



e-ISSN: 2319-8753 | p-ISSN: 2347-6710

IJRSET

International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 12, Issue 3, March 2023

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.423

9940 572 462

6381 907 438

ijirset@gmail.com

www.ijirset.com



An Trends of Event Detection by Analyzing Social Media Platforms Data

MEIYALAKAN.K, GOKULAPRIYA.R, PAVITHARA.S, KAVIYA.C, SHARIKA.B,

Assistant Professor, Department of CSE, Mahendra Institute of Technology, Mahendhirapuri, Mallasamudram, Namakkal DT, India

Department of CSE, Mahendra institute of technology, Mahendhirapuri, Mallasamudram, Namakkal DT, India

ABSTRACT: A growing number of web users around the world have started to post their opinions on social media platforms and offer them for share. Building a highly scalable evolution prediction model by means of evolution trend volatility plays a significant role in the operations of enterprise marketing, public opinion supervision, personalized recommendation, and so forth. However, the historical patterns cannot cover the systematical time-series dynamic and volatility features in the prediction problems of a social network. This paper proposes an effective approach to deal with popularity trend prediction for microblog events based on SSA and Bi-LSTM model. Firstly, the algorithm of singular spectrum analysis is used to extract trend components of popularity time series from Weibo posts. Then, after time series denoising by SSA model, the event popularity is divided into different trend states by using the box plot analysis. Finally, the paper exploits the Bi-LSTM model to deal with popularity trend prediction with a sequence to label model. Meanwhile, the comparative experiments on two real datasets with three existing methods are conducted. The experimental results show that our model performs best on both datasets with respect to various metrics, which demonstrates the superiority of our proposal. In the future, this paper will explore how to follow some KDD standards, e.g., CRISP-DM standard, to optimize the process of the system. Meanwhile, the other language datasets will also be considered to verify and improve the system performance on cross-language.

I.INTRODUCTION

Social media platforms, for example, Weibo and Facebook, enable web users to post their views in the virtual communities behind screens. This new type of communication has been well accepted and acclaimed for its apparent advantages of low cost and low user interaction risk. However, the obvious benefits have created potential problems that almost everybody may encounter. An increasing number of social users indulge in passing unverified gossip without doing anything meaningful in the virtual community. Therefore, establishing a highly extensible propagation prediction model has recognized the wisdom of employing social media analysis problems. The related research aims to provide reasonable technique for avoiding public opinion outbreaks and maintaining social stability.



Fig 1: Social media platforms

Information popularity prediction in social networks is usually related to the information diffusion models and information-carrying patterns. The primary idea of popularity analysis is to design or learn a prediction model that can accurately reflect a hot event's information propagation law. On the other hand, the temporal evolution of event popularity can be separated into two types (stationary and nonstationary) according to its trend degree of fluctuation. This work is based on the assumption that a hot event's evolution prediction accuracy is closely correlated to its propagation type in a social network. Some of the current excellent literature on prediction and detection problems in engineering applications pays particular attention to deep learning [1–3]. It is now well established from a variety of studies (e.g., Lin, 2020 [1]) that a small dataset can train a high-precision model. Some studies of engineering problems benefit a lot by using hybrid deep learning models [2] or deep fusion models [3]. On the other hand, dynamic linear models (DLM) [4, 5] are quite suitable to be employed for the prediction problem concerning their feature extensibility. DLM has a strong interpretability compared with the deep learning models. Hence, this study seeks to obtain a general popularity prediction framework based on DLM, which will help address the evolutionary prediction problem of information propagation dynamics based on stationary theory and time-series characteristics.

