2022 8th International Conference on Web Research (ICWR)

Event Detection in Twitter by Weighting Tweet's Features

Parinaz Rahimizadeh, Mohammad Javad Shayegan*
Department of Computer Engineering
University of Science and Culture
Tehran, Iran
parinazrahimizadeh@yahoo.com, shayegan@usc.ac.ir

2022 8th International Conference on Web Research (ICWR) | 978-1-6654-6626-4/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICWR54782.2022.9786242

Abstract— **In recent years, people spend much time on social networks. They use social networks as a place to comment on personal or public events. Thus, a large amount of information is generated and shared daily in these networks. Such a massive amount of information help authorities to accurately and timely monitor and react to events. This unique specification prevents further damages, especially when a crisis occurs. Thus, event detection is attracting considerable interest among social networks research. Since Twitter is one of the most popular social networks that potentially prepare an appropriate bed for event detection, this study has been conducted on Twitter. The main idea of this research is to differentiate among tweets based on some of their features. For this purpose, the proposed methodology applies weights to the three features, including the followers' count, the retweets count, and the user location. The event detection performance is evaluated by scoring potential clusters based on weighting the three mentioned features. The results show that the average execution time and the precision of event detection in the proposed approach have been improved by 27% and 31%, respectively, in comparison to the base method. Another result of this research is detecting more events (including hot events and less important ones) in the presented method.**

Keywords— Event detection; Twitter; weighting; clustering; social network

I. Introduction

Social networks are one of the largest sources of information. By the expanse of these networks, the need for systems that can extract useful information from such a large amount of data is felt. Twitter is one of these social networks that has received much attention recently. Twitter users can receive and post text messages of up to 280 characters, named tweets. Twitter has 330 million active users monthly who generate more than 500 million tweets daily. Thus, a large amount of information is exchanged in this social network that can be utilized as an essential source for reporting real-world events [1-4].

One of the important characteristics of Twitter is its real-time nature. Twitter users share and discuss different kinds of information, from daily personal events to important and global ones, in real-time. Recent studies have shown that reporting and discussing events that users are experiencing are one of the common usages of social networks. These events may contain critical contents that describe the situations throughout a crisis. Monitoring the critical events, crisis management, and decision making can be done via social streams. These capabilities enable authorities to analyze the general situation of an event and make the right decision [3–6]. In the event detection domain, we believe that tweets differ from each other in their weights, and all tweets should not be weighed the same. As the earlier studies have not considered the combination of the three features, including the followers' count, the retweets count, and user location, the novelty and the main idea of this study is to consider weighting the mentioned tweets features for event detection purposes. For clarification, these three features are defined as follow:

The followers count: When a user posts a tweet with high followers, it is of higher value than one with fewer followers, and it can be allocated a higher weight. Influencer users usually publish accurate information; it means the validity of their posts is higher than an ordinary user. These users have more audiences, too.

The retweets count: When a tweet is retweeted frequently, it means the tweet contains important material that can perhaps be event-related. Thus, such a tweet has a higher value than a tweet that is not retweeted and can be allocated a proportionate weight.

The user location: When the probable location of the tweet's author is near the probable location of an event, the tweet's value can be considered higher. Locals usually have quick access to the location of the event, and they can publish the news and its details sooner. Accordingly, such a tweet can be allocated proportional weight. This study aimed to improve event detection performance by weighting the three features mentioned above and utilizing the base method presented in [7].

II. Literature Review

Many studies worked on event detection; this section briefly presents the most relevant ones. Cui, M.Zhang, Liu, Ma and K.Zhang [8] have utilized tags on Twitter as an indicator of events. They have presented a classified algorithm according to three features of hashtags, including instability, Twitter meme possibility, and authorship entropy. Based on these features, hashtags are categorized, and breaking events are detected. In another study, Li, Sun and Datta [9] have proposed a scalable segment-based event detection system, named Twevent. This system consists of three main components: I) tweet segmentation, II) event segment detection, and III) event segment clustering. On the other hand, Ozdikis, Senkul and Oguztuzun [10] have introduced a method Based on lexico-semantic expansion of tweets for improving event detection performance on Twitter. The implemented semantic expansion method is based on first-order and second-order (syntagmatic and paradigmatic, respectively) relationships among words. Moreover, in another study, Pradhan, Mohanty and Lal [11] have detected events by the Bag of Words technique. In this method, a three-phase incremental clustering algorithm was presented for grouping similar tweets effectively. They also

978-1-6654-6626-4/22/$31.00 ©2022 IEEE

offered a heuristics method, named EAAS (Event And Aspects Selection), for detecting an event and its aspects.

