

АНАЛИЗ И ОБРАБОТКА БОЛЬШИХ ДАННЫХ НА ОСНОВЕ СОЦИАЛЬНЫХ СЕТЕЙ

Турсинкулова Индира

Университет менеджмента и технологий будущего Магистрант 2ого курса

tursinkulovai@gmail.com

Райимова Шахноза Озод кизи

Университет менеджмента и технологий будущего Магистрант 2ого курса

shaxnozarayimova96@gmail.com

<https://doi.org/10.5281/zenodo.15467530>

Аннотация. Данная статья представляет собой аналитический обзор современных методов анализа и обработки больших данных, генерируемых в социальных сетях. Исследование рассматривает основные подходы, технологии и инструменты, применяемые для извлечения полезной информации из массивов социальных данных. Особое внимание уделяется методологическим аспектам работы с большими данными в контексте социальных сетей, существующим проблемам и перспективным направлениям развития данной области как в глобальном контексте, так и в реалиях Узбекистана и России.

Ключевые слова: большие данные, социальные сети, анализ социальных сетей, обработка данных, интеллектуальный анализ данных, машинное обучение, сентимент-анализ, визуализация данных, прогнозная аналитика.

ANALYSIS AND PROCESSING OF BIG DATA BASED ON SOCIAL NETWORKS

Abstract. This article presents an analytical review of modern methods for analyzing and processing big data generated in social networks. The study examines the main approaches, technologies, and tools used to extract useful information from social data arrays. Special attention is paid to the methodological aspects of working with big data in the context of social networks, existing problems, and promising directions for the development of this field both in the global context and in the realities of Uzbekistan and Russia.

Keywords: big data, social networks, social network analysis, data processing, data mining, machine learning, sentiment analysis, data visualization, predictive analytics.

IJTIMOYIY TARMOQLAR ASOSIDA KATTA MA'LUMOTLARNI TAHLIL QILISH VA QAYTA ISHLASH

Annotatsiya. Ushbu maqola ijtimoiy tarmoqlarda yaratilgan katta ma'lumotlarni tahlil qilish va qayta ishlashning zamonaviy usullarining analitik sharhini taqdim etadi. Tadqiqot ijtimoiy ma'lumotlar massivlaridan foydali ma'lumotlarni ajratib olish uchun qo'llaniladigan asosiy yondashuvlar, texnologiyalar va vositalarni ko'rib chiqadi. Katta ma'lumotlar bilan ishlashning metodologik jihatlariga, ijtimoiy tarmoqlar kontekstida mavjud muammolarga va ushbu sohani rivojlantirishning istiqbolli yo'nalishlariga, ham global kontekstda, ham O'zbekiston va Rossiya realliklarida alohida e'tibor qaratilgan.

Kalit so'zlar: katta ma'lumotlar, ijtimoiy tarmoqlar, ijtimoiy tarmoqlar tahlili, ma'lumotlarni qayta ishlash, ma'lumotlarni intellektual tahlil qilish, mashinali o'rganish, sentiment-tahlil, ma'lumotlarni vizualizatsiya qilish, prognoz tahlili.

ВВЕДЕНИЕ

В современном цифровом мире социальные сети стали неотъемлемой частью повседневной жизни миллиардов людей. Ежедневно пользователи генерируют колоссальные объемы данных через публикации, комментарии, лайки, репосты и другие формы взаимодействия в социальных медиа. Социальные сети характеризуются скоростью, объемом, ценностью, разнообразием и достоверностью – пятью "V" больших данных (velocity, volume, value, variety, veracity) [1]. Эти массивы информации представляют собой богатый источник для анализа и извлечения ценных знаний о поведении пользователей, их предпочтениях, мнениях и социальных взаимодействиях. Актуальность темы обусловлена стремительным ростом влияния социальных сетей на различные аспекты общественной жизни – от потребительского поведения до политических процессов. По данным исследований, в Узбекистане к 2023 году насчитывалось 5,35 миллионов пользователей социальных сетей, что составляет 15,3% населения страны [2]. В России социальные сети также имеют широкое распространение, причем такие компании как телекоммуникационные операторы, финансовые учреждения и ритейлеры активно используют большие данные из социальных сетей для анализа клиентского поведения, управления лояльностью и персонализации услуг [3].