They implemented their system into Twitter data which was classified into two types, i.e., tweets and retweets, and converted them into temporal images. They first transformed these social media graphs into adjacency matrices and then generated grayscale images from these matrices by using the downsampling technique. They also applied a multilayer neural network to analyze the model's performance. Qiu et al. proposed a simple but efficient single-pass clustering method for event detection on social media. Firstly, they developed a matching dataset based on the similarity of the tweet–event pair. Then, they grouped the same event-related tweets using the incremental clustering approach. Their model selected several active events from the existing event library whenever a new tweet was encountered. The corresponding similarity of the new tweet and the candidate events was calculated to determine which event group the new tweet should go to. After that, the existing event library was updated. Ali et al. presented a sentence-level multiclass event detection model for the Urdu language. They used deep learning models and word embedding, one-hot encoding, TF, TF/IDF, and dynamic embedding-based feature vectors to evaluate their model's performance. Their model detected 12 different events. They found the best accuracy with the TF/IDF-based scheme. They also compared their approach with traditional machine learning models, and their DNN approach outperformed others. The preprocessing step removed all the non-Urdu words and sentences, URLs, and special symbols. Additionally, stopwords elimination, tokenization, and sentence filtering were also applied. Generated features were fed into the embedding layer, and output from the embedding layer was fed into the neural network's fully connected/dense layer.



II.RELATED WORK

Social networks play an essential role in our daily life. People join multiple social network platforms, for example, Facebook and Twitter, to enjoy different services. Information propagation through online social networks has proved to be a powerful tool in many situations. The review in [6] on information propagation has highlighted several application disciplines ranging from biology to social sciences, mathematics, physics, and computer science [7]. The classical application in this field is the virus spread analysis in ecology [8], biology [9], and marketing [10].

Previous studies of information propagation prediction in social networks refer to one of the following three tasks: predicting information popularity [11–14], foretelling user influence [15–18], and divining information diffusion paths (links) [19–22]. Some of the literature focuses on the user influence in the social analysis [15, 17]. Some significant studies are concerned with link prediction to reveal the evolution of real social networks [19, 20]. Much of the current literature on the propagation prediction pays particular attention to information popularity, since it can clearly and intuitively reveal the genuine impact of a hot event by employing statistics. For example, the popularity of video content on social network platforms TikTok and YouTube is often regulated by the statistics such as views, followers, favorites, shares, and downloads [23, 24]. The popularity prediction focuses on the whole trend of information diffusion, for example, the propagation ranges [25] and lifespans [26, 27], to provide the valuable decision-making supports for network public opinion monitoring and guidance. This study sets out to propose the universal scheme of the information popularity prediction problems with time-series characteristics.

Weibo 1 is the unique microblog platform of China which can satisfy regular users' demands to send messages across the country. Any public opinion that might spread in society would certainly rely on the Weibo platform in China. There will always be hundreds of millions of users forwarding their posts to Weibo from the portal, forums, moments, and other media if the content is novel or valuable. Weibo social network is popularly accepted as a monitoring center for public opinion in the era of big data. This study takes Weibo as the carrier to deal with popularity prediction considering its influence on all social networks in China.

The popularity prediction in social networks is a typical time-series problem, in which solutions can provide useful decision supports for avoiding negative propaganda, rumor, hotspot disposal, and so forth. The current research on Weibo popularity prediction is mainly based on the methods of epidemic models [8, 9], classification models [3, 1], and regression models [2, 3]. Such approaches highlight the requirement for time-series analysis of social networks. Time-series analysis can help researchers realize the random mechanism of generating time feature sequences, set up the data generation model, and predict the future possible values of time series. However, a systematic understanding and process of how to analyze time-series problems in the evolution prediction of a social network is still lacking.

