ISSN: 2096-3246
Volume 51, Issue 01, Jan, 2019

https://doi.org/10.5281/zenodo.12707033

Social Media Big Data: Scope & Challenges

VIJAYA BHASKAR MADGULA, K BALAJI SUNIL CHANDRA, JANGILI RAVI KISHORE

Assistant Professor [1,2,3]

viswanath.p002@gmail.com, vijaya.bhaskar2010@gmail.com, KISHORE_EDIGA@GMAIL.COM

Department of Computer Science and Engineering, Sri Venkateswara Institute of Technology, N.H 44, Hampapuram, Rapthadu, Anantapuramu, Andhra Pradesh 515722

Keywords:
Twitter, Tweets, election, prediction, LDA, Social Influence

ABSTRACT

In the age of big data, social media has emerged as a prime example of a representative and relevant data source. The rapid expansion of the Internet has given rise to a new breed of individuals—or maybe a whole new social order—that is enthralled by the prospect of connecting with others via shared experiences and information. Facebook, Instagram, Google, Tumblr, Flickr, Twitter, and LinkedIn are just a few of the many prominent online programmes and websites that contribute to social media data. This data has expanded to several domains, including e-commerce, e-business, e-tourism, hobbies, friendship, education, health, and day-to-day employment. New and intriguing problems have emerged in fields such as social media and networks as a result of the merging of big data technology with conventional machine learning algorithms. Issues with data processing, storage, and representation, as well as its usage for pattern mining, user behaviour analysis, data visualisation, and tracking, are at the heart of these new issues.

ISSN: 2096-3246
Volume 51, Issue 01, Jan, 2019

https://doi.org/10.5281/zenodo.1270703

Introduction

There is tremendous value in Big Data analytics, which is why the term "Big Data" is now trending. Traditional technologies are unable to store, process, or analyse the massive amounts of data sets created every minute by consumers and companies across the globe. Millions of data sources now provide data at a breakneck pace. All throughout the globe, you may find these data sources. Social media sites and networks are among the most extensive data repositories. As an example, consider Facebook: daily data generation exceeds 500 terabytes. Media files, text messages, and more are all part of this data set. The merging of social media and big data is what gives rise to social big data. With a heavy emphasis on social media, social big data will therefore rely on the examination of massive volumes of data that may originate from several dispersed sources. Data mining, machine learning, statistics, graph mining, information retrieval, linguistics, natural language processing, the semantic web, ontologies, and big data computing are all included in social big data analytics, which means it is inherently multidisciplinary. Many other fields may benefit from them as well, including health and political trends and forecasts, hobbies, e-business, cybercrime, counterterrorism, social network analysis, time-evolving opinion mining, and human machine interactions.
"Those processes and methods that are designed to provide sensitive and relevant knowledge to any user or company from social media data sources when data sources can be characterised by their different formats and contents, their very large size, and the online or streamed generation of information." This provides a working definition of social big data.

ISSN: 2096-3246
Volume 51, Issue 01, Jan, 2019

https://doi.org/10.5281/zenodo.12707033

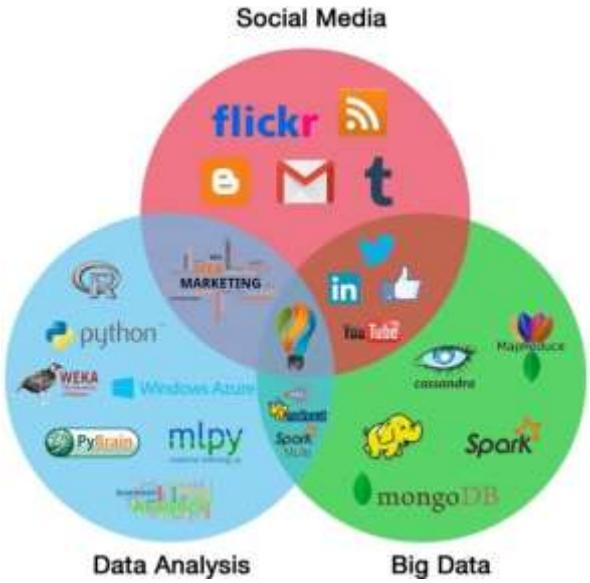

Fig.1. The conceptual map of Social Big Data.[1]

You can see a simplified diagram of the three main categories of social big data in the image.

One, using social media as an inherent data source.