McCreadie, Macdonald, Ounis, Osborne and Petrovic [12] have proposed a new scalable method for detecting events. In this method, a new strategy of lexical key partitioning is used to distribute the event detection process among several machines. Furthermore, Choi and Park [13] have presented a method for detecting emerging topics by using High Utility Pattern Mining (HUPM) in Twitter streams. In the HUPM method, both factors of words frequency and word utility are taken into account to detect topics in the pattern generation process. Besides, Nguyen, Ngo, Vo and Cao [14] have introduced a novel method for detecting hot topics on the Twitter data stream. They detected hot topics by incremental clustering, which used named entities and central centroids.

Kaleel and Abhari [15] have presented a novel method for detecting events from tweet clusters based on locality sensitive hashing. In this method, events are detected by matching the event's keywords on cluster labels. Boettcher and Lee [16] have also presented a method for detecting local events, called EventRadar. In this method, the average tweet frequency of keywords is estimated per day in and near a potential event zone. These estimations are then used to categorize whether the keywords are local event-related. On the other hand, Unankard, Li and Sharaf [17] have presented a method for early detection of emerging events in social networks, called LSED. In this method, they utilized the mentioned locations in the tweet text for identifying the event's location.

Katragadda, Virani, Benton and Raghavan [1] have presented a new real-time model in which time-evolving graphs were used to detect events in Twitter streams. They also used a topic evolution model for finding credible events and removing noise. Besides, Barros, Cardoso-Pereira, Loureiro and Ramos [18] have presented a novel real-time method for detecting events on Twitter, which is based on the entropy calculation of the content of tweets. They used the phase transition of bigrams entropy detection for identifying events. In another study, Phuvipadawat and Murata [19] have presented a real-time method for detecting and tracking breaking news on Twitter. In this method, researchers have improved grouping results by boosting proper nouns' scores. Groups also have been ranked based on popularity, reliability, and freshness factors. Asadi, Yari and Shayegan [20] have also improved the performance of event detection in Twitter streams by considering retweet features in a thesis that has been done at the University of Science and Culture. The main idea of this study was based on differentiating between tweets and retweets. In addition, Kumar, Liu, Mehta and Subramaniam [7] have introduced a novel method for solving the challenges of event detection in real-time Twitter streams. In this study, compression distance and single-pass clustering were used to detect events effectively. Despite the different studies on event detection, the idea of the hybrid weighting of tweet's features has not been used yet. Hence, this research aimed at investigating event detection performance by the presented idea. Table I shows a summary and comparison of the previously studied works.

III. Methodology

In this section, research steps, including collecting data, preprocessing data, and event detection process, are presented in detail, and the proposed method is compared with the base method presented in [7].

A. Execution Environment

The current research had been implemented with the help of Python programming language and in PyCharm IDE. Also, The Execution environment on which the research was run had the following specifications:

CPU Core i5; Ram 4 GB; Hard Disk 477 GB; OS Windows 8 x64.

B. Data Collection

The first stage of the study is collecting data. This study has been done on Twitter. For collecting Twitter data, we used Twitter API that Twitter Company provided. In this study, Persian tweets were investigated, and all data were extracted randomly. At first, about 150000 Persian tweets were studied randomly, and 50 top-most frequent words in these tweets were extracted. After that, the main data were extracted by these 50 words to cover all topics and not just focus on special ones. In this research, Twitter data for six days, including 2019.7.31, 2019.8.1, 2019.8.11, 2019.11.15, 2020.1.3, and 2020.1.8, were extracted. Overall, the Obtained dataset consisted of 600000 tweets.

C. Data Preprocessing

In this research, before entering the primary process of event detection, there is a preprocessing stage that its steps have been shown in Fig. 1.

For determining the weights of the follower's count and retweets count features, a sample of tweets about 250000 was first labelled based on whether they are events. After labelling the sample tweets, we utilized the Information Gain method for weighting these two tweets' features [22].

D. The Main Process of Event Detection

The suggested event detection method in this study is based on the method in [7]. The general framework of the suggested method has been presented in Fig. 2.