Traditional predictive models are usually related to information cascade theory [6], divided into graph-based and non-graph-based ones. The former explore the dynamics of information propagation by individual nodes starting from an initial set of nodes and spreading through the network based on a cascade model, for example, linear threshold models [4] and independent cascade models [5]. The latter mathematically study diffusion using population-based dynamics, for example, SIR models [6]. Some recent attention has focused on the application of incorporating both traditional prediction models and time-series factors have shown an example to apply the time series of dynamic data in the task of popularity analysis. This method focuses on the regression modeling of time series corresponding to the propagation of the user-generated content. The investigation pattern of user-generated content popularity over time is supported by Hu et al. [10], who utilize regression models and three time-series features to recognize content changes. Matsubara et al. [13] accompanied additional examination of popularity prediction by applying the SpikeM pattern to fit the above-reported time-series models. The published studies have focused on using a temporal feature as an analytical tool rather than concentrating on the model theory of time series in social networks. Some tiny differences that appeared in the popularity of prediction research may lead to consequential large variations in approach.

III.METHODS

Maintaining security from the analysis of social media data has been challenging since social media contains different data types such as text, images, and videos. Researchers have employed various methods in analyzing these data. Although there are some differences in the methodologies, they all follow the general model shown in Data collection, preprocessing, processing, classification, and visualization are the general steps in this event detection method. In this field of study, most works are based on the collected textual data. That is why it is far more challenging to extract meaningful information from those texts, and hence a data preprocessing step is widely followed. The preprocessed data then move on to the processing steps. The processing steps include different algorithms and other techniques for extracting features used to detect or predict events. Depending on these features, contents are classified as events or not. The last step is to visualize the result, and this visualization can be numeric, textual, or graph-based.

This article reviewed recent event detection work that used social media data. First, we gathered 120 papers from various sources. These sources included Google Scholar, IEEE Explorer, Science Direct, and other scholarly article search engines. Totals of 70, 20, and 30 of the publications we obtained were from Google Scholar, IEEE Explorer, and Science Direct, respectively. We went over the abstracts of these publications after collecting them, and deleted irrelevant ones. Duplicate papers were also eliminated. As a result, we had 67 distinct publications on social media event detection. We thoroughly examined these 67 publications and identified research gaps in the event detection domain. This review primarily focuses on elaborately showing these gaps for future research works. We divided the papers into four categories to provide a full overview of these publications' working procedures.

The articles were categorized as shallow-machine-learning-based, deep-machine-learning-based, rule-based, and other approaches. This classification was made based on the methods utilized. Support vector machine (SVM), K-nearest neighbor (KNN), naïve Bayes (NB), random forest (RF), etc., were classified as shallow-ML-based approaches. In contrast, long short-term memory (LSTM), convolutional neural network (CNN), deep neural network (DNN), and recurrent neural network (RNN) were classified as deep-ML-based approaches. Rule-based approaches include rule-based methods along with rule-based classifiers. Clustering algorithms, the term interestingness technique, and the graph-based process were all included in the other approaches section. These studies covered a wide range of event detection topics: disease events, natural disaster events, rumor events, journalistic events, wellness events, traffic events, city events, etc. Following the categorization, we merged similar types of event detection articles in the same paragraph and linked them. After that, we evaluated the performance of these articles. Tables and bar charts are used to depict the performance evaluations. We also include a qualitative assessment using the Delphi technique, which ultimately provides an overall comparison of our studied articles.

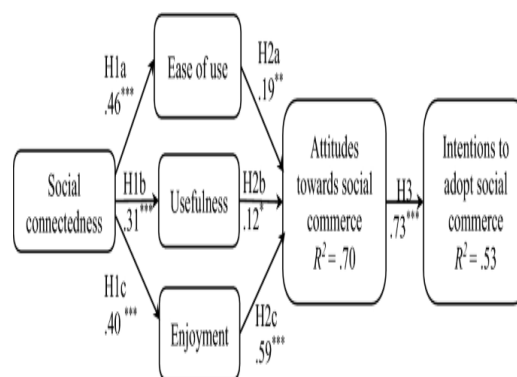


Fig 2: Work Flow

In this section, a detailed description of all the articles is presented. This detailed discussion includes their associated models, their analyzed data, selected features, etc. Different approaches such as shallow machine learning (ML), deep ML, rule-



based, and so on were adopted in this event detection task. Shallow ML includes methods such as SVM, KNN, RF, NB, etc. Deep ML algorithms include DNN, RNN, LSTM, CNN, etc. In the other approaches, sensor tree propagation, microblog clique (MC), bursty event detection in Twitter streams (BEaTS), and so on were used. The following subsections describe the articles adopting these approaches in detecting events.