3. Data analysis, which is a collection of techniques and procedures used to extract and analyse information, and 2. big data, which is a paradigm for parallel and large processing. The idea of blending such locations is reflected in the junctions between these groups. Some machine learning frameworks, for instance, have been built on top of big data technology; this is evident at the confluence of big data and data analysis. Apps in the marketing and e-health industries, for example, rely heavily on social media data, and the idea behind these apps is founded on the idea of data analysis meeting social media. Some social networking apps are using big data technologies to build their web platforms; LinkedIn, Facebook, and YouTube are just a few examples. These technologies include MongoDB, Cassandra, Hadoop, and more. Finally, the knowledge extraction and exploitation aim, shown in the middle of the previous diagram, is the only objective of any social big data application. [1]

Here are some of the opportunities and challenges associated with social networking big data:

Big data social impact modelling and analysis are the cornerstones of this field. exploring the features and processes of social networks via modelling; the challenge of optimising influence with large data; study of social impact in large-scale, dynamic networks; study of social influence in networks with heterogeneity; study of casual relationships in such networks.

• Technological advancements: social media advertising and recommendation systems powered by big data; modelling, techniques, and tools for influence propagation; methods for community detection using big data; methods for analysing user behaviour in conjunction with social influence evaluation; methods for differentiating between positive, negative, and controversy influence; studying the effects of social networks on human behaviour in online communities; analysing human behaviour in online communities using big data; and modelling the impacts of online communities on online communities.

• Security: developing novel secure solutions for designing, supporting, and operating social networks; analysing threats and vulnerabilities in these platforms; developing secure social networking applications with social influence analysis; preventing malware from propagating in these platforms; modelling the secure mechanisms of these networks;

ISSN: 2096-3246
Volume 51, Issue 01, Jan, 2019

https://doi.org/10.5281/zenodo.1270703

and so on.

data, safe social network apps with big data, social network security with big data, models, techniques, and tools for social network security testing, social network spam with big data, and social network harmful information propagation detection.
• Trust: concepts, methodologies, and tools for assessing social network trust; trust management in social networks using big data; and trust assessment in social networks using big data.
• Confidentiality: Confidentiality in processing and analysing social media big data; Confidentiality in social media networks using big data; Confidentiality in social media models, procedures, and resources. [2] in

1. Research challenges

Researchers in the fields of social science, computer science, and academic financing face a plethora of obstacles when it comes to social media scraping and analytics. These obstacles include:
• Scraping—though data from social media platforms is available through application programming interfaces (APIs), big data providers like Facebook and Google are making it harder for academics to get their hands on all of their "raw" data. Additionally, very few social data sources offer researchers and academic institutions affordable data. Access to the data of news services like Bloomberg and Thomson Reuters usually comes at a premium cost. Alternatively, researchers may apply to join Twitter's Data Grants programme and have access to public tweets and historical data. With more than 500 million tweets every day, this large dataset might provide valuable insights.
• There are still many issues and research concerns with data cleansing, which involves cleaning unstructured textual data (such as normalising text). This is especially true for high-frequency streaming real-time data.
Researchers are progressively combining and integrating innovative data sources, including as social media data, real-time market and consumer data, and geographical data, to conduct holistic analyses.
To prevent users from trying to "suck" all the valuable data from the database, it is important to secure the data, resolve ownership and intellectual property issues (for example, most publishers' terms of service forbid storing scraped data), and provide users varying degrees of access.
Foreign languages, foreign terms, slang, spelling mistakes, and the natural evolution of language continue to provide a multitude of obstacles to advanced social media data analysis for opinion mining (e.g., sentiment analysis) in the realm of data analytics.
• Analytics dashboards—to access feeds or programme analytics models in a programming language like Java, many social media sites demand that users develop APIs. Although these abilities are common among computer scientists, they are usually beyond the scope of most researchers in the social sciences. In order to provide what may be described as "deep" access to "raw" data, non-programming interfaces are necessary. This is true for many tasks, including setting up APIs, integrating comprehensive sources, creating analytical models, and merging social media feeds.
The purpose of data visualisation is to convey information clearly and effectively using graphics by representing data in a way that is both understandable and easy to understand. Visualisation is gaining importance due to the large amount of data involved. the third

2. Related work

ISSN: 2096-3246
Volume 51, Issue 01, Jan, 2019

https://doi.org/10.5281/zenodo.12707033
A large amount of data, both organised and unstructured, is created every day by social media. The massive amounts of unstructured data are too much for traditional methods like Relational Database Management Systems (RDBMS). Data with a modest quantity of structured information may be processed using RDBMS. The current systems are slow and cumbersome because of how slowly they handle data. Also, there is a limit on how big data may be. All of the computing power is concentrated on a small number of devices. By design, the machines can't handle errors.

Using a social network's centrality to forecast spam detection—for example, the likelihood that a person will publish spam on a social network—is described in the article "Using Social Network Analysis for Spam Detection" by DeBarr and Wechsler [5]. Using graph-based metrics to enhance spam categorization on a microblogging site is another approach that Wang [6] mentioned in another study concerning Twitter.
A basic unsupervised algorithm may be used for spam identification, according to research by Mehta et al. [7]. The statistical features of efficient spam profiles are used by this method. According to the report, these features contribute to the development of a spam detection system that is both very accurate and very fast. Thanks to technological advancements, social networks like Tickle, Facebook, MySpace, LinkedIn, Friendster[8], and many more have millions of users who use them for both personal and professional networking.
People spend a lot of time working on Facebook, MySpace, or Twitter, and it's not easy to figure out which method to use for each. Some examples of these methods include collaborative filtering, friend graph analysis, classification, behavioural analysis, and so on.

