Научная статья
Original article
УДК 004.55

ИССЛЕДОВАНИЕ МЕТОДОВ АНАЛИЗА АУДИТОРИИ СОЦИАЛЬНЫХ СЕТЕЙ

RESEARCH OF METHODS FOR ANALYZING THE AUDIENCE OF SOCIAL NETWORKS

Ефимов Егор Александрович, студент 2 курс, Институт Информационных технологий, МИРЭА - Российский Технологический Университет, Россия, г. Москва

Моругова Маргарита Владимировна, студент 2 курс, Институт Информационных технологий, МИРЭА - Российский Технологический Университет, Россия, г. Москва

Efimov Egor Aleksandrovich, student, 2nd year student, Institute of Information Technologies, MIREA - Russian Technological University, Russia, Moscow

Morugova Margarita Vladimirovna, student, 2nd year student, Institute of Information Technologies, MIREA - Russian Technological University, Russia, Moscow

Аннотация: Социальные сети как массовое явление начали зарождаться в 90-х годах предыдущего столетия, одновременно ознаменовав новую эпоху социализации персональных данных. Пользователи добровольно публикуют информацию о себе, а с учётом увеличения популярности социальных ресурсов, соразмерно увеличиваются и хранящиеся в них данные. На

основании этих данных крупные бренды могут вести прямой диалог с заинтересованными пользователями, поддерживая лояльность клиентов и распространяя позитивное мнение о бренде, что ведёт к прямому увеличению продаж. Упущение подобной информации равносильно полному бездействию в направлении развитии рекламы, а значит, социальные сети становятся уникальной средой для решения социальных, исследовательских и бизнес-задач. Для рассмотрения возможностей анализа подобных данных, следует выделить основные направления для аналитиков, а также методы, используемые в этих направлениях.

S u m m a r y : Social networks as a mass phenomenon began to emerge in the 90s of the previous century, simultaneously marking a new era of socialization of personal data. Users voluntarily publish information about themselves, and taking into account the increasing popularity of social resources, the data stored in them also increases proportionally. Based on this data, large brands can conduct a direct dialogue with interested users, maintaining customer loyalty and spreading a positive opinion about the brand, which leads to a direct increase in sales. The omission of such information is equivalent to complete inaction in the direction of advertising development, which means that social networks are becoming a unique environment for solving social, research and business problems. To consider the possibilities of analyzing such data, it is necessary to identify the main areas for analysts, as well as the methods used in these areas.

Ключевые слова: социальная сеть, анализ аудитории, web-сайт, анализ контента.

Keywords: social network, audience analysis, website, content analysis.

Введение

Прежде чем рассмотреть возможности анализа социальных сетей, следует глубже понять важность данных изучений. В традиционных средствах массовой информации пользователи являются простыми потребителями контента, в то время как социальные сети позволяют пользователям быть

потенциальным публикатором информации в ходе межличностного взаимодействия. Основная идея подобных сетей – быть в тесном контакте с аудиторией, что и является главным мотиватором распространения информации.

Также, стоит учесть, что любой анализ бессмысленный, без определённых выводов. Так, например, западные аналитики на основе социальных сетей производят анализ их использования и увеличение их популярности за последнее время [1].

Последним по очереди, но не по значимости, является сам способ сбора данных. Для сбора используются Web API, веб-ориентированных прикладных программных интерфейсов для встраивания инструментов аналитики. Основными проблемами в получении информации социальных сетей являются следующие аспекты:

1. Приватность данных – для анализа систем следует учитывать, что не все данные доступны для анализа. Пользователь может скрыть уровень детализации своего профиля;
2. Ограничения платформ – некоторые социальные сети в целях безопасности ограничивают количество передаваемых запросов, что затрудняет работу с данными за счёт следующего пункта;
3. Размерность данных – с учетом большого количества пользователей, может потребоваться параллельный сбор данных.
4. Слабая структурированность данных – затрудняет понимание полученного кластера для обработки, от чего в некоторых случаях автоматический сбор сводится на нет, чем обусловлена популярность профессий, связанных с инженерией и аналитикой данных [2].