1) Removing URLs and Tweets Containing Mentions: Some tweets have a link in their text that these links have the same format. For this reason, in the base method [7], tweets containing URLs are sometimes mistakenly classified in the same cluster. Therefore, for avoiding such mistakes and improve the cluster's quality, tweets' ULRs were removed in this study. Also, studying the data of three days showed that, on average, from the total tweets of a day, only 0.2% of tweets containing mentions were event-related, and most of them contained noisy materials (Fig. 3). Hence, in the current study, for removing such noisy tweets and

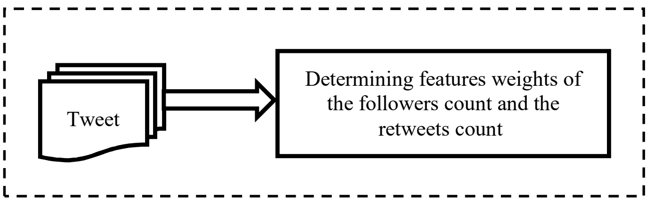

Fig. 1. Preprocessing of the suggested method

TABLE I. SUMMARY AND COMPARISON OF PREVIOUS WORKS

Ref.	***Method***		
	Category [21]	***Name***	***Advantages***
[8]	Thematic	Using Hashtag	• Remove the advertising hashtags • Almost high accuracy • Good performance
[9]	Thematic	Segment based	• Scalable • Make detected events easy to interpret • High precision and recall • Greatly reduces the noise
[10]	Thematic	Based on semantic expansion of tweet contents	• Language independent • Higher F-Score • High processing speed
[11]	Thematic	Using BOW technique	• High accuracy • Efficient performance
[12]	Thematic	Using lexical key partitioning strategy	• Scalable
[13]	Thematic	Using HUPM	• Higher performance • Shorter runtime • Higher recall
[14]	Thematic	Based on Incremental clustering using named entities and central centroids	• Higher Performance • Higher quality clustering • Shorter runtime
[15]	Thematic-Spatial	Using locality sensitive hashing (LSH)	• significantly improved runtime • No need of supplying clusters quantity in advance
[16]	Thematic-Spatial	Detecting local events with the help of location and estimating the average tweet frequency of keywords	• Real-time • High precision
[17]	Thematic-Spatial	Using location	• Efficient performance • Higher precision
[1]	Temporal	Using time evolving graphs	• Eliminate noise • Nearly real-time • Low computational overhead
[18]	Thematic-Temporal	Based on entropy calculation of the content of the tweets	• Correctly detect events that occur almost at the same time • Good overall performance • Real-time
[19]	Thematic-Temporal	Based on grouping with the help of boosting scores on proper nouns	• Improve grouping results • Improve the similarity comparison for short-length messages • Real-time
[20]	Thematic-Temporal	Using single-pass clustering with the help of retweets	• No need for additional processing for retweets • 40% performance improvement
[7]	Thematic-Temporal	Using single-pass clustering	• Nearly real-time • Scalable • Usable for noisy streams • Higher F1 score

increasing the speed of tweets processing, tweets containing mentions have not been processed.

2) Determining the Score of the User Location Feature: In the suggested method, the score of the user location was obtained by the method used in [17], along with some changes that will be explained in detail.

a) Extraction of user location: In this study, we utilized the registered location in the profile information as the user location. Extraction of mentioned locations in the tweet text: Extracting location from a text is considered as one of the challenging stages in this study. As a result, two methods were combined to get all the mentioned locations in the tweet text. The first method is Named Entity

Recognition (NER). The second method is using hashtags in the text. We applied the second method because sometimes the location name is not mentioned in the text of an event-related tweet that is related to a place, whereas it is mentioned in one of the hashtags. Usually, if an event is related to a location, users are more likely to mention that location name in one of the hashtags. Hence, in the current study, a combination of these two methods was applied to extract the mentioned locations in the tweet text.

b) User location scoring criterion: After extracting the user location and the mentioned locations in the tweet text, they were searched in the Gazetteer database, which contains all geographical locations around the world. Finally, the correlation of locations was assigned by (1).

$$\text{LocCorrelateScore} = \alpha_1(F(\text{uContinent,eContinent})) + \alpha_2(F(\text{uCountry,eCountry})) + \alpha_3(F(\text{uState,eState})) + \alpha_4(F(\text{uCity,eCity})) \quad (1)$$

In (1), $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0.25$, and also ucontinent, ucountry, ustate, and ucity are related to the user location and econtinent, ecountry, estate, and ecity are related to the mentioned locations in the tweet text. For calculating (1), if each level has the same value, it is assigned the score. In other words, if x = y, f(x,y) = 1, otherwise f(x,y) = 0.

