

МЕТОДИКА ОБНАРУЖЕНИЯ ЭХОКАМЕР В СОЦИАЛЬНЫХ СЕТЯХ

Моликов И.А.

Научный руководитель – кандидат технических наук, Русак А.В.

Университет ИТМО

moliksey@yandex.ru

Введение

Эхокамеры - неявные сообщества пользователей социальных сетей, в которых продвигается единый нарратив. В таких группах люди взаимодействуют исключительно с единомышленниками, что усиливает поляризацию и искажает информационное поле. На данный момент эхокамеры рассматриваются в большинстве случаев как проблема рекомендательных алгоритмов, а не как реально существующий социальный кластер. А в тех случаях, когда эхокамеры определяются как реальные сущности, действующие методики основываются на анализе связей, явных хэштегов или тематических кластеров [4]. В данной работе предлагается мультимодальный подход, объединяющий анализ социальных связей и текстовой информации. В работе используются методы машинного обучения для определения тональности и тематики текста и для выделения собственно эхокамер. В результате работы создано приложение, использующее нижеперечисленные методы, которое анализирует информационную среду пользователя и предлагает ему альтернативные мнения и темы.

Основная часть

В основе предлагаемого подхода лежит гипотеза о том, что информационный пузырь может быть определен через пересечение трех измерений: социальных связей, тематики и эмоциональной окраски. Любое из этих измерений в отдельности не может полностью описать феномен эхокамер. Подписки и реакции могут указывать на дружескую группу или сообщество объединенное географически. Одна и та же тема может иметь различную эмоциональную окраску, а единый эмоциональный окрас без социальных связей может быть просто показателем консенсуса нескольких пузырей.

В работе ищется кластер социальных групп, затем отделяются группа кластеров, обсуждающих определенную тему, и последним этапом определяется отношение каждого кластера группы к этой теме.

Для выделения сообществ применены алгоритмы: Louvain (69 классов), Leiden (95 классов) и InfoMap (38 035 классов) [3]. Эти алгоритмы ориентированы на большие графы. Louvain и Leiden сформировали немного крупных кластеров, тогда как InfoMap выявил более разреженную структуру, что делает его лучшим вариантом для поставленной задачи.

Содержательная валидация осуществляется на основе метрики когерентности, оценивающей семантическую близость ключевых слов внутри темы. Сопоставление структурных кластеров с их тематическими профилями позволяет выявить группы, демонстрирующие аномально высокую тематическую сфокусированность, что является одним из маркеров эхокамеры.

В качестве основного инструмента тематического моделирования выбран BERTopic, совмещающий векторизацию предложений на основе SBERT, снижение размерности при помощи UMAP, кластеризацию (HDBSCAN) и извлечение ключевых слов методом c-TF-IDF. Данный подход позволяет не только автоматически

определять число тем в корпусе, но и получать интерпретируемые тематические профили для каждого выделенного сообщества.

Для определения тональности высказываний используется трансформерная модель RuBERT, адаптированная для русского языка и дообученная на корпусе тонально размеченных постов [1]. Модель выбрана из-за ее способности учитывать широкий контекст, распознавать иронию и сложные оценочные конструкции [2].

В качестве интерфейса взаимодействия с программой разработано веб-приложение.

Выводы

В результате работы разработан мультимодальный подход к обнаружению эхокамер в социальных сетях. Показано, что использование комбинации трех модальностей позволяет более точно идентифицировать замкнутые информационные пузыри по сравнению с традиционными подходами, основанными только на одном измерении. Предложенный пайплайн обработки данных может быть применен в также иных в задачах изучения социальных структур и информационных потоков.

Литература

1. Иконников С. А. Обзор методов анализа тональности текста // Студенческая наука: лучшие исследования и разработки 2024. - 2024. - С. 31.
2. Гаврилин П. С. Анализ тональности текстов на базе трансформеров // Современные проблемы проектирования, производства и эксплуатации радиотехнических систем: сборник научных трудов. - 2024. - С. 141.
3. Traag V. A., Waltman L., van Eck N. J. From Louvain to Leiden: guaranteeing well-connected communities // Scientific Reports. - 2019. - Vol. 9, № 1. - P. 5233.
4. Галаганова С. Г., Турусина Т. В. Технологии анализа социальных сетей с целью выявления социальных трендов // Человеческий капитал. - 2023. - № 1. - С. 169.