Решение

Для анализа данных используются различные инструменты, которые используются в зависимости от типов данных, принятых на обработку. Среди них рассмотрим анализ графов, текстов, аудио объекты, а также графических, таких как видео объекты и изображения.

Наиболее простой в анализе метод является анализ графов. В социальных сетях содержится множественные данные пользователей с их персональными данными, включающие их даты и города рождения, предпочитаемые сообщества и медиаконтент. Эти сущности образуют собой множество связей, которые позволяют построить социальные графы, в которых узлами окажутся социальные объекты, а рёбрами – связи между ними.

При анализе характеристик объектов, стоит выделить их метрики, которые в целом, можно разделить на три объемные группы [3]. Первая группа – определение логики организации социальной сети, которая выделяет признаки организации объектов. Они отображают характер взаимоотношений одного социального объекта с другим, позволяющие определить степень их взаимодействия. К этой группе причисляют такие понятия, как множественность, взаимность и соседство.

Метриками второй группы являются особенности связи отдельных социальных объектов или графа в целом. При помощи которых находятся граничные связи (сильные, или же слабые), а также определяется значимость отдельного объекта и доля прямых связей в сети по отношению к общему числу возможных.

К метрикам третьей группы относят сегментационные метрики, отображающие характеристики отдельных сегментов социального графа. Они используются для определения особо плотных групп в пределах единой сети.

Исходя из анализа графов, можно получить информацию о пользователе по его предпочтениям, кругу общения и социальной принадлежности. Благодаря данной информации можно выделить рекомендации для пользователя и неявные связи, однако метод зависим от внутренней организации конкретной социальной сети.

Анализ текстов. Пользователи ежедневно оставляют сообщения в социальных сетях, отражающие их личную позицию касательно затрагиваемых вопросов. Анализ такого рода контента позволяет выявить настроения пользователей относительного того или иного вопроса. Наиболее

часто в анализе текстов используются методы определения тональности, основанные на словарях и машинном обучении.

При использовании метода словарей, составляются так называемые словари тональности, основанные на многочисленных лингвистических анализах, по которым и производится поиск лексики. После чего количество негативно и позитивно окрашенной лексики оценивается и вычисляется общая тональность текста. Недостатком такого метода является значительный объем исходных затрат для определения лексических правил определения выражения по интонации.

В основе машинного обучения [4] лежит классификатор, который обучается на заранее размеченных текстах, после чего полученные алгоритмы постоянно проверяются и дорабатываются при использовании на полученных от пользователей данных. Данный метод требует большого количества данных для обучения, однако, зачастую, машинная интерпретация, обученная необходимым образом, выделяет ключевые моменты лучше, чем это сделал бы человек, а также позволяет автоматизировать процесс обработки данных.

Анализ аудио объектов [5], к которым относится как непосредственно музыка, так и лекции, аудиокниги и записанные голосовые сообщения, и прочие информационные данные такого формата. Аудио составляют относительно небольшую часть социальных сетей, по большей части они вынесены в отдельные сегменты или даже сторонние приложения самой социальной сети. Это связано в первую очередь с тем, что информацию человек в основном считывает визуально, а не на слух. Тем не менее, можно провести анализ и аудио объектов. К некоторым таким объектам прилагаются текста, а также расшифровки, что могут послужить дополнительным источником информации.

Наиболее выделяющейся и обобщающей характеристикой для аудиоинформации является жанр, к примеру той же музыки. Нередко жанр прослушиваемых песен отражаются на слушателе в повседневной жизни и его выборах.

В наши дни набирает популярность метод анализа аудио посредством интеллектуального анализа нейронными сетями. Для начала искусственный интеллект обучают на основании кластера информации с заранее размеченными данными, после чего следует классификация самих аудио объектов пользователя. Благодаря выявлению частотных совпадений пользователю может быть предложена схожая музыка или схожая аудиокнига.

Кроме того, не стоит забывать и о распознавании речи из аудиосообщений, работающих на основе двух групп методов распознавания речи, а именно параметрических и лингвистических. Первая группа направленна на математическое преобразование сигнал с его дальнейшим преобразованием. Вторая же группа основывается на контексте высказывания и его обработке.

