


User Activity Classification and Domain-Wise Ranking Through Social Interactions

Ravindra Kumar Singh, National Institute of Technology, Jalandhar, India

 <https://orcid.org/0000-0003-1142-1954>

Harsh Kumar Verma, National Institute of Technology, Jalandhar, India

ABSTRACT

Twitter has gained a significant prevalence among users across numerous domains in the majority of the countries and among different age groups. It serves a real-time micro-blogging service for communication and opinion sharing. Twitter is sharing its data for research and study purposes by exposing open APIs that make it the most suitable source of data for social media analytics. Applying data mining and machine learning techniques on tweets is gaining more interest. The most prominent enigma in social media analytics is to automatically identify and rank influencers. This research is aimed to detect the user's topics of interest in social media and rank them based on specific topics, domains, etc. A few hybrid parameters are also distinguished in this research based on the post's content, post's metadata, user's profile, and user's network feature to capture different aspects of being influential and used in the ranking algorithm. Results concluded that the proposed approach is effective in both the classification and ranking of individuals in a cluster.

KEYWORDS

Apache Kafka, Apache Spark, Elasticsearch, Real-Time Analytics, Social Media Analytics, User Classifications, User Ranking

1. INTRODUCTION

Social media has evolved very dynamically and has become a primary source of feedback, trends, debate, and sentiments across various domains (Singh et al., 2020; Grover et al., 2019). The various social media platforms have become an important channel in collaboration and sharing opinions, thoughts, and experiences (Jansen et al., 2009) due to its publicly broadcasting and uncontrolled interactions on the posts. Messages posted on twitter could be viewed by everyone. Most of the users are interested in multiple domains and sharing their opinion on them. On the contradictory side, few users are very much inclined towards a specific domain and share their viewpoints related to that domain, they generally have a huge follower base and their posts are serving as food for thought or point of discussions among followers and the whole network. Such users are known as influencers. This social impact could be observed in all the business segments, online advertising and promotions have been considered as an important aspect to maintain a good brand reputation on social media platforms (Vernier et al., 2018). Business leaders are looking for different ways to effectively promote their offerings and be competent in social media networks. In such cases, Influencers play a crucial

DOI: 10.4018/IJSDA.20220701.oa5

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

role to help the companies to reach the correct audience/customers. In this scenario, one of the biggest challenges is to find a way to determine the most influential user on the social media network.

Indeed, it's not possible to have a context-independent universal influencer, so automatically discovering the influencer in a specific domain, location or any other stringent criteria is an add on challenge (Montangero et al., 2015). Most of the approaches to solve this problem are simply magnitude-based and taking the absolute count of the concerned feature to rank, that is not an effective approach because these are not highly reliable. A high number of followers or friends don't guarantee the higher engagement on the posts, similarly, user-provided information too could not be reliable, sometimes users fake it. As an example, the user's short bio section is not reflecting the core interest of the user even sometimes it is misleading the facts.

So this research is mainly focused to classify the user's social media activities and rank them on various domains, particularly politics, sports, cinema, business, technology, and others. Additionally, this research has distinguished a few hybrid parameters based on the post's content, post's metadata, user's profile, and user's network feature to capture different aspects of the influencer's ranking.

The rest of the article is arranged as follows. Section 2 presents a brief overview of related work regarding user activity classifications in social media and ranking influencers of specific topics. Section 3 presents an overview of the research design and methodology. Section 4 describes the experimental setup parts of this research by describing the datasets, methods of preprocessing, selections of machine learning models, and their evaluation parameters. Section 5 focuses on the experimentation part along with the analysis and discussion of the results. Finally, section 6 concludes the research and also presents recommendations for further work.

2. BACKGROUND

Twitter is not only utilized for communication and opinion sharing but also considered as a source of recommendation systems and promotion activities. In such cases, finding the users interested in the concerned field or domain is very evident, and boosting the effectiveness of the recommendations and promotions. User's categorization based on their interests will help this cause and limit the targeted users from the huge user base of Twitter. Predicting the behavior and interests of web users is an evolving area of research, and it is a very challenging task to reliably classify users among various categories (Rahman et al., 2019). User behavior classification, profiling, modeling, and prediction for various use cases in different domains, such as commerce, banking, trend analysis, education, medicine, etc., are the hottest area of research in the data analytics field (Sawita Yousukkee, 2016).

The research highlighted the user's interest from Twitter trends using a modeling approach to tweets (Shahzad et al., 2017). They collected the tweets of particular trends and label it with an appropriate category and utilized it for model training after preprocessing of the tweets and later on these models are predicting the categories of the given tweets. This approach is using the Support Vector Machine (SVM) for the predictions and classifies the topics of interest of a user on Twitter. This approach could be utilized in information filtering and prediction systems, especially in personalized recommendation systems, etc. Another research has designed a real-time system for Twitter user profiling based on a supervised machine learning approach to categorize Twitter users into various interest categories like Politics, Entertainment, Entrepreneurship, Journalism, Science & Technology, and Healthcare (Raghuram et al., 2016) based on Tweet-based, User-based and Time-series based features. They utilized numerous classifiers like Support Vector Machines, Naive-Bayes, k-Nearest Neighbours, Decision Tree, and Logistic Regression, and obtained up to 89.82% accuracy in classification. A characteristics analysis of online user's behavior concludes that feature extraction techniques, such as principal component analysis (PCA), independent component analysis (ICA), and self-organizing maps (SOM), helps in detecting anomalies in user behavior (Deshpande et al., 2017). A research investigated the user behavior in an e-commerce site for predicting the buying intention

13 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/article/user-activity-classification-and-domain-wise-ranking-through-social-interactions/281696

Related Content

Spreadsheet-Based Orchestration for Describing and Automating Web Information Access Processes

Jun Fujima, Shohei Yoshihara and Yuzuru Tanaka (2011). *Interdisciplinary Advances in Adaptive and Intelligent Assistant Systems: Concepts, Techniques, Applications, and Use* (pp. 40-62).

www.irma-international.org/chapter/spreadsheet-based-orchestration-describing-automating/46284

Innovation Diffusion Among Heterogeneous Agents: Exploring Complexity with Agent-Based Modelling (ABM)

Nazmun N. Ratna, Anne Dray, Pascal Perez, R. Quentin Grafton and Tom Kompas (2008). *Applications of Complex Adaptive Systems* (pp. 113-141).

www.irma-international.org/chapter/innovation-diffusion-among-heterogeneous-agents/5136

A Web-Enabled, Mobile Intelligent Information Technology Architecture for On-Demand and Mass Customized Markets

M. Ghiassian and C. Spera (2011). *Interdisciplinary Advances in Adaptive and Intelligent Assistant Systems: Concepts, Techniques, Applications, and Use* (pp. 83-123).

www.irma-international.org/chapter/web-enabled-mobile-intelligent-information/46286

Evolutionary Oppositional Moth Flame Optimization for Renewable and Sustainable Wind Energy Based Economic Dispatch: Evolutionary OMFO for R and SW Energy-Based Economic Dispatch

Sunanda Hazra and Provas Kumar Roy (2019). *International Journal of Applied Evolutionary Computation* (pp. 65-84).

www.irma-international.org/article/evolutionary-oppositional-moth-flame-optimization-for-renewable-and-sustainable-wind-energy-based-economic-dispatch/237975

Stabilization of Mechanical Systems with Backlash by PI Loop Shaping

Ahmad Taher Azar and Fernando E. Serrano (2016). *International Journal of System Dynamics Applications* (pp. 21-46).

www.irma-international.org/article/stabilization-of-mechanical-systems-with-backlash-by-pi-loop-shaping/156513