МЕТОДОЛОГИЯ И АНАЛИЗ ЛИТЕРАТУРЫ

Методология анализа больших данных социальных сетей базируется на интеграции нескольких научных дисциплин, включая компьютерные науки, статистику, социологию и лингвистику. Для систематического исследования аналитических подходов к большим данным социальных сетей применяется методика систематического литературного обзора (SLR), позволяющая идентифицировать, классифицировать и сравнивать существующие методы анализа [4]. Данный подход структурирует исследовательский процесс в три фазы: планирование, проведение исследования и документирование результатов. Основополагающей концепцией в данной области является анализ социальных сетей (Social Network Analysis, SNA). SNA представляет собой исследовательский метод, используемый для визуализации и анализа отношений и связей между объектами или индивидами в сети [5]. Этот метод позволяет изучать как формальные, так и неформальные связи, выявлять влиятельных акторов и определять паттерны коммуникации. В контексте больших данных SNA дополняется специализированными методами и технологиями, способными обрабатывать массивные объемы неструктурированной информации. Социальная сеть состоит из конечного набора акторов (социальных объектов), которые связаны между собой определенными отношениями, образуя структуру, которую можно анализировать с помощью теории графов [5].

Сбор данных из социальных сетей представляет собой первый этап аналитического процесса. Основными методами сбора данных являются: API-доступ, при котором многие социальные платформы предоставляют программные интерфейсы, позволяющие исследователям и разработчикам получать доступ к данным (доступность API от Twitter, Facebook и новостных сервисов привела к «взрыву» сервисов данных и программных инструментов для сбора и анализа данных социальных медиа [6]); веб-скрейпинг, представляющий метод извлечения данных с веб-страниц социальных сетей с помощью

автоматизированных инструментов (этот подход используется, когда API-доступ ограничен или недоступен); потоковая передача данных, когда некоторые платформы позволяют получать данные в режиме реального времени через потоковые API.

Современные подходы к анализу больших данных социальных сетей включают: анализ контента и сентимент-анализ, где обработка естественного языка (NLP), сентимент-анализ и анализ социальных сетей являются ключевыми техниками аналитики больших данных, используемыми в социальных медиа, что в конечном итоге способствует принятию организационных решений [7]; анализ структуры сети, включающий изучение топологических свойств социальных сетей, включая плотность связей, централизацию, кластеризацию и идентификацию сообществ (существует два основных типа анализа социальных сетей: эго-сетевой анализ, фокусирующийся на отношениях конкретного индивида, и полный сетевой анализ, исследующий отношения между всеми узлами в сети [5]); предиктивную аналитику, использующую исторические данные социальных сетей для прогнозирования будущего поведения пользователей, распространения информации или трендов; визуализацию сетей и обнаружение аномалий, выявляющее необычные паттерны или поведение в социальных сетях, что может указывать на мошенничество, спам или другие нежелательные явления.

В Узбекистане и России развиваются собственные подходы к исследованию социальных сетей, учитывающие региональную специфику. В Узбекистане социальные сети и мессенджеры стали важным источником новостей для населения. Facebook занимает первое место среди социальных сетей, а Telegram – среди мессенджеров [8]. Это создает уникальный контекст для анализа распространения информации и общественного мнения. В России социальные сети также играют значительную роль.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Анализ литературы показывает активное развитие исследований в области больших данных социальных сетей. Исследования в этой области экспоненциально росли с 2014 года, причем системы поддержки принятия решений и эвристические методы являются ключевыми направлениями [9]. Основные тенденции включают интеграцию методов машинного обучения, глубокого обучения и аналитики графов для извлечения знаний из социальных данных. Важной особенностью современных исследований является междисциплинарный подход. Социальные сети характеризуются пятью "V" больших данных, что требует применения методов аналитики больших данных для понимания коммуникационных паттернов между пользователями и их поведения [10]. Исследователи совместно работают над развитием методов компьютерных наук, статистики и социальных наук для решения сложных проблем анализа.