Shallow-ML-based approaches are mainly supervised learning methods. An event's characterization is predefined in these methods, and events are detected or predicted using the predefined labels or characteristics. Hence, these approaches are suitable for specific domains such as celebrating, protesting, religious, disaster, etc. On the other hand, these shallow ML methods are not ideal for new fields except those defined earlier. They cannot learn from the new events. For this reason, these approaches are not appropriate for unsupervised clustering, or they can not work without predefined attributes. Conventional machine learning algorithms such as SVM, random forest, KNN, NB, etc., are examples of shallow machine learning. The event detection articles utilizing shallow ML approaches are described in this portion. In addition, describes features, languages, and evaluation metrics used in the shallow ML approaches.

IV.RESULT ANALYSIS

he most promising and vital sector of event detection is detecting disruptive events. It is essential to maintain the overall social balance and harmony. Researchers have been using social media to detect disruptive events. Alsaedi and Burnap [8] proposed a novel disruptive event detection approach from real-time Twitter data. The model took Twitter posts as input, and after processing, it created a table of the detected events of a specific region with sub-events at a particular time. The model preprocessed posts using stemming and stopword elimination processes, then the post was applied to the NB classifier for separating event and non-event related tweets. Next, tweet features were calculated and applied to an online clustering algorithm for generating clusters. For this clustering, the tweet's similarity was calculated with the already-existing clusters and then assigned a group based on the calculation. Dey et al. [3] also proposed a model for predicting events (celebrating, protesting, religious, and neutral) from Bengali Facebook posts using Bernoulli naïve Bayes (BNB) classifier. Hence, the collected Bengali Facebook posts were cleaned using preprocessing steps (tokenization and stopwords removal). Preprocessed data were applied to the features extraction process, and they extracted common and event-specific word and phrase frequency along with sentiment score.

Eventually, all these features were applied to the BNB model for predicting event types. Similarly, Hossny and Mitchell proposed a language-independent method that utilized commonly used Twitter keywords for events such as protests, using a binary classification model. They used keywords from word pairs and a spike detection filter to identify context from the preprocessed data. They binarized the event count vector and every word-pair vector. Finally, they used the Jaccard similarity metric for measuring similarity. The highest valued word pairs were used as features for model training, and eventually, they trained the NB classifier with the extracted features. For detecting local violent events in South Africa, Kotzé et al. proposed an approach by analyzing Whatsapp messages employing machine learning, and showed how local violent events could be recorded and coded over time. They applied LR with unigram and Word2Vec features in detecting those events. At first, they preprocessed the data by removing stopwords, cleaning, tokenization, and syntactic parsing, and then used the NLTK stopwords list for the English language and translated English stopwords into their appropriate African words. For annotation purposes, they used four annotations: land grab, farm attack, protest, and crime.

