Chapter 9
Social Media Content Analysis:
Machine Learning

D. Sudaroli Vijayakumar
https://orcid.org/0000-0001-8270-6223
PES University, Bangalore, India

Senbagavalli M.
https://orcid.org/0000-0003-3806-8257
Alliance University, Bangalore, India

Jesudas Thangaraju
Mahendra Engineering College (Autonomous), Namakkal, India

Sathiyamoorthi V.
https://orcid.org/0000-0002-7012-3941
Sona College of Technology, Salem, India

ABSTRACT

Today's wealth and value are data. Data, used sensibly, are making wonders to make wise decisions for individuals, corporates, etc. The era of spending time with an individual to understand them better is gone. Individual's interests, requirements are identified easily by observing the activities an individual performs in social media. Social media, started as a tool for interaction, has grown as a platform to make and promote business. Social media content is unavoidable as the data that are going to be dealt with is huge in volume, variety, and velocity. The demand for using machine learning in analysing social media content is increasing at a faster pace in identifying influencers, demands of individuals. However, the real complexity lies in making the data from social media suitable for analysis. The type of data from social media content may be audio, video, image. The chapter attempts to give a comprehensive overview of the various pre-processing methods involved in dealing the social media content and the usage of right algorithms at the right time with suitable case examples.

INTRODUCTION

The influence of social media in one's life has achieved massive growth. It is very hard to find an individual who does not use Facebook and twitter daily thus contributed a lot in terms of connectivity. It is also been estimated that more than 80% of online users use this social network. Thus, social networks like any network consist of users who are connected to each other through a common web-based application. Nodes here in these social networks are the users and the connectivity is formed through the various commonalties or matches among users through some pattern. This connectivity has improved the way of collecting social data. This drift in technology has eventually replaced the conventional approaches to computational intelligence. The power of social media is so strong, so organizations use this as a tool to identify accurately their competitors, business trends etc (Beier and Wagner, 2016; Senbagavalli and Tholkappia, 2016). This conversion of adopting computational intelligence on social network assist one to do better decision making not only in business but also to spread news faster, take precautions against natural calamities, etc. This social media content analysis gained a lot of popularity due to the closeness results towards the real happenings. This proximity between real and predicted results is made possible because of the huge amount of data that is been used to get insights. The number of users using social networks is increasing at a constant rate so the percentage of accuracy on social media analysis will be incremented.

Despite the numerous benefits, doing a social network analysis involves many complexities as the type of data that has to be handled comes in various forms. Making the data suitable for the analysis itself is a huge task in social networks. It is because most of the data that we obtain through social networks is unstructured. The computational approaches in social networks still in its inception because of the complexities involved in understanding the human language as it varies from person to person.

Another challenge is in the amount of data that must be structured and make it understandable (Chaudhary et al., 2016; TAM, 2016). Assume an organization is having 100 employees. Every employee is having their own set of friends forming around 6 friends' group. In group 1 an employee discusses her resignation to her set of friends on Monday. Tuesday a person from group 2 asks her about her resignation. It was shocking news as the person from group 2 informed her that the whole organization knows about her resignation. This was made possible because of the interrelationship among the six groups. So, to identify who are friends with whom, an analysis process is involved that starts off with the construction of a graph with 100 nodes and edges are drawn as per the friendship node 1 share. This will create a huge number of connections that make a supercomputer unable to analyse this huge data. Another distinctive complexity is in preserving the individual's privacy (Mo & Li, 2015; Bowcott, 2015). With such a huge connectedness, it becomes nearly impossible to maintain an individual's privacy.

Considering the challenges discussed above, it is really a cumbersome activity to make the data suitable for analysis. The type of data can be of various forms. Data collection from multiple sources should necessarily perform data analysis to get some conclusions about the data. What type of analysis should be carried out to what type of data and how to collect and clean different types of data? Answers to these questions come in the form of understanding the different types of data available online. In Section 2 overview the different forms of social data available and the various steps associated with making the data suitable for analysis. Thus, section 2 focuses on the complexities associated with pre-processing social media data. Once the data is collected and cleaned, what type of analysis must be performed? In section 3 discussion on various types of analysis methods and approaches along with specific use cases for the approaches. After getting an insight into the different analysis methods, depending on the problem

Related Content

Computationally Efficient and Effective Machine Learning Model Using Time Series Data in Different Prediction Problems
Dwiti Krishna Bebartaand Birendra Biswal (2021). *Handbook of Research on Automated Feature Engineering and Advanced Applications in Data Science (pp. 218-233).*
www.irma-international.org/chapter/computationally-efficient-and-effective-machine-learning-model-using-time-series-data-in-different-prediction-problems/268757

Applications of Artificial Intelligence and Machine Learning in Achieving SDG 6, 7, and 14
Arti Saxena, Rajeev Kumar, Vijay Kumarand Jyoti Chawla (2024). *Methodologies, Frameworks, and Applications of Machine Learning (pp. 90-107).*
www.irma-international.org/chapter/applications-of-artificial-intelligence-and-machine-learning-in-achieving-sdg-6-7-and-14/342650

Convolution Neural Network Architectures for Motor Imagery EEG Signal Classification
Nagabushanam Perattur, S. Thomas George, D. Raveena Judie Dollyand Radha Subramanyam (2021). *International Journal of Artificial Intelligence and Machine Learning (pp. 15-22).*
www.irma-international.org/article/convolution-neural-network-architectures-for-motor-imagery-eeg-signal-classification/266493

ACO_NB-Based Hybrid Prediction Model for Medical Disease Diagnosis
Amit Kumar, Manish Kumarand Nidhya R. (2021). *Handbook of Research on Disease Prediction Through Data Analytics and Machine Learning (pp. 526-536).*
www.irma-international.org/chapter/aconb-based-hybrid-prediction-model-for-medical-disease-diagnosis/263336

Features Selection Study for Breast Cancer Diagnosis Using Thermographic Images, Genetic Algorithms, and Particle Swarm Optimization
Amanda Lays Rodrigues da Silva, Maíra Araújo de Santana, Clarisse Lins de Lima, José Filipe Silva de Andrade, Thifany Ketuli Silva de Souza, Maria Beatriz Jacinto de Almeida, Washington Wagner Azevedo da Silva, Rita de Cássia Fernandes de Limaand Wellington Pinheiro dos Santos (2021). *International Journal of Artificial Intelligence and Machine Learning (pp. 1-18).*
www.irma-international.org/article/features-selection-study-for-breast-cancer-diagnosis-using-thermographic-images-genetic-algorithms-and-particle-swarm-optimization/277431