилизацию психоэмоционального состояния жертвы. Манипуляции производятся через электронную почту, телефонные звонки, СМС, онлайн чаты, социальные сети, форумы и другие электронные каналы связи. В данной работе рассматривается Интернет агрессия, возникающая в социальных сетях. В качестве основного способа анализа контента и текста применяется наивный байесовский классификатор. Он очень доступен в реализации, относительно прост в эксплуатации, имеет высокую скорость обучения и дальнейшей целенаправленной классификации. Кроме этого, Интернет-травля — определяется как систематическое применение цифровых устройств с выходом в Интернет для запугивания жертвы через отправку сообщений, публикацию компрометирующих постов или другие аналогичные акты. Обычно, акты Интернет агрессии осуществляются через электронные письма, социальные сети и сервисы обмена быстрыми сообщениями.

Накопление данных

Накопление данных — это применение совокупности алгоритмов для определения закономерностей из заранее подготовленных данных для формирования знаний, которые в дальнейшем будут применяться в процессах принятия решений [1, с. 1]. Накопление данных включа-

ет в себя формирования, корреляций в криминалистических данных, обнаружение, сортировку дампа данных в группы по признаку сходства, обнаружение группы скрытых фактов и определение закономерностей информации, которая может привести к полезным предсказаниям (прогнозирование).

Процесс дата майнинга

Отбор данных из оперативной базы данных выполняется сразу после получения первого набора информации [2, с. 70]. Результаты отбора записываются в отдельный файл. Происходит более детальная очистка — удаляются дубликаты и противоречивые записи. Автоматически справляются грамматические ошибки и опечатки. Следующим этапом происходит кодирование. Это процесс преобразования информации в единый формат. Затем, происходит главный шаг — поиск закономерностей и выделяющейся информации. После завершения данного события формируется отчёт, в котором отражены главные тенденции, исключения, основные особенности и выдвигаются автоматически сгенерированные гипотезы.

API социальных сетей

API позволяет извлекать данные из профилей социальных сетей в стандартизированном виде, например — JSON [9, с. 87]. «ВКонтакте» позволяет настраивать и использовать свой API для работы с учётными записями, рекламными кампаниями, мини приложениями, базой данных учебных заведений, документами, списком друзей, групп, поиском объекта в списке «Мне нравится», сообщениями, заметками, статусами и статистике по странице (пользователь и группа). Связь между API и разделами в социальной сети происходит через протокол HTTP [3, с. 124]. При правильной настройке и дальнейших манипуляциях, ответы API запросов в дальнейшем можно будет сохранять в форматах JSON и XML [4, с. 147]. API других социальных сетей работают почти этим же принципам. Больше всего из общедоступных API возможностей социальных сетей нас интересуют — индексация публикаций, поиск постов по временному фильтру и ограничение по количеству возвращаемых публикаций.

Сбор текстовых данных

В попытке извлечь полезную информацию, операторы интеллектуального анализа текста обычно применяются для команд системного анализа объемов данных в виде текста (на естественном языке), обнаружения шаблонов использования и определения лингвистической лексики. Процесс текстового интеллектуального анализа проходит в несколько этапов, где каждый из которых это полноценный, комплексный процесс. На вход подаются данные, которые имеют следующие характеристики: большой размер базы данных, высокая размерность слов, имеет зависимость (одна фраза зависит от другой), публикации неоднозначны, может содержаться шум (аббревиатуры, сокращения слов, термины и орфографические ошибки) [5, с. 94].

Извлечение информации из набора текстовых документов, таких как веб-страницы и API-ответы требует нескольких взаимосвязанных процессов — методы извлечения, фильтрации и распределения весов (коэффициентов). Результаты взвешивания данных обрабатываются с помощью методов интеллектуального анализа. Далее, применяется наивный байесовский классификатор. Это алгоритм, который выполняет обработку числовых данных с использованием байесовской вероятности. Такая классификация статистических данных способа относительно точно предсказать направления вектора и даже верные веса [6, с. 40].