3. Social Big Data Applications:-

Users and businesses alike may benefit from applying social big data analysis to data sources found in social media in order to uncover actionable insights that can inform better decision-making. To better understand the company's operations and the market, as well as to back up decision-making, business intelligence entails analysing crucial business data using a variety of methods, systems, processes, and applications. approaches developed in this domain have found use in a variety of contexts, including healthcare, e-commerce, marketing, and security. More recently, these approaches have been used to address social big data.[1]
i. Marketing: Experts in the field of marketing think that cloud computing and big data analytics in social media provide a once in a lifetime chance for companies to get feedback from a large audience, which can then be used to refine more conventional marketing approaches.
ii. Analysis of criminal activity: This kind of behaviour is very situational and characterised by recurrent patterns. In other words, places that have characteristics that make criminal activity easier will have a higher concentration of criminals. The goal of analysing criminal records is to spot trends in criminal activity so that more crimes and their connections may be found. Law enforcement organisations may greatly benefit from the information that can be gathered via data mining methods.
iii. Epidemic intelligence: This refers to the ability to detect, evaluate, and confirm any threats to public health in a timely manner and then provide the necessary warnings. The field encompasses methods for the ongoing and automatic monitoring of online sources such as social networks, blogs, digital news media, and government documents that include unstructured free text or media content.
iv. Visualisation based on user experiences: Better user experiences and services may be

ISSN: 2096-3246
Volume 51, Issue 01, Jan, 2019

https://doi.org/10.5281/zenodo.1270703

achieved by visualising big data from social media. As an example, there are several ways to format the massive amounts of numerical data, which are often presented in tabular form. As a result, it becomes easier for users to grasp. Success in many fields depends on the capacity to visualise huge data in a timely manner; this is true in many areas of medicine, cyber and national security, disaster management, and business.[1]

4. Conclusion

5. The volume of data produced by social media platforms has been enormous in the last several years. Big data analytics in social media: a review of the field, a look at its potential uses,

difficulties and tasks associated with large data. Accelerated progress in several scientific fields, as well as increased profitability and success, may be possible with improved analysis of the massive amounts of data that are becoming accessible. With the right use of big data analytics, almost any department may reap the advantages, whether it's marketing, criminal research, epidemic intelligence, user experiences-based visualisation, etc.

References

1. G. Bello-Orgaz et al. “Social big data: Recent achievements and new challenges”, Information Fusion 28, 2016, pg 45–59.
2. Sancheng Peng, Guojun Wang, Dongqing Xie “Social Influence Analysis in Social Networking Big Data: Opportunities and Challenges”, IEEE, October 2016, PP(99):12-18
3. Bogdan Batrinca, Philip C. Treleaven, “Social media analytics: a survey of techniques, tools and platforms”, AI & SOCIETY volume 30, Springer, 2015, pages89–116.
4. Jitendra Pandey, Minimol Anil Job “Proposed Framework for Spam Recognition in Big Data for Social Media Networks in Smart Environment”, IEEE, 2019
5. DeBarr, D., Wechsler, H.: Using Social Network Analysis for Spam Detection. In: Chai, S.-K., Salerno, J.J., Mabry, P.L. (eds.) SBP 2010. LNCS, vol. 6007, Springer, Heidelberg (2010), pp. 62–69.
6. Wang, A.H.: Don’t follow me: Spam detection in twitter. In: Proceedings of the 2010 International Conference on Security and Cryptography (SECRYPT), IEEE, 2010, pp. 1–10.
7. Mehta, B., Hofmann, T., Fankhauser, P.: Lies and propaganda: detecting spam users in collaborative filtering. In: Proceedings of the 12th International Conference on Intelligent User Interfaces, ACM (2007) pp. 14–21.
8. Danah, M.B., Friendster and publicly articulated social networking. Proceedings of Extended Abstracts on Human Factors in Computing Systems, (CHI ‘04), Vienna, Austria, 2004, pp: 1279-1282. DOI:10.1145/985921.986043
9. S. Yardi, D. Romero, G. Schoenebeck, Detecting spam in a Twitter network, First Monday 15 (1) (2009).
10. De Wang, Danesh Irani, and Calton Pu. A social-spam detection framework. In Proceedings of the 8th Annual Collaboration, Electronic Messaging, Anti-Abuse and Spam Conference, pages 46–54. ACM, 2011.
11.