3) Determining Active Clusters: In the suggested method, the score of the user location was obtained by the method used in [17], along with some changes that will be explained in detail.

4) Clustering Tweets: In this study, when a new tweet enters the system, the value of the tweet's distance from other tweets in a cluster can be calculated by compression

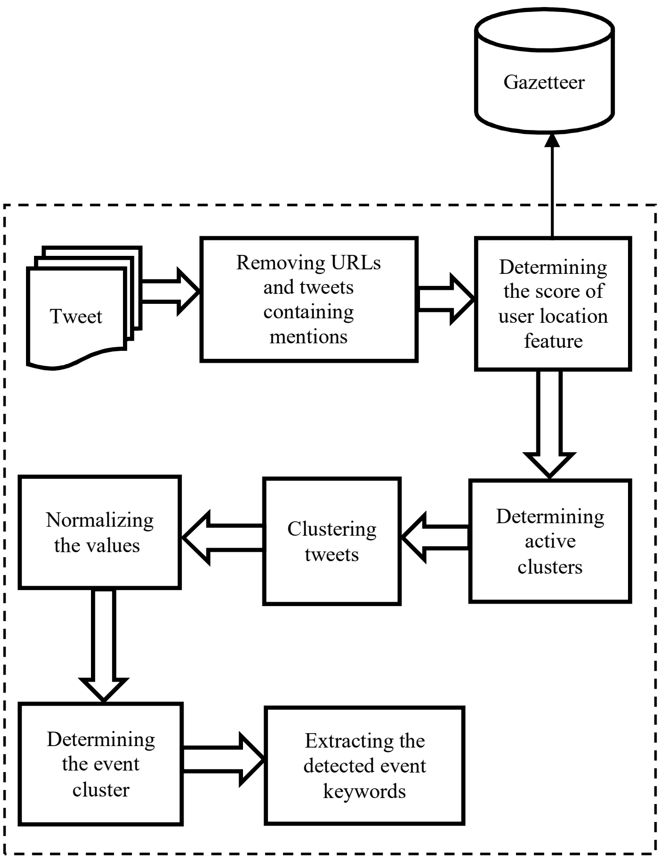

Fig. 2. General framework of the suggested event detection process

distance [23]. If the calculated value is less than the threshold value (D_t) that has been obtained by trial and error, the tweet will be added to that cluster; otherwise, a new cluster will be created and the tweet will be added to it. In this study, each tweet is considered as a document. C is any compressor, C(x) is the compressed size of the tweet x. Thus, the distance between the two tweets x and y is D(x,y) that in (2), it has been defined [24]:

$$D(x, y) = \frac{C(xy)}{C(x)+C(y)} \quad (2)$$

In (2), C(xy) is the obtained compression by merging the two tweets.

5) Normalizing the Values: The follower's count and the retweets count values should be normal to be used in the presented (3), for investigating whether a cluster is an event. Thus, by the min-max method, the values of these two features were normalized.

6) Determining the Event Cluster: All the detected clusters by the algorithm cannot be events. Thus, it needs to utilize a method for identifying the clusters of events from the other clusters. In the current study, we weighed three features of the tweet for this purpose. The following formula has been presented based on research [19].

$$Score_{(c)} = \sum_i ((W_{followers-count} \times n_{followers-count_i}) + (W_{retweets-count} \times n_{retweets-count_i}) + LocCorrelateScore_i) \quad (3)$$

In (3), $n_{followers-count_i}$ is the follower's count of the i^{th} tweet's publisher in the cluster, and $n_{retweets-count_i}$ is the retweets count of the i^{th} tweet in the cluster. $W_{followers-count}$, $W_{retweets-count}$ and $LocCorrelateScore_i$ are the weights of the follower's count, the retweets count, and the user location features, respectively, that have been obtained in the previous stages. Finally, (3) has been utilized to investigate whether a cluster represents an event or not. When the score of each cluster is higher than the threshold value ($score_t$), that cluster will be considered as an event. The threshold value has been obtained by trial and error.

