information

Editorial

Social Media Mining and Analysis: A Brief Review of Recent Challenges

Nirmalya Thakur

Department of Computer Science, Emory University, Atlanta, GA 30322, USA; nirmalya.thakur@emory.edu

Social media platforms are a type of web-based applications that are built on the conceptual and technical underpinnings of Web 2.0 and enable the generation and dissemination of user-generated content. Social media platforms provide a seamless means for users to connect, communicate, and collaborate with each other. Social media is digital as it exists fully on the Internet or via outlets connected to the Internet. Social media is a collection of several social media platforms, such as Twitter, Facebook, Instagram, TikTok, YouTube, Sina Weibo, and SnapChat, just to name a few. While there have been multiple social media platforms developed in the last few years, all such platforms use the same underlying architecture based on Web 2.0 technologies. This implies they are built on an Internet framework that lets countless users collaborate on content production and dissemination. Users do not only consume digital content from the Web; they also contribute to its creation. Social media significantly differs from traditional Web 1.0 communication methods and online interaction mediums such as email and text messaging. Web 1.0 systems are more linear in terms of content distribution and cannot readily sustain massive and heavily engaged networks. Web 1.0 technologies, in contrast to Web 2.0, do not allow numerous users to retrieve and change uploaded material. To summarize, social media differs from previous online networks and online interaction mediums in that they are far more open, participatory, flexible, robust, and creative [1].

One well-known aspect of social media is that, unlike randomized networks, where two nodes in the network have a predetermined chance of being linked to each other, the framework of social media is scale-free. A scale-free network is one in which the degree distribution of a network follows a power law. A network's degree distribution is termed a power law when the probability of a node having degree k is proportional to $k^{-\gamma}$ where the value of γ is empirically determined. Such networks are distinguished by exceptionally high-degree hub nodes, whereas the vast majority of nodes have significantly lower degrees. Few individuals impact large groups on a social media platform because they have a considerably greater number of connections or friends, whereas most average users possess significantly fewer connections or friends in comparison. The small-world feature is another essential property shared by social media networks. Even in a very vast and sparse scale-free network like the Web, two nodes will probably be connected by a very short route. This may not always seem obvious, given that most links in social networks are densely packed. However, the existence of hubs in scale-free networks can also confer the small-world property, provided a connection in the majority of the network remains clustered. Social media platforms like real-life social networks comprise virtual connections between individuals, so the small-world phenomena always apply to such platforms [2].

Every day, massive volumes of user-generated material are uploaded on social media platforms. This pattern is expected to grow in the years to come. As a result, it is vital for content creators, users, and services on social media platforms to understand how to generate, mine, manage, analyze, and utilize large amounts of data. The popularity of social media platforms has been growing exponentially in the last few years. At present, 4.9 billion people use social media, and this number is expected to rise to 5.85 billion by

Citation: Thakur, N. Social Media Mining and Analysis: A Brief Review of Recent Challenges. *Information* **2023**, *14*, 484. https://doi.org/10.3390/info14090484

Received: 15 August 2023
Accepted: 27 August 2023
Published: 31 August 2023

2027. On average, a social media user has 8.4 social media accounts and spends about 145 min across different social media platforms daily [3].

The exponential growth of social media platforms in the last few years has been driven by several research challenges. Gundecha et al. [1] and Jones et al. [4] presented various challenges related to the mining and analysis of content production, alteration, and dissemination on different topics on social media. These challenges in this area of research are summarized as follows:

- Information Diffusion: Understanding and analyzing the multimodal patterns of information exchange and dissemination on social media, including investigating misinformation, memes, rumors, conspiracy theories, and viral posts. Researchers in this field have studied different information diffusion models, such as the independent cascade model, the threshold model, the susceptible–infected model, and the susceptible–infected–recovered model.
- Privacy: Investigation of privacy concerns of social media users. The conflicting desires of social media users have led to new challenges in this area of research. Specifically, users desire to have as many friends or connections as possible, and at the same time, they wish to be as private as possible. Social media users have been subject to different privacy concerns, such as stalking, cyberbullying, phishing, scamming, and clickjacking.
- Trust: Studying the various degrees to which the general public trusts the information available via social media. Prior works in this field have mostly interpreted trust in terms of integrity, ability, and benevolence.
- Sentiment Analysis and Opinion Mining: Investigation of the sentiment, emotion, or opinion associated with different topics of conversation on social media. For instance, companies or organizations may be able to comprehend the opinions of consumers regarding goods, brand perception, new product awareness, degrees of perceived acceptance of new products, and reputation by using concepts of sentiment analysis and opinion mining.
- User Migration: Understanding when and why users move from one social media network to another. It is crucial for social media platforms to retain their current users while also attracting new ones. Studying how users select social media platforms has significant ramifications. Understanding migratory patterns may assist a social media platform in three ways: (1) generating revenue through applicable recommendations, (2) increasing traffic via shared content, and (3) expanding its network of committed users.
- Location-Based Social Networks (LSBNs): An LSBN is an architecture that comprises individuals connected via the interdependencies obtained from their locations in the physical world and their location-tagged content on social media platforms, such as photos, videos, and texts. The instantaneous position of a person at a specific timestamp and the location history make up the physical location in this case. Interdependency comprises information, such as common interests, behavior, and operations, deduced from a person's geographical location and location-tagged data.
- Social Recommendations: The idea behind social recommendations is that individuals who are linked or connected on social media platforms are more likely to have shared or comparable likes and dislikes, and users are readily swayed by their friends or connections on such platforms over random recommendations. Social recommendations aim to reduce the issue of information overload.
- Community Analysis: A community may also be described as a collection, network, harmonious subgroup, or component, depending on the situation and the social media platform being studied. Social media allows people to develop and grow their virtual connections, so community analysis in this context helps to infer the multimodal characteristics of these virtual connections.
- Influence Modeling: It is crucial to understand whether a social media platform is homophily-driven or influence-driven. For instance, in the advertising sector, if a social media platform is influence-driven, influencers on that platform should be identified, and they may be given incentives to spread the word about a certain product or