Алгоритм работает по следующему принципу. На ввод подаётся частота целевых и противоположных фильтров. Рассчитывается вероятность возникновения обоих. Применяется условие, после которого происходит классификация обоих по заданному признаку. Для того, чтобы каждый фильтр получил вес, необходимо, чтобы он прошёл через алгоритмы отбора с обучающим набором. В нём будут признаки необходимого фильтра, его прямые и косвенные признаки.

Наивный байесовский метод проходит обучение постепенно и хорошо работает с репрезентативной статистикой. Но у него есть большой недостаток — это размер результирующего признака, который занимает много памяти. Здесь появляется задача на оптимизацию, которая решается через метод минимизации размера вектора. Наивный байесовский алгоритм в контексте задачи определения подверженности цели Интернет травли работает следующим образом.

Первым шагом задаётся n-мерный вектор, в котором есть совокупность значений «X», «X» в свою очередь равен совокупности значений от «x1» до «xn». Далее, задаётся вектор «m», в который записываются категории (фильтры) «A». То есть, от «A1» до «Am». «Am» формирует набор категорий «C». Когда появляется неизвестная категория, то классификатор определяет апостериорной вероятностью на основе условия элемента «x».

Соответственно, наивный байесовский классификатор сформирует отдельную группу признаков «C(x+1)». Следующие признаки, которые не похожи на все предыдущие будут иметь аналогичный инкрементирующий коэффициент. Таким образом, совокупность («C1» | «X»)> P («Cj» | «X») при условии, что длина меньше или равна коэффициентам «i» и «j», но «i» и «j» между собой не равны. Следующей задачей становится формирование зависимости фильтров, выполнение которой выглядит следующим образом:

$$P(C1|X) = \frac{\left(P(X|C1) * P(C1)\right)}{P(X)}$$

«Px» станет константой всех категорий, а «P(X|C1)» надо будет максимизировать. Априорные вероятности вычисляются через расчёт «P(Ci)», где P(Ci) — это результат деления «Si» на «S». «Si» — количество обучающих данных категории «Ci», а «S» — общий инпут. Следующая задача — сокращение времени «P(X|Ci)», которое будет потрачено на вычисления. «X» будет рассматриваться как переменное значение атрибута. Данная задача решается с помощью следующей формулы.

$$P(X|Ci) = n(к = 1)P(и^{к}|\ Ci)$$

Метод исследования

Данное исследование состоит из нескольких этапов. Сначала происходит сбор данных журнала из социальной сети, предварительная обработка (очистка) информации, которая была получена в результате структурированного сканирования [7, с. 49]. Затем, распределение данных с применением байесовской классификации и машинного обучения. Если подробно, то данная процедура состоит из шести этапов: сбор данных, предобработка, оценка важности весов в контексте документа, валидация информации, классификация и сбор результатов.

Манипуляции с данными

Собрано 1 367 записей, после предварительной обработки это количество сократилось до 472. Процесс выглядел следующим образом. Выполняется вход в учётную запись социальной сети, запрашивается токен для манипуляций с API и выполняется логический поиск. Фильтрация происходит благодаря применению ключевых слов, после чего результаты агрегируются в JSON-файл [8, с. 96].

JSON-файл конвертируется в CSV-таблицу, в нём будут проходить предварительная обработка и очистка от избыточных данных (первый этап — вручную, второй — с применением машинного обучения). Отдельно обрабатываются дубликаты идентификаторов, удаляются специальные символы ссылок и запросов, теги, хеш-значения, изображения и токены. Создаётся словарь, где указываются фильтры, к ним применяются весовые коэффициенты и исключения.

Предобработка — это комплексный процесс, во время которого происходит 9 последовательных автоматических манипуляций: импорт данных, удаление повторяющихся идентификаторов и специальных символов, токенизация, формирование словаря, стемминг, сохранение базы данных, преобразование в формат «ARFF», предпроцессинг через алгоритмы машинного обучения и классификация. «ARFF» формат нужен для того, чтобы осуществить «WEKA»-обработку.