7) Extracting the Detected Event Keywords: The discovered events are usually described by the top-most frequent words of its tweets. Hence, in this study, the top

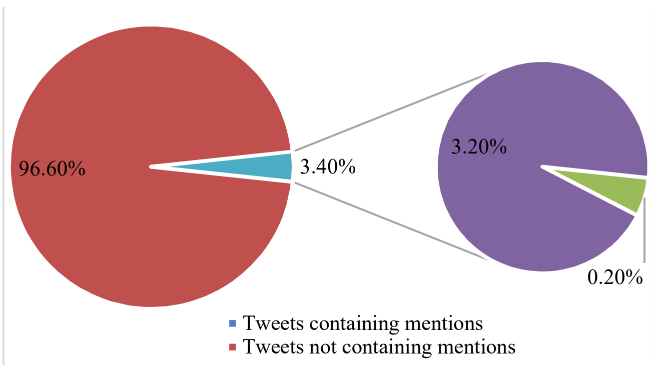

Fig. 3. Studying the effect of tweets containing mentions on event detection

keywords of each event were extracted as its description. After that, these words were matched the ground truth to verify the detected event.

IV. Findings

In this section, the obtained results from testing the suggested method are analyzed.

A. Results of Studying the Tweets Containing Mentions

Based on Fig. 3 and studying the tweets containing mentions during three days of the dataset, the tweets containing mentions comprised only 3.4% of total tweets of a day on average. Also, the average number of tweets containing mentions that were event-related comprised only 0.2% of the total tweets of a day. It is because most of the tweets containing mention are replying to another person's tweet, which is not event-related. Thus, it can be concluded that removing the tweets containing mentions will have no great negative impact on event detection. However, removing them helps noisy data to be removed significantly, which makes the event detection process faster.

B. Results of Event Detection

Table II shows the results of event detection in the collected dataset.

As it can be seen in Table II, most of the events have been detected in both base and suggested methods, but several events have been detected only with the suggested method. These events were of less importance than the other events. This issue shows that the suggested method can detect more events, including hot events and less important ones. The reason for that is that the noisy tweets are ignored in the proposed method by removing URLs and tweets containing mentions, so other tweets have more opportunities to expand their clusters and this leads more clusters to have a chance to be detected as an event. In general, in our proposed method, event clusters are identified by scoring potential clusters with the weighting tweet's features.

C. Results of Studying Location

In this study, two methods were combined to extract all the locations from the tweet text. The first method was utilizing the tweet text to extract location names by NER. The second method was investigating the tweet hashtags. The reason for utilizing both of these methods was that sometimes location names are mentioned only in hashtags or in both texts and hashtags. Fig. 4 shows the importance of using this combined method. As it can be seen in Fig. 4, 68.61% of location names have been mentioned only in tweet texts, 10.72 % only in hashtags, and 20.67% in both tweet texts and hashtags. From the mentioned statistics we can conclude that near 89% of the location names are at least mentioned in the tweet text. However, for covering all the cases, a combination of these two methods was utilized.

D. Results of Execution Time

Based on Fig. 5, the execution time of the event detection process in the base method follows a linear form. It is because of the equal number of extracted tweets on different days. On the other hand, the execution time of the event detection process in the suggested method had some ups and downs, which were caused by the difference in tweets processing time to investigate the location score. It is evident that if a tweet has more mentioned locations, it needs more time to be processed in the Gazetteer database. Chart peaks represent days in which most of its tweets mentioned many locations, and chart troughs represent days in which most of its tweets did not mention any location.