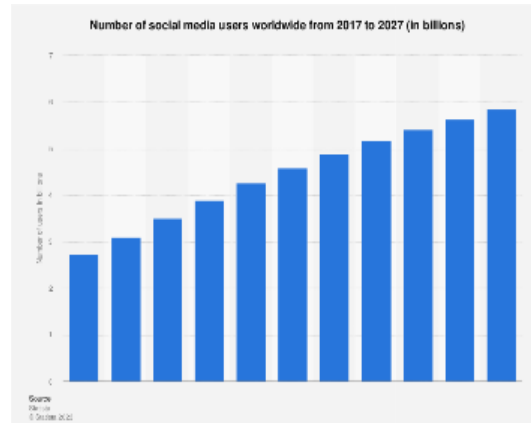


Fig 3: Result Analysis

Apart from the event mentioned above, deep ML techniques have been used to detect other events on social media. Some have detected temporal events while others have tracked events and so on. Chen et al. proposed an online event tracking and detection model using neural networks and similarity metrics. When a tweet arrived, its similarity metric was calculated with the existing tweets. If it already existed, information in the memory was updated, and it created a new id if the event was not in the memory. For similarity metric learning, they employed a function $f(T1, T2)$. It returned a value closer to 1 if the events were similar or a value close to 0 if they were dissimilar. A tweet's low dimension representation was calculated for event tracking, and an entry from memory was selected using the nearest neighbor. If the computed similarity exceeded the threshold for the selected entry, the tweet was relevant and updated. Otherwise, a new entry was created in the memory. Aldhaferi and Lee proposed a framework for temporal event detection on social media.

V.CONCLUSIONS

This review paper focused on user-generated social media data, e.g., text, images, etc., to detect events from these contents. Due to the diversity of users and their shared contents, there has been a trend to use social media for different event-related issues such as announcing, organizing, and collecting man force or gathering for events. These events often create disruptive situations that must be detected earlier to avoid any casualties. Hence, we have tried to accumulate various event detection schemes from the last decade so that new researchers can glean ideas from them and employ new robust models to detect events. From 2009 to 2021, we systematically selected 67 research articles and presented a detailed analysis of targeted events, details on the methods used to identify those events, their performance metrics, and performance analysis. We found that most of the models follow a general approach with data collection, preprocessing, processing, classification, and visualization.

REFERENCES

1. Fang, S.W.; Zhao, N.; Chen, N.; Xiong, F.; Yi, Y.H. Analyzing and predicting network public opinion evolution based on group persuasion force of populism. *Phys. A Stat. Mech. Its Appl.* **2019**, *525*, 809–824. [Google Scholar] [CrossRef]
2. Mao, Y.B.; Bolouki, S.; Akyol, E. On the Evolution of Public Opinion in the Presence of Confirmation BIAS. In Proceedings of the 2018 IEEE Conference on Decision and Control (CDC), Miami, FL, USA, 17–19 December 2018. [Google Scholar]
3. Wang, F.; Li, H.T.; Zuo, J.; Wang, Z. Evolution of online public opinions on social impact induced by NIMBY facility. *Environ. Impact Assess. Rev.* **2019**, *78*, 106290. [Google Scholar] [CrossRef]
4. Mu, L.; Jin, P.Q.; Zhao, J.; Chen, E.H. Detecting evolutionary stages of events on social media: A graph-kernel-based approach. *Future Gener. Comput. Syst.* **2021**, *123*, 219–232. [Google Scholar] [CrossRef]