Валидация данных происходит с помощью десятикратной перекрёстной проверки, которая возможна благодаря машинному обучению. Классификация происходит в 6 шагов — импорт результатов предобработки, проставление новых весовых коэффициентов, индексация, классификация через условие, определение флага и вывод. На инпут подаются данные, прошедшие предварительную предобработку. Следующий ручной ввод — обучающий набор информации. Происходит «TF-IDF»-взвешивание и 10-кратная перекрёстная проверка, классификация с применением наивного байесовского критерия. После чего запускаются алгоритмы сравнения агрегированного набора данных с автоматически модифицированным словарём.

Результат

Объём рассматриваемого датабанка — 472 публикации. Данные разделены на 2 части: первая половина — для обучения, вторая — для тестирования. Первая половина постов необходима для точечной минимизации ложноположительных срабатываний. Чем чаще они встречаются, тем больше времени разработчику ПО или оператору ЭВМ придётся проводить времени занимаясь ручным анализом событий. После первого запуска прикладного программного обеспечения и детального анализа флагов, 21.48 % событий оказалось ложнопозитивными. Далее, происходило редактирование фильтров и настройка флагов. После повторного запуска на обоих наборах данных, частота ложных реагирований снизилась до 4.56 %, что почти в 5 раз эффективнее первой итерации.

Заключение

Обнаружение актов Интернет травли как исследовательский ИТ-процесс состоит из трёх основных шагов и относительно легко осуществляется. На первом этапе необходимо создать учётную запись [10, с. 125], зайти в неё, зарегистрироваться в API-меню, получить токен, затем, изучить документацию API, сформировать JSON-файл и применить его к проекту. На втором этапе производится предварительный анализ, аналитика результатов и манипуляции машинного обучения, после чего получается структурированный аутпут. Третий шаг — «WEKA»-алгоритмы, «TF-IDF»-взвешивание, 10-кратная перекрёстная проверка и классификация. В результате выполнения данных этупов, оператор ЭВМ узнает, насколько сильно Интернет агрессия имеет место быть в социальной сети. Дальнейшие исследования по этой проблеме могут быть связаны с более детальным анализом Интернет травли в других социальных сетях, масштабированием онлайн сегмента и применением более специализированных алгоритмов интеллектуальной обработки данных, например, «C45», «K-Menas» и «SVM».

ЛИТЕРАТУРА

1. Черкасов Д.Ю., Иванов В.В. Машинное обучение. // Наука, техника и образование. 2018. № 1. С. 1–3.
2. Мухамаднева К.Б. Машинное обучение в совершенствовании образовательной среды. // Образование и проблемы развития общества. 2020. № 4. С 70–77.
3. Ахромов Я. Разработка интерфейсов к Интернет-приложениям на основе технологии API. // Новые информационные технологии в автоматизированных информационных системах. 2007. № 1. С. 124–126.
4. Широбокова С.Н., Стрельцова Е.А. Сравнительный анализ возможностей API социальных сетей по критерию функциональной полноты. // Инновационная наука. 2016. № 1. С. 147–151.
5. Арсеньтева Н.В., Скрипин А.А., Скрипина И.И. Data mining: методы, этапы, применение и значение в современном мире. // Форум молодых учёных. 2024. № 6. С. 94–101.
6. Баранчиков А.И., Федосова Е.Б. Применение методов data mining для анализа и выявления закономерностей в реляционных базах данных. // Радиотехнические и телекоммуникационные системы. 2023. № 1. С. 40–45.
7. Сикулер Д.В. Поиск данных для апробации интеллектуальных алгоритмов и технологий. // Символ науки. 2022. № 4. С. 49–55.
8. Исрар Т.Б. Искусственный интеллект в решении актуальных социальных и экономических проблем XXI века. // Наука и реальность. 2023. № 4. С. 96–102.
9. Кускарова О.И. Социальные сети как новая форма коммуникации: добро или зло. // Архонт. 2022. № 5. С. 87–92.
10. Селезнёв Р.С., Скрипак Е.И. Социальные сети как феномен информационного общества и специфика социальных связей в их среде. // СибСкрипт. 2013. № 2. С. 125–131.

Журнал «Современная наука: актуальные проблемы теории и практики»