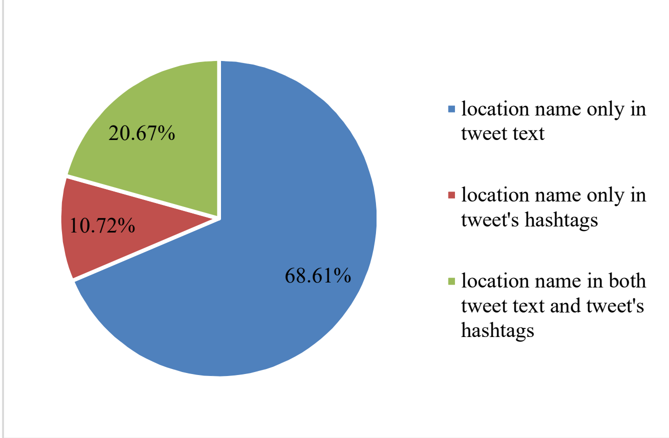

Fig. 4. Results of studying location name in three case

TABLE II. Event Detection Results

Date	*Event Detection*			
	Event keywords*	***Event description***	***Base method***	***Suggested method***
2019/7/31	Remove-currency-zero	Removing 4 zeros from national currency	✓	✓
	Caspian sea	Protest against Iran's share in the Caspian sea	✓	✓
	Broadcast	Broadcasting day	✗	✓
2019/8/1	Sanction-Zarif-Foreign Minister	The US sanctions Iran's foreign minister	✓	✓
	#project2533	2533 civil projects of deprivation elimination in Sistan and Baluchestan	✗	✓
	#increase-capacity-threat-health #combat-congress-with-monopoly	Reactions to medical field capacity in universities	✓	✓
2019/8/11	Iraq-#Iraq_will_not_burn-#Iran_and_Iraq_can_not_be_separated	Reactions to demonstrations in Iraq	✓	✓
2019/11/15	petrol	Petrol rationing	✓	✓
2020/1/3	Assassination-General-Soleimani #harsh_revenge	The assassination of General Soleimani	✓	✓
2020/1/8	Missile-Base-Iran-#harsh_revenge	Iran missile attack to USA base in Iraq	✓	✓
	Ukrainian airplane	Ukrainian airplane crash	✓	✓

*All the presented Event keywords in this table are translated from Persian to English.

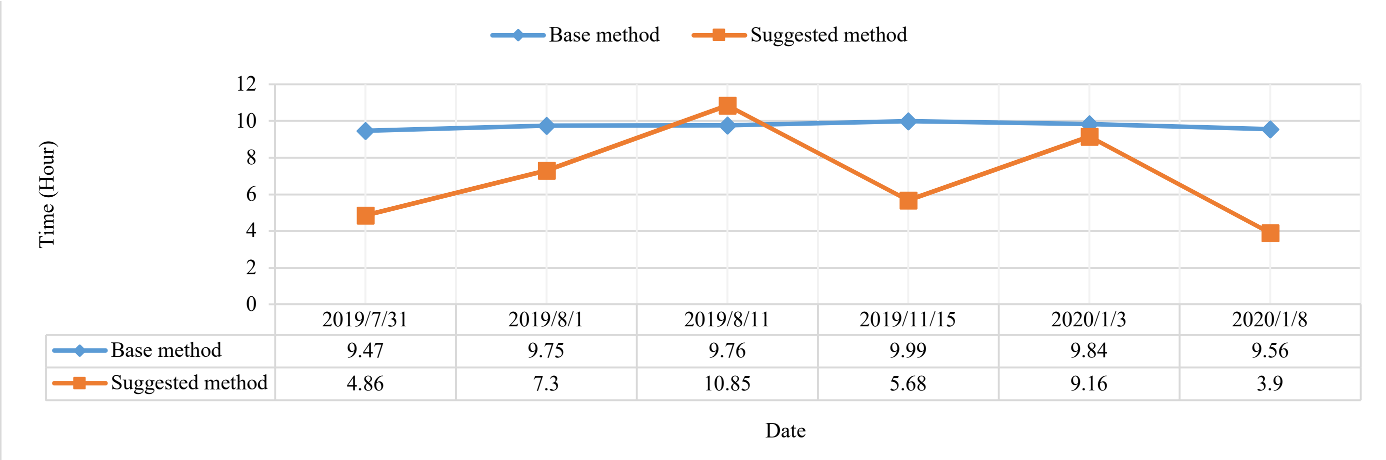

Fig. 5. Comparing execution times

According to Fig. 5, the average execution time of event detection has decreased in the suggested method than the base method. Overall, the execution time of event detection has improved by 27% on average. The main reason for this improved execution time was ignoring or not processing tweets containing mentions that were mostly noisy and were not event-related. Ignoring tweets containing mentions was a simple process that can help the execution speed of event detection significantly. Another reason was utilizing the method which weighted the tweet's features for detecting event clusters. It is due to scoring potential event clusters by weighting the tweet's features helps to identify event clusters quicker than the base method which, uses user diversity of a cluster for detecting event clusters.