5. Mu, L.; Jin, P.Q.; Zheng, L.Z.; Chen, E.H. EventSys: Tracking Event Evolution on Microblogging Platforms. In Proceedings of the 23rd International Conference on Database Systems for Advanced Applications, Gold Coast, QLD, Australia, 21–24 May 2018. [Google Scholar]
6. Mu, L.; Jin, P.Q.; Zheng, L.Z.; Chen, E.H.; Yue, L.H. Lifecycle-Based Event Detection from Microblogs. In Proceedings of the Web Conference 2018, Lyon, France, 23–27 April 2018. [Google Scholar]
7. Xia, H.S.; An, W.Y.; Li, J.Z.; Zhang, Z.P. Outlier knowledge management for extreme public health events: Understanding public opinions about COVID-19 based on microblog data. *Socio-Econ. Plan. Sci.* **2020**, 100941. [Google Scholar] [CrossRef]
8. Xiong, J.; Feng, X.D.; Tang, Z.W. Understanding user-to-User interaction on government microblogs: An exponential random graph model with the homophily and emotional effect. *Inf. Process. Manag.* **2020**, *57*, 102229. [Google Scholar] [CrossRef]
9. Yang, J.; Zou, X.M.; Zhang, W.; Han, H.Y. Microblog sentiment analysis via embedding social contexts into an attentive LSTM. *Eng. Appl. Artif. Intell.* **2021**, *97*, 104048. [Google Scholar] [CrossRef]
10. Xu, D.L.; Tian, Z.H.; Lai, R.F.; Kong, X.T.; Tan, Z.Y.; Shi, W. Deep learning based emotion analysis of microblog texts. *Inf. Fusion* **2020**, *64*, 1–11. [Google Scholar] [CrossRef]
11. Hajjem, M.; Latiri, C. Combining IR and LDA Topic Modeling for Filtering Microblogs. *Procedia Comput. Sci.* **2017**, *112*, 761–770. [Google Scholar] [CrossRef]
12. Xu, F.; Sheng, V.S.; Wang, M.W. Near real-time topic-driven rumor detection in source microblogs. *Knowl.-Based Syst.* **2020**, *207*, 106391. [Google Scholar] [CrossRef]
13. Yin, F.L.; Wu, J.L.; Shao, X.Y.; Wu, J.H. Topic reading dynamics of the Chinese Sina-Microblog. *Chaos Solitons Fractals X* **2020**, *5*, 100031. [Google Scholar] [CrossRef]
14. Yin, F.L.; Pang, H.Y.; Xia, X.Y.; Shao, X.Y.; Wu, J.H. COVID-19 information contact and participation analysis and dynamic prediction in the Chinese Sina-microblog. *Phys. A Stat. Mech. Its Appl.* **2021**, *570*, 125788. [Google Scholar] [CrossRef]
15. Pan, J.S.; Li, Y.Q.; Hu, H.P.; Hu, Y. Modeling collective behavior of posting microblogs by stochastic differential equation with jump. *Phys. A Stat. Mech. Its Appl.* **2021**, *578*, 126117. [Google Scholar] [CrossRef]
16. Zeng, Z.M.; Huang, C.Y. Research on public opinion heat trend prediction model of emergent infectious diseases based on BP neural network. *J. Mod. Inf.* **2018**, *38*, 37–52. [Google Scholar]
17. Huang, Y.J.; Chen, F.J.; You, D.D. Research on prediction of network public opinion based on hybrid algorithm and BP neural network. *Inf. Sci.* **2018**, *36*, 24–29. [Google Scholar]
18. Jing, N.; Hu, Y.; Han, X.S. Trend of COVID-19 network attention based on ARIMA and LSTM. *China Saf. Sci. J.* **2020**, *12*, 37–42. [Google Scholar]
19. Feng, Y.; Lv, H.X.; Xu, H.Y.; Wang, R.B.; Zhang, Y.G. The Network Trend Prediction Model of Public Opinion Events Based on SDZ-LSTM. *Inf. Stud. Theory Appl.* **2021**, *44*, 158–163. [Google Scholar]
20. Mughees, N.; Mohsin, S.A.; Mughees, A.; Mughees, A. Deep sequence to sequence Bi-LSTM neural networks for day-ahead peak load forecasting. *Expert Syst. Appl.* **2021**, *175*, 114844. [Google Scholar] [CrossRef]
21. Liang, T.; Zhao, Q.; Lv, Q.Z.; Sun, H.X. A novel wind speed prediction strategy based on Bi-LSTM, MOOFADA and transfer learning for centralized control centers. *Energy* **2021**, *230*, 120904. [Google Scholar] [CrossRef]
22. Shahid, F.; Zameer, A.; Muneeb, M. Predictions for COVID-19 with deep learning models of LSTM, GRU and Bi-LSTM. *Chaos Solitons Fractals* **2020**, *140*, 110212. [Google Scholar] [CrossRef] [PubMed]



INNO SPACE
SJIF Scientific Journal Impact Factor
Impact Factor: 8.118



ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



www.ijirset.com

Scan to save the contact details