Анализ видео и изобразительных объектов представляет собой группировку алгоритмов обработки данных пользователей по их предпочтениям, применяющуюся к контенту на странице пользователя, количеству просмотра или прослушивания файлов, а также, особенно заинтересованные в качественных рекомендациях пользователи, часто используют внутренние инструменты (положительные или же отрицательные оценки контента) для помощи алгоритмам машинного обучения.

Заключение

С развитием социальных сетей в большие медиа-платформы, увеличилась и потребность в обработке данных, генерируемые пользователями в ходе использования этих самых сетей, что приводит к новым алгоритмам и методологиям анализа информации. При использовании текущих технологий наиболее доступным методом анализа остаётся анализ графов, поскольку он не требует большого количества вычислительных мощностей. Однако направление BigData [6] и технологии, связанные с этим направлением наиболее перспективны с точки зрения глубокого понимания поведения пользователей, что даёт стимул крупным компаниям инвестировать в это направление.

Литература

1. Kossinets, G., & Watts, J., Duncan. (2006). Empirical analysis of an evolving social network. Science, 311(5757), 88–90. (Дата обращения 23.04.2022).
2. The Popularity of Data Science Software [Электронный ресурс]. – URL: https://r4stats.com/articles/popularity/ (Дата обращения 21.04.2022).
3. Смирнова Ольга Сергеевна, Петров Алексей Иванович, Бабийчук Георгий Александрович Основные методы анализа, используемые при исследовании социальных сетей // Современные информационные технологии и ИТ-образование. 2016. №3-1. (дата обращения 19.04.2022).
4. Менщиков А.А., Комарова А.В., Гатчин Ю.А., Полев А.В. Разработка системы автоматического категорирования тематики страниц веб-ресурса // Программные системы и вычислительные методы. – 2016. – № 4. – С. 383 - 391. (дата обращения: 07.03.2022).
5. Nagesh Singh Chauhan. Audio Data Analysis Using Deep Learning with Python [Электронный ресурс]. – URL: https://levelup.gitconnected.com/audio-data-analysis-using-deep-learning-part-1-7f6e08803f60 (дата обращения 23.04.2022).
6. Dr. Ayesha Banu, Dr. Md. Yakub. Evolution of Big Data and tools for Big Data analytics // Journal of Interdisciplinary Cycle Research. 2020. №12(10). (дата обращения: 25.04.2022).

Literature

1. Kossinets, G., & Watts, J., Duncan. (2006). Empirical analysis of an evolving social network. Science, 311(5757), 88–90. (Accessed 23.04.2022)
2. Popularity of software for data processing [Electronic resource]. – URL: https://r4stats.com/articles/popularity / (Accessed 21.04.2022).
3. Smirnova Olga Sergeevna, Petrov Alexey Ivanovich, Babiychuk Georgy Alexandrovich The main methods of analysis used in the study of social networks // Modern information technologies and IT education. 2016. №3-1. (accessed 04/19/2022).

4. Menshchikov A.A., Komarova A.V., Gatchin Yu.A., Polev A.V. Development of a system for automatic categorization of the subject of web resource pages // Software systems and computational methods. – 2016. – No. 4. – p. 383 - 391. (accessed: 07.03.2022).
5. Nagesh Singh Chauhan. Analysis of audio data using Deep Learning using Python [Electronic resource]. – URL: https://levelup.gitconnected.com/audio-data-analysis-using-deep-learning-part-1-7f6e08803f60
6. Dr. Aisha Banu, MD Yakub. The evolution of Big data and tools for Big data analysis // Journal of Interdisciplinary Cycle Research. 2020. No.12(10). (accessed: 04/25/2022).

© Ефимов Е.А., Моругова М.В., 2022 Научно-образовательный журнал для студентов и преподавателей «StudNet» №5/2022.

Для цитирования: Ефимов Е.А., Моругова М.В., ИССЛЕДОВАНИЕ МЕТОДОВ АНАЛИЗА АУДИТОРИИ СОЦИАЛЬНЫХ СЕТЕЙ// Научно-образовательный журнал для студентов и преподавателей «StudNet» №5/2022.