Несмотря на значительный прогресс, область сталкивается с рядом вызовов: масштабируемость, требующая обработки постоянно растущих объемов данных и разработки все более эффективных алгоритмов и инфраструктурных решений; качество данных, где исследователи отмечают проблемы с ненадежными и неполными данными в социальных сетях, включая сленг, аббревиатуры и орфографические ошибки, что затрудняет точную классификацию [4]; конфиденциальность, поскольку в эпоху онлайн-больших данных и социальных медиа защита конфиденциальности пользователей в

социальных сетях рассматривается как важная проблема [1]; интерпретация результатов, где извлечение значимых выводов из сложных аналитических моделей остается сложной задачей, требующей экспертных знаний; многоязычность, так как в Узбекистане, как и в других странах Центральной Азии, каждый второй житель читает новости на русском языке, при этом молодые пользователи в возрасте от 19 до 43 лет чаще потребляют новости на узбекском языке [8], что создает дополнительные сложности для анализа.

ЗАКЛЮЧЕНИЕ

Анализ и обработка больших данных на основе социальных сетей представляет собой динамично развивающуюся область исследований, объединяющую методы компьютерных наук, статистики и социальных наук. Современные подходы к анализу больших данных социальных сетей включают разнообразные методы – от анализа контента и сентимент-анализа до исследования структуры сетей и предиктивной аналитики. В ходе исследования было выявлено, что в Узбекистане и России формируются собственные подходы к анализу социальных сетей, учитывающие специфику местных платформ и языковые особенности. При этом сохраняются общие методологические принципы, характерные для глобального научного сообщества. Основными вызовами в области остаются вопросы масштабируемости, качества данных, конфиденциальности и интерпретации результатов. Перспективными направлениями развития являются интеграция с другими источниками данных, совершенствование алгоритмов обработки естественного языка, развитие распределенных технологий анализа и персонализированной аналитики.

REFERENCES

1. Bello-Orgaz, G., Jung, J. J., Camacho, D. (2016). Social big data: Recent achievements and new challenges. *Information Fusion*. 28, 45-59.
2. Stieglitz, S., Mirbabaie, M., Ross, B., Neuberger, C. (2018). Social media analytics – Challenges in topic discovery, data collection, and data preparation. *International Journal of Information Management*. 39, 156-168.
3. Bykanova, O., Bykanov, D., Khusainova, A., Mahaieva, E. (2023). Five Ways How Big Data is Applied in Russia. *Datafloq*. April 12.
4. Amini, A., Karimi, N., Jandaghi, G. (2022). Big data analytics meets social media: A systematic review of techniques, open issues, and future directions. *Telematics and Informatics*. 62, 101734.
5. Tripathy, B. K., Mitra, A. (2019). Big Data Techniques in Social Network Analysis. *Journal of Information Technology Applications and Management*. 25(4), 113-132.
6. Ahmed, Y. A., Ahmad, M. N., Ahmad, N., Zakaria, N. H. (2019). Social media analytics: A survey of techniques, tools and platforms. *AI & Society*. 34, 617-636.
7. Agrawal, D., Budak, C., El Abbadi, A., Georgiou, T., Yan, X. (2014). Big Data in Online Social Networks: User Interaction Analysis to Model User Behavior in Social Networks. *Lecture Notes in Computer Science*. 8381, 1-16.
8. Ghaffar, M. A., Hassan, M., Mand, A. (2022). The role of social networks in the information space of Uzbekistan. *EU Reporter*. October 17.

9. Esfahani, H. J., Tavasoli, K., Jabbarzadeh, A. (2019). Big data and social media: A scientometrics analysis. *International Journal of Data and Network Science*. 3(3), 145-164.
10. Meng, X., Zhu, Y., Zhang, L. (2022). Social-Network-Sourced Big Data Analytics. *IEEE Transactions on Big Data*. 8(2), 321-334.