E. Results of Validity (Precision)

Precision is one of the criteria that has been considered for comparing the suggested method performance with the base method. Fig. 6 shows the results. Precision is defined as (4):

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{4}$$

Fig. 6 shows the values of event detection precision in both suggested and base methods for each day separately. The results show that the suggested method increases the precision of event detection on average. The improvement happens as a result of ignoring tweets containing mentions and removing URLs from tweets texts because the noisy tweets will not be processed, and the remaining tweets will have more chance to expand their relative cluster.

The reason for the significant difference between the two methods on some dates, like 2019/7/31, was the existence of numerous tweets containing mentions in those dates. Hence, by ignoring these tweets in the suggested method, the effect of these kinds of noise can be prevented. As a result, the precision of event detection has improved 31% on average than the base method.

V. Conclusion and Future Works

Different kinds of social networks that have been popular recently are rich sources of information. Information from these social networks can be used for detecting events

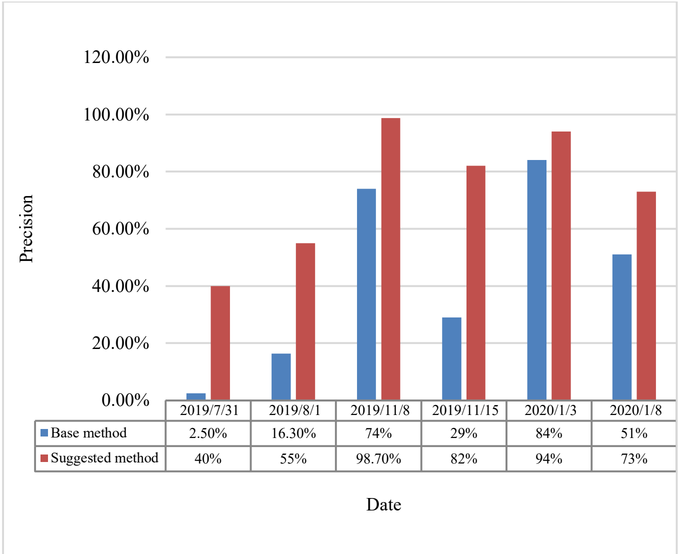

Fig. 6. Comparing the precision of methods

so the authorities can use them to have more accurate and timely reactions in critical conditions. Twitter is one of the popular social networks which based on its entity and function, can be used for informing events. Furthermore, the identified gap from the previous studies led us to investigate event detection performance by considering the three tweet features, including the followers' count, the retweets count, and the user location. In general, the main idea and the novelty of this study was weighting the mentioned tweet's features for event detection. Finally, by implementing the base and suggested method, we concluded that the suggested method was able to detect more events, including hot and less important ones. The average execution time in the suggested method has improved by 27% than the base method. Also, the average precision of event detection has improved by 31% than the base method. Generally, it can be said that the suggested method had better performance than the base method.

Event detection is a domain with many potentials. For future researches, another social network like Telegram that is very popular among Iranians can be studied. Another suggestion is using the combined information of several social networks. Also, the effect of other tweets' features in event detection can be studied.

REFERENCES

[1] S. Katragadda, S. Virani, R. Benton and V. Raghavan, "Detection of event onset using Twitter," in 2016 International Joint Conference on Neural Networks (IJCNN), Vancouver, BC, Canada, 2016, pp. 1539–1546.

[2] S. Petrovi´c, M. Osborne and V. Lavrenko, "Streaming first story detection with application to Twitter," in Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, Los Angeles, CA, USA, 2010, pp. 181–189.

[3] T. Sakaki, M. Okazaki and Y. Matsuo, "Tweet analysis for real-time event detection and earthquake reporting system development," IEEE Transaction on Knowledge and Data Engineering, vol. 25, no. 4, pp. 919–931, Apr. 2013.

[4] X. Zhou and L. Chen, "Event detection over twitter social media streams, " VLDB Journal, vol. 23, no. 3, pp. 381–400, June 2014.

[5] A. Guille and C. Favre, "Event detection, tracking, and visualization in Twitter: a mention-anomaly-based approach," Social Network Analysis and Mining, vol. 5, no. 1, p. 18, May 2015.

[6] W. Dou, X. Wang, W. Ribarsky and M. Zhou, "Event detection in social media data," Proc. IEEE VisWeek Work. Interact. Vis. Text Anal. Driven Anal. Soc. Media Content, Seattle, WA, USA, 2012, pp. 971–980.