service among other users to increase sales for that product or service. However, if a social media platform is homophily-driven, then specific people should be sought to increase sales for that product or service. Most social networks combine homophily and influence. Therefore, differentiating them and following specific and applicable measures (for instance, for advertising a product or service) is a challenge.

- Topic Modeling: Techniques for topic modeling are frequently used in natural language processing to extract topics and linguistic information and underlining patterns from unordered data. Performing topic modeling on data obtained from social media helps to identify the different topics of conversations in virtual communities and trends of the same over a period of time.

The ubiquitousness of social media platforms makes them a rich resource for the investigation of a wide range of research questions related to one or more of the above-mentioned focus areas, as can be seen from the recent works in this field, which focused on the mining and analysis of conversations on social media related to ChatGPT, the Russia–Ukraine war, cryptocurrency markets, virtual assistants, abortions, loneliness, housing needs, fake news, religion, cognitive impairment, gender identity, elections, fall detection, drug safety, pregnancy, food insufficiency, indoor localization, education, MPox outbreak, and COVID-19 outbreak [5–7]. These works related to the mining and analysis of data from social media platforms have been complemented by research in the areas of software development, algorithm design, algorithm development, and programming languages in the last few years. As a result of this research, several libraries, tools, applications, frameworks, and software platforms are available today that make the mining and analysis of data from social media platforms seamless and faster than ever before [1].

To conclude, mining and analysis of data from social media holds the potential to contribute towards the timely advancement of research and development in a wide range of disciplines. The objective of this Editorial is to announce the Special Issue titled "Recent Advances in Social Media Mining and Analysis" for the Journal of Information (ISSN 2078-2489). This Special Issue aims to present and disseminate successful research projects related to the mining and analysis of data from social media with a specific focus on one or more of the above-mentioned challenges and related areas of research in this field, with a solid methodological structure, and where the novel findings differentiate the developed and adopted solutions.

Conflicts of Interest: The author declares no conflict of interest.

References

1. Gundecha, P.; Liu, H. Mining Social Media: A Brief Introduction. In *2012 TutORials in Operations Research*; INFORMS: Catonsville, MD, USA, 2012; pp. 1–17.
2. Bhattacharya, A. A Multi-Agent Model to Study the Effects of Crowdsourcing on the Spread of Misinformation in Social Networks. Masters Thesis, University of Cincinnati, Cincinnati, OH, USA, 2023. Available online: http://rave.ohiolink.edu/etdc/view?acc_num=ucin1684770124758418 (accessed on 14 August 2023).
3. Belle Wong, J.D. Top Social Media Statistics and Trends of 2023. Available online: https://www.forbes.com/advisor/business/social-media-statistics/ (accessed on 14 August 2023).
4. Jones, I.; Liu, H. Mining Social Media: Challenges and Opportunities. In Proceedings of the 2013 International Conference on Social Intelligence and Technology, State College, PA, USA, 8–10 May 2013; pp. 90–99.
5. Thakur, N. Sentiment Analysis and Text Analysis of the Public Discourse on Twitter about COVID-19 and MPox. *Big Data Cogn. Comput.* **2023**, *7*, 116. [CrossRef]
6. Thakur, N. A Large-Scale Dataset of Twitter Chatter about Online Learning during the Current COVID-19 Omicron Wave. *Data* **2022**, *7*, 109. [CrossRef]
7. Thakur, N.; Han, C.Y. A Multimodal Approach for Early Detection of Cognitive Impairment from Tweets. In *Human Interaction, Emerging Technologies and Future Systems V*; Springer International Publishing: Cham, Switzerland, 2022; pp. 11–19. ISBN 9783030855390.