[7] S. Kumar, H. Liu, S. Mehta and L. V. Subramaniam, "Exploring a scalable solution to identifying events in noisy Twitter streams," in Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015 - ASONAM '15, Paris, France, 2015, pp. 496–499.

[8] A. Cui, M. Zhang, Y. Liu, S. Ma and K. Zhang, "Discover breaking events with popular hashtags in Twitter," in Proceedings of the 21st ACM International Conference on Information and Knowledge Management - CIKM '12, Maui, HI, USA, 2012, pp. 1794.

[9] C. Li, A. Sun and A. Datta, "Twevent," in Proceedings of the 21st ACM International Conference on Information and Knowledge Management - CIKM '12, Maui, HI, USA, 2012, pp. 155.

[10] O. Ozdikis, P. Senkul and H. Oguztuzun, "Semantic expansion of Tweet contents for enhanced event detection in Twitter," in 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, Istanbul, Turkey, 2012, pp. 20–24.

[11] A. K. Pradhan, H. Mohanty and R. P. Lal, "Event detection and aspects in Twitter: A BoW approach," In: Fahrnberger G., Gopinathan S., Parida L. (eds) Distributed Computing and Internet Technology. ICDCIT 2019. Lecture Notes in Computer Science, vol. 11319, Cham, Switzerland: Springer, 2019, pp. 194-211.

[12] R. McCreadie, C. Macdonald, I. Ounis, M. Osborne and S. Petrovic, "Scalable distributed event detection for Twitter," in 2013 IEEE International Conference on Big Data, Santa Clara, CA, USA, 2013, pp. 543–549.

[13] H.-J. Choi and C. H. Park, "Emerging topic detection in twitter stream based on high utility pattern mining," Expert Systems with Applications, vol. 115, pp. 27–36, Jan. 2019.

[14] S. Nguyen, B. Ngo, C. Vo and T. Cao, "Hot topic detection on twitter data streams with incremental clustering using named entities and central centroids," in RIVF 2019 - Proceedings: 2019 IEEE-RIVF International Conference on Computing and Communication Technologies, Danang, Vietnam, 2019, pp. 1–6.

[15] S. B. Kaleel and A. Abhari, "Cluster-discovery of Twitter messages for event detection and trending," Journal of Computer Science, vol. 6, pp. 47–57, Jan. 2015.

[16] A. Boettcher and D. Lee, "EventRadar: A real-time local event detection scheme using Twitter stream," in 2012 IEEE International Conference on Green Computing and Communications, Besancon, France, 2012, pp. 358–367.

[17] S. Unankard, X. Li and M. A. Sharaf, "Emerging event detection in social networks with location sensitivity," World Wide Web, vol. 18, no. 5, pp. 1393–1417, Sep. 2015.

[18] P. H. Barros, I. Cardoso-Pereira, A. A. F. Loureiro and H. S. Ramos, "Event detection in social media through phase transition of bigrams entropy," in 2018 IEEE Symposium on Computers and Communications (ISCC), Natal, Brazil, 2018, pp. 1–6.

[19] S. Phuvipadawat and T. Murata, "Breaking news detection and tracking in Twitter," in 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology, Toronto, ON, Canada, 2010, pp. 120–123.

[20] M. Asadi, A. Yari and M. J. Shayegan, "Performance improvement of event detection in Twitter's streaming text using retweets (In Persian)," M.S. thesis, Dept. Comp. Eng., University of Science and Culture, Tehran, Iran, 2018.

[21] A. Goswami and A. Kumar, "A survey of event detection techniques in online social networks," Social Network Analysis and Mining, vol. 6, no. 1, Nov. 2016.

[22] Y. Zhang, Z. Xu and Q. Yang, "Predicting popularity of messages in Twitter using a feature-weighted model," International Journal of Advanced Intelligence, pp. 1–7, 2012.

[23] E. Keogh, S. Lonardi, and C. A. Ratanamahatana, "Towards parameter-free data mining," in Proceedings of the 2004 ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '04, Seattle, WA, USA, 2004, p. 206.

[24] S. Kumar, H. Liu, S. Mehta, and L. V. Subramaniam, "From Tweets to Events: Exploring a Scalable Solution for Twitter Streams," arXiv Prepr. arXiv1405.